UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"
Scuola di Dottorato di Ricerca in Fisica - XXXV ciclo
Tesi di Dottorato

Measurement of W boson production in association with jets and WW vector boson pair production at the LHCb experiment.

Relatrice
Prof.ssa Donatella Lucchesi
Correlatore
Dott. Lorenzo Sestini

Candidato
Davide Zuliani

Coordinatore Scuola di Dottorato
Prof. Giulio Monaco

Anno Accademico 2022/2023

Contents

Abstract

This thesis addresses the measurements of W boson production in association with jets, and WW vector boson pair production at the LHCb experiment. For both analyses, a high-energetic muon and a least one or two jets are required in the LHCb acceptance. The $W+$ jets analysis performs a measurement of the cross-section, extending previous results obtained at a center-of-mass energy of 8 TeV. In the WW measurement, one W boson is required to decay hadronically into two jets, and the cross-section is measured for the first time in the forward region studied by the LHCb experiment. In addition, studies on high-energetic electron reconstruction performance in view of future LHCb upgrades are presented. Finally, insights on new techniques for jet identification based on new innovative algorithms are presented.

This is the end,
beautiful friend.
This is the end,
my only friend.

(The End, The Doors)

Introduction

The LHCb experiment at the Large Hadron Collider (LHC) at CERN, has been designed to perform b- and c-hadron physics measurements. The detector is similar to a single-arm spectrometer that covers the forward region of proton-proton collisions, where the b-hadron cross-section is enhanced. The main activities of the LHCb Collaboration are dedicated to the study of these heavy flavors. In the latest years, after the discovery of the Higgs boson and any missing signal of supersymmetry or any other new physics at very high energy, it is becoming more and more clear that the electroweak sector needs to be investigated with higher precision. The excellent LHCb detector performance is opening the possibility to perform high precision measurements even though the data collected are not yet enough to be competitive with the other LHC experiments. On the other hand, LHCb covers the very forward region with respect to the beam axis, where electroweak measurements can be more sensitive to modification induced by new physics. With its cons and pros, LHCb is starting to contribute in a unique way to the test of the electroweak sector of the Standard Model. The primary goal of this thesis is to study the WW bosons production in the forward region. That measurement will allow Standard Model tests together with the possibility to reconstruct the Higgs boson in this final state. Before arriving to that stage, an intermediate step is necessary, the determination of the $W+$ jets cross section that is missing at LHCb at a center-of-mass energy $\sqrt{s} = 13$ TeV.

This measurement by itself is important since it provides a test of the Standard Model and allows constraining Parton Distribution Functions, *i.e.* functions describing the probability of a parton having a certain fraction of the proton momentum, in the forward region of proton-proton collisions. The precise determination of the Parton Distribution Functions is one of the major elements limiting the precision of crucial electroweak measurements like the W boson mass.

Data collected at the LHCb experiment during the Run 2 (2016-2018) data-taking campaign, corresponding to an integrated luminosity $\mathcal{L} = 5.4$ fb^{-1} are analyzed to reconstruct $W+$ jets and $W(\rightarrow \mu\nu)W(\rightarrow$ jets) final states. The procedure requires a muon and at least one or two jets (depending on the final state) to be reconstructed in the LHCb acceptance. The $W(\rightarrow e\nu)W(\rightarrow$ jets) is not considered due to detector limitations, as discussed. The second part of the thesis is devoted to the investigation of possible future improvements to increase the LHCb sensitivity to electroweak physics. The first one concerns the capability of reconstructing highly energetic electrons. LHCb is proposing an upgrade of the Electromagnetic Calorimeter (ECAL) and the occasion to design it to be fully efficient for electroweak measurements is an occasion not to be missed. The second possible improvement could come from new technology: the exploitation of quantum computing technologies to have insight into jet substructure that could help to identify jets coming from different sources.

The thesis is organized as follows:

- Chapter 1: is devoted to the description of the theoretical framework. Particular care is taken in describing the Electroweak theory, Quantum Chromodynamics, and the physics of hadron colliders, highlighting the importance of $W+$ jets and $W(\rightarrow \mu\nu)W(\rightarrow$ jets) measurements in the

forward region of proton-proton collisions. Finally, an overview of the main experimental results for $W+$ jets and WW measurements is presented.

- Chapter 2: focuses on the description of the LHCb experiment. LHCb is a forward spectrometer initially designed to study heavy-flavor hadron physics in the forward region of proton-proton collisions. The phase space region studied by LHCb is complementary with respect to the other two LHC experiments, ATLAS and CMS. In this Chapter, an accurate description of the LHCb detector is presented, focusing on the several sub-systems used in the measurements.
- Chapter 3: describes the algorithms used to reconstruct the main physics objects of these measurements. Particularly, algorithms for muon reconstruction and jets reconstruction and identification are presented and discussed in detail, together with their performance.
- Chapter 4: in this Chapter, the $W+$ jets cross-section measurement is presented. In this analysis, the muon coming from the W boson decay is reconstructed, together with at least one or two jets in the LHCb acceptance. The cross-section measurement is performed in a fiducial region defined by putting requirements on the kinematics of the events. Differential cross-section measurements are presented, as a function of the muon and the leading jet kinematic observables. Results are compared with theoretical predictions obtained at next-to-leading order accuracy. It is the first measurement of $W+$ jets at $\sqrt{s} = 13$ TeV in the forward region, extending the previous results obtained at the LHCb experiment using Run 1 data at $\sqrt{s} = 8$ TeV.
- Chapter 5: the measurement of the $W(\rightarrow \mu\nu)W(\rightarrow \text{jets})$ process is presented in this Chapter. In this analysis, one W boson is required to decay semileptonically into a muon and a neutrino, while the other W boson is required to decay hadronically into two jets. The same fiducial region defined in Chapter 4 is used in this analysis. The main background contributions come from $W+$ jets and QCD processes, which have been already analyzed in Chapter 4. Machine Learning techniques are employed to separate WW events from $W+$ jets events. Despite the measurement being statistically limited, this analysis poses solid foundations for future measurement in the next data-taking campaigns.
- Chapter 6: is devoted to studies of electron reconstruction in the context of future LHCb upgrades. The analysis presented in Chapter 5 does not consider events where a W boson decays into an electron and a neutrino, the main motivation being the limitations in the electron reconstruction performance of the ECAL. After a brief overview of the plans for future LHCb upgrades, this Chapter describes studies of high-energetic electron reconstruction with the upgraded ECAL and the impact on jets reconstruction performance of a new proposed experiment layout.
- Chapter 7: an exploratory study for new jet identification techniques is presented. New tools of Quantum Machine Learning, where a quantum circuit replaces a Deep Neural Network, are applied to a jet identification task performed at the LHCb experiment. The performance of these new algorithms is evaluated in view of finding possible improvements with respect to state-of-the-art Machine Learning algorithms.

Chapter 1

Vector bosons: theory and experimental results

The main purpose of this thesis is the evaluation of the cross-sections of $W^{\pm}$+ jets and WW processes. To properly understand the impact of these measurements, a theoretical introduction is necessary. This Chapter is indeed devoted to introduce this theoretical framework. The structure of the Chapter is the following:

- Section 1.1 introduces the Standard Model of particle physics and its particle content;
- Sections 1.2 and 1.3 introduces the main aspects of the Electroweak theory and Quantum Chromodymanics;
- Section 1.4 deals with the physics of hadron colliders, with particular care in understanding the impact of Parton Distribution Functions in the current measurements;
- Sections 1.5 and 1.6 go deeper in the physics of $W^{\pm}$+ jets and WW, to properly understand the measurements presented in Chapters 4 and 5;
- Section 1.7 describes the most recent experimental results for $W^{\pm}$+ jets and WW processes;
- finally, Section 1.8 explains the motivations behind the two analysis presented in Chapters 4 and 5.

1.1 The Standard Model of particle physics

The Standard Model (SM) of particle physics [1–4] is the current theory that describes particles and their interactions. It is based on the equations of motions of fields with a Lagrangian density, and this Lagrangian density invariance under given transformations describes the symmetry of the system. The SM is a Quantum Field Theory based on the symmetry group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, where each gauge group describe a precise kind of interaction between particles: $\mathrm{SU}(3)_C$ is responsible for the strong force, related to the theory of colour charge called Quantum Chromodynamics (QCD), $\mathrm{SU}(2)_L$ defines the isospin invariance of the weak interactions while $\mathrm{U}(1)_Y$ is responsible for the local phase invariance of electromagnetic interactions. The last two groups unify to form the Electroweak theory, described in the next Section.

1.1.1 Particle content

The SM theoretical framework is based on its particle content. Both matter and forces are described in terms of particles, respectively called *fermions* and *bosons*.
Fermions are point-like particles with spin 1/2, where the spin denotes their intrinsic angular momentum. They are divided into two groups called *leptons* and *quarks*, with both groups arranged in a typical "tabular" prescription: up-type and down-type quarks, charged leptons and neutrinos, each with three generations. Leptons consist of three families of charged leptons (e,μ, and τ) each carrying unit charge, respectively paired to a neutrino (ν_e, ν_μ and ν_τ). Quarks on the other side carry a fractional charge, with the up-type quarks (u,c and t) carrying a +2/3 charge, each one paired with a down-type quark (d, s and b), carrying a -1/3 charge.
Bosons are spin 1 particles that behave as the mediator of forces described by the SM: electromagnetic, weak, and strong forces. So far the SM offers no description of the gravitational interaction. The interaction happens through an exchange of a boson between two particles, conserving the charge related to that boson. While quarks interact through all forces, leptons couple only to electromagnetic and weak bosons. Four bosons are responsible for mediating forces: the photon γ, an electrically-neutral and massless boson responsible for the electromagnetic interaction; the charged $W^\pm$ and the neutral Z, massive bosons responsible for the weak force; the gluon g, a massless particle which mediates the strong force. Finally, the Higgs boson H (and its respective Higgs field) is responsible for the mass of fermions and bosons. The particle content of the SM is presented in Tab. 1.1.

	Force	Particle	Mass/GeV	Spin	q/e
Bosons	Electromagnetism	γ	0	1	0
	Weak	$W^\pm$	80.385 ± 0.015	1	± 1
		Z	91.1876 ± 0.0021	1	0
	Strong	g	0	1	0
	Higgs	H	125.9 ± 0.04	0	0
	Generation	Particle	Mass/MeV	Spin	q/e
Leptons	1	e	$0.510998928 \pm 0.000000011$	1/2	-1
		ν_e	-	1/2	-1
	2	μ	$105.6583715 \pm 0.0000035$	1/2	-1
		ν_μ	-	1/2	-1
	3	τ	1776.82 ± 0.16	1/2	-1
		ν_τ	-	1/2	-1
Quarks	1	u	2.3 ± 0.7	1/2	$+2/3$
		d	4.8 ± 0.5	1/2	$-1/3$
	2	c	1275 ± 25	1/2	$+2/3$
		s	95 ± 5	1/2	$-1/3$
	3	b	$173070 \pm 520 \pm 720$	1/2	$+2/3$
		t	4180 ± 30	1/2	$-1/3$

Table 1.1: Particle content of the SM. Particles are divided in bosons and fermions, with fermions divided into leptons and quarks. For each boson (fermion), their force (generation), mass, spin, and electric charge (q) are presented [5].

1.2 Electroweak theory

The Electroweak (EW) theory describes electromagnetic and weak interactions as manifestations of the same force. The gauge group of the EW theory is

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

where $\mathrm{SU}(2)_L$ describes the isospin invariance and $\mathrm{U}(1)_Y$ is the gauge group of the Quantum Electron Dynamics (QED) theory, which preserves the local phase invariance. The EW theory can be described by a Lagrangian:

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{f}} + \mathcal{L}_{\text{Higgs}} \tag{1.2}$$

where $\mathcal{L}_{\text{gauge}}$ describes the interaction through vector bosons and $\mathcal{L}_{\text{f}}$ describes the particle content, *i.e.* fermions, and $\mathcal{L}_{\text{Higgs}}$ describes the Higgs mechanism, through which SM particles acquire mass. The gauge lagrangian $\mathcal{L}_{\text{gauge}}$ is

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} W^i_{\mu\nu} W^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.3}$$

where the field strength tensors $W^i_{\mu\nu}$ and $B_{\mu\nu}$ are defined as

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\nu W^i_\mu - \partial_\mu W^i_\nu - g_W \varepsilon^{ijk} W^j_\mu W^k_\nu \\ B_{\mu\nu} &= \partial_\nu B_\mu - \partial_\mu B_\nu \end{aligned} \tag{1.4}$$

where g_W is the coupling constant of the theory. $B_{\mu\nu}$ is invariant under $U(1)$, while $W_{\mu\nu}$ is invariant under $SU(2)$. The fermion lagrangian $\mathcal{L}_{\text{f}}$ describes the coupling between the massless boson and fermions. Fermions are divided into left-handed (Ψ_L) and right-handed (ψ_R): while left-handed fermions are formed by doublets grouping the corresponding fermions of the same generation, right-handed ones are singlets. The fermion Lagrangian is given by

$$\mathcal{L}_{\text{f}} = \sum_{fermions} \bar{\Psi}_L \gamma^\mu D^L_\mu \Psi_L + \bar{\psi}_R \gamma^\mu D^R_\mu \psi_R, \tag{1.5}$$

where $D^{L,R}_\mu$ are the covariant derivatives defined as

$$\begin{aligned} D^L_\mu &= \partial_\mu + i g_W \frac{\tau_i}{2} W^i_\mu + i g_B Y B_\mu \\ D^R_\mu &= \partial_\mu + i g_B Y B_\mu \end{aligned} \tag{1.6}$$

where τ_i are the Pauli matrices, g_B is a coupling constant and Y is the weak hypercharge. This Lagrangian is invariant under $SU(2) \times U(1)$ transformations, and it is evident from Eq. 1.6 that left-handed fermions interact with both W and B fields, whereas right-handed fermions interaction only with the B field. In the Lagrangians described so far, all fermions and the four-vector fields (W^1, W^2, W^3, B) are massless. Introducing a mass term in the form

$$-m\bar{\psi}\psi = -m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R) \tag{1.7}$$

breaks the $SU(2)_L$ gauge invariance of the theory. In order to properly enable fermions and bosons to acquire their mass, but keeping the invariance under $SU(2)_L \times U(1)_Y$, the last term $\mathcal{L}_{\text{Higgs}}$ is necessary. $\mathcal{L}_{\text{Higgs}}$ introduces the Higgs scalar field Φ [6, 7], a $SU(2)$ doublet that breaks the electroweak symmetry by acquiring a non-zero vacuum expectation value. By interacting with the Higgs field, bosons and fermions acquire mass. Particularly, $\mathcal{L}_{\text{Higgs}}$ is described as

$$\mathcal{L}_{\text{Higgs}} = \mathcal{L}_{H,\text{bosons}} + \mathcal{L}_{H,\text{fermions}} \tag{1.8}$$

where

$$\mathcal{L}_{H,\text{bosons}} = (D^L_\mu \Phi)^\dagger (D^{L\mu}\Phi) - \mu^2 \Phi\Phi^\dagger - \lambda(\Phi^\dagger\Phi)^2 \tag{1.9}$$

with μ and λ real constants, while $\mathcal{L}_{H,\text{fermions}} = \mathcal{L}_{H,\text{leptons}} + \mathcal{L}_{H,\text{quarks}}$, where

$$\begin{aligned} \mathcal{L}_{H,\text{leptons}} &= -g_N \bar{\Psi}_{NL} \psi_{NR} \Psi + \text{h.c.} \\ \mathcal{L}_{H,\text{quarks}} &= -\bar{\Psi}'_{ML} G_{MN} \psi_{NdR} \Phi - \bar{\Psi}'_{ML} \tilde{G}_{MN} \psi_{NuR} \tilde{\Phi} + \text{h.c.} \end{aligned} \tag{1.10}$$

The couplings g_N, G_{MN} and $\tilde{G}_{MN}$ are proportional to the fermion masses, that are acquired after the EW symmetry breaking [8–13]. The masses of the leptons are simply defined as

$$m_N = \frac{g_N \nu}{\sqrt{2}} \tag{1.11}$$

where ν is the vacuum expectation value of the Higgs, while quarks masses are obtained after diagonalizing G_{MN} and $\tilde{G}_{MN}$, obtaining

$$\begin{aligned} M^u &= \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{bmatrix} = \frac{\nu}{\sqrt{2}} V_L G_{MN} V_R^\dagger \\ M^d &= \begin{bmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{bmatrix} = \frac{\nu}{\sqrt{2}} \tilde{V}_L \tilde{G}_{MN} \tilde{V}_R^\dagger \end{aligned} \tag{1.12}$$

The matrix that connects the interaction eigenstates with the mass eigenstates is called the Cabibbo-Kobayashi-Maskawa matrix V_{CKM} [14, 15], and it is given by

$$V_{CKM} = \tilde{V}_L V_L \tag{1.13}$$

V_{CKM} allows for mixing between flavor and mass eigenstates across the quark generations, and it is described by three rotation angles and one complex phase. Also, bosons acquire mass through the EW symmetry breaking. Indeed, the four massless fields (W^1, W^2, W^3, B) can be combined to obtain four "new" fields defined as

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \pm i W^2_\mu) \\ Z_\mu &= \cos(\theta_W) W^1_\mu - \sin(\theta_W) B_\mu \\ A_\mu &= \sin(\theta_W) W^1_\mu + \cos(\theta_W) B_\mu \end{aligned} \tag{1.14}$$

where θ_W is the Weinberg angle [2]. While the A_μ field describes the photon and therefore is massless, the other fields defining $W^\pm$ and Z bosons are massive. The masses of the bosons are defined as

$$\begin{aligned} m_W &= \frac{1}{2}\nu g_W \\ m_Z &= \frac{1}{2}\nu \sqrt{g_W^2 + g_B^2} \end{aligned} \tag{1.15}$$

1.3 Strong interaction and QCD

The theory of Quantum Chromodynamics (QCD) describes strong interactions. QCD introduces a new charge called "color charge", and interactions between quarks happen through the exchange of gluons. In contrast with QED, which describes interactions through the exchange of photons, the gluons in QCD also have color charges and interact with each other through the emission and absorption of gluons. There are three different types of color charge - red, green, and blue - and their corresponding

anti-colors. Quarks have a single color, while gluons are composed of a color and/or an anti-color. The gauge group of QCD is $SU(3)_C$, and its Lagrangian is given by

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_a(i\gamma^\mu D_\mu^{ab} - m\delta^{ab})\psi_b - \frac{1}{4}F_{\mu\nu}^A F_A^{\mu\nu}, \tag{1.16}$$

with the covariant derivative defined as

$$D_\mu^{ab} = \delta_\mu^{ab}\partial_\mu + ig_s G_\mu^A T_A^{ab} \tag{1.17}$$

and the gluon field G_μ^A propagating term given by

$$F_{\mu\nu}^A = \partial_\mu G_\nu^A - g_s f_{ABC} G_\mu^B G_\nu^C. \tag{1.18}$$

The field ψ carries a color charge, with the quantum number of the color charge given by a and b; as there are $N_C = 3$ colors, these indices go from 1 to 3. g_s is the coupling constant of the theory, which is associated to a coupling strength $\alpha_s = g_s^2/4\pi$, while the structure constants are f_{ABC}, with indices A,B and C running from 1 to $N_C^2 - 1 = 8$. Finally, the generators of the gauge group, T_A, can be represented as 3×3 matrices, called Gell-Mann matrices. The Lagrangian is symmetric under $SU(3)$ gauge transformations.
As with QED, the coupling strength of the strong interaction varies depending on the energy Q of the process considered, as

$$\alpha_s(Q^2) = \frac{\alpha_s(Q_0^2)}{1 + \beta\alpha_s(Q_0^2)\ln(\frac{Q^2}{Q_0^2})} \tag{1.19}$$

where Q_0^2 is a reference scale and $\beta = \frac{11N_C - 2N_f}{12\pi}$. So, for $N_C = 3$ and the number of quarks flavors $N_f = 6$, $\beta > 0$, which means that unlike QED (where $\beta < 0$), the strong interaction runs to weaker values at higher energies. Therefore, at higher energies, the value of α_s is sufficiently small (*e.g.* $\alpha_s(M_Z) = 0.1184 \pm 0.0007$) [5] to allow the use of perturbative approaches. In this energy regime, perturbative QCD (pQCD) is used and quarks can be considered free particles, as the coupling to each other is small [16, 17]. This feature of QCD is called "asymptotic freedom".

1.4 Hadron collider physics

In order to produce new particles and study SM processes (or even possible extensions of SM), a great amount of energy is needed in a very well-defined region: such energy, which is afterward used to create particles, can be produced at colliders. Depending on the particles that collide, there are lepton and hadron colliders. Hadron colliders use bound colourless states of quarks that are accelerated up to a defined energy before reaching collisions, where the energy accumulated by the hadrons is then used to produce new particles. Together with the so-called "hard process", the collisions result in additional soft-QCD processes, a typical signature of hadron colliders that differs them from lepton colliders, which have a typically cleaner environment. Hadron colliders typically collide protons (or protons with anti-protons). Protons are made of three valence quarks, *uud*, but due to irradiation and exchange of gluons also virtual "sea quarks" take part in defining the proton structure. The constituents of the proton are generally called partons. Given the complex structure of the proton, when two protons collide only one parton from one proton collides with one parton from the other proton: this allows for $q\bar{q}$ interactions but also quark-gluon interactions, the latter more likely to happen at higher energies.

A typical process at hadron colliders is a "$2 \to 2$" process, where two initial state particles interact to produce two final state particles. The initial state can be quarks or gluons, while the final state

depends on the specific process studied. For example, let's consider the process $q\bar{q} \to Wg$, where a quark and an anti-quark interact to produce a W boson and a gluon g, which will be the main topic of Chapter 4: the "partonic" cross-section is defined as [18]

$$\sigma(q\bar{q} \to Wg) = \frac{1}{64\pi^2 s}\frac{p_f}{p_i}\int |\mathcal{M}|^2 d\Omega \tag{1.20}$$

where $\sqrt{s}$ is the center-of-mass energy, p_i and p_f are the momenta of the initial and final state particles, and $d\Omega$ is the element of solid angle. $|\mathcal{M}|$ is the matrix element, which encodes the physics of the process. The matrix element $|\mathcal{M}|$ is computed using "Feynman rules", which relate a pictorial representation of the physics process (the so-called "Feynman diagrams") to the actual computation of the matrix element. The Feynman diagram for the $q\bar{q} \to Wg$ process is shown in Fig. 1.1.

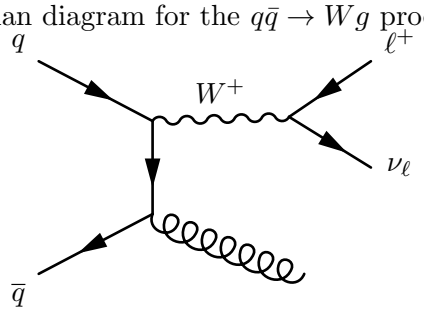


Figure 1.1: Feynman diagram for $q\bar{q} \to Wg$.

To get a perfect SM prediction of a given physics process, in principle all Feynman diagrams should be summed. Given that Feynman diagrams can be drawn with infinite complexity (*i.e.* with more and more interaction vertices), this task is almost impossible. Fortunately, at hadron colliders, the collision energies are so high that the QCD coupling strength α_s is very small, following Eq. 1.19: therefore, more complex diagrams contribute less to the total amplitude of the process. At this stage, perturbative methods can be used to make predictions, and the accuracy of the perturbative method is given by the considered order of the coupling strength: at Leading Order (LO) only one power of α_s is considered, while at Next-to-Leading order (NLO) diagrams are computed considering two powers of α_s. For typical EW processes such $W+$ jets and WW, NLO predictions are considered, as it will be presented in Chapter 4 and 5.

1.4.1 Parton Distribution Functions

The partonic cross-sections are computed assuming parton-parton interactions, but in order to model the hadron-hadron interaction, one has to take into account the probability for a specific parton of one hadron to interact with another parton coming from the other hadron. At the high energy typical of hadron colliders (at the order of the TeV), the factorization theorem [19] and QCD asymptotic freedom allow computing cross-sections involving hadrons. Basically, the partonic cross-section is weighted by the probability that protons contain initial-state partons of the required energy. This "factorization" between the partonic cross-section and the hadron/parton contribution is possible because the interactions between different partons in the same proton are small, compared to the interactions between partons from different protons. Considering the $q\bar{q} \to Wg$ example described above, the factorisation theorem can be expressed as [19]:

$$\sigma_{pp\to Wg} = \sum_{q,\bar{q}} \int dx_q dx_{\bar{q}} f_{q/p}(x_q, Q^2) f_{\bar{q}/p}(x_q, Q^2) \sigma_{q\bar{q}\to Wg}(Q^2), \tag{1.21}$$

where $\sigma_{q\bar{q}\to Wg}$ is the partonic cross-section and $\sigma_{pp\to Wg}$ is the proton-proton cross-section. The terms $f_{a/p}(x_a, Q^2)$ for $a = q, \bar{q}$ are called Parton Distribution Functions (PDFs), which represent the

probability for a parton a to carry a fraction of the proton momentum x_a [20, 21], as a function of the transferred momentum squared Q^2 during the interaction. Therefore, PDFs describe the probability for a given initial state to be available for the collision. While the partonic cross-section can be computed from the matrix elements of a given process, up to a given perturbative order, PDFs have to be determined by fitting experimental data coming from different experiments and evolving the PDFs to the relevant scale of the interaction Q^2. An example of the difference of PDFs as a function of Q^2 for different quarks is shown in Fig. 1.2.

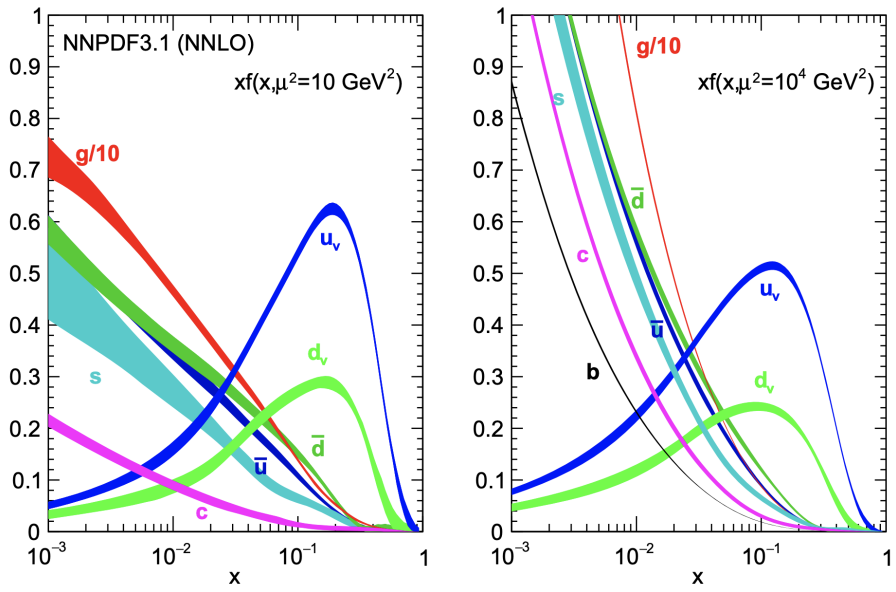


Figure 1.2: PDFs computed at next-to-next-to-leading order (NNLO) by the NNPDF group for different scales (10 and 1000 GeV^2), as a function of proton momentum fraction x [22].

PDFs can be calculated directly from lattice QCD. However, this approach currently yields large uncertainties. Additionally, the process of evolving the PDFs to the relevant scale of the interaction involves solving the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [23–25], which can be computationally intensive. Therefore, to obtain accurate PDFs, it is necessary to fit the experimental data over a wide range of x values and over a wide range of scales Q^2. This can be challenging, as the data used to fit the PDFs may come from different experiments with different systematic uncertainties and biases. Particularly, for small values of x, PDFs are not well-constrained. In this aspect, LHCb is the only experiment that can play a key role in understanding and potentially constraining PDFs in this region. A precise measurement of relevant physics processes, such as W boson production in association with jets described in Chapter 4, will help in constraining PDFs in this (x, Q^2) range.

1.5 Vector bosons + jets production

At hadron colliders, W and Z boson may be produced in association with jets. A typical process for $W+$ jets production is:

$$pp \to W^{\pm} + \text{jets} \to l^{\pm}\nu_l + \text{jets} \tag{1.22}$$

where here the semi-leptonic decay of the $W^{\pm}$ boson is considered. The interaction producing $W^{\pm}$+jets can be represented by several Feynman diagrams; typical LO Feynman diagrams are shown in Fig. 1.3.

The sum of all possible diagrams for $W^{\pm}+ \geq$N jets gives the inclusive cross-section for N jets. The measurements of processes such as $W^{\pm}$+N jets are important tests of pQCD: in high-energy particle

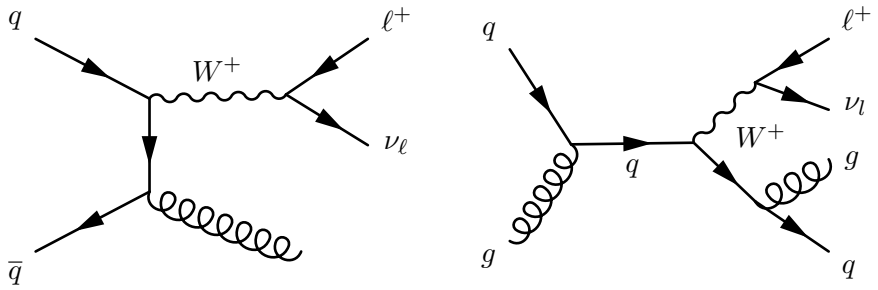


Figure 1.3: Typical LO Feynman diagrams for W+jets production: $W+$ 1 jet (left) and $W+$ 2 jet (right).

collisions, such as those involving the production of $W^{\pm}$ bosons, the strong nuclear force exhibits a smaller coupling constant α_s, which allows for the use of pQCD to accurately model the interaction. The inclusion of more partons in the process allows the testing of higher orders of pQCD. Finally, $W^{\pm}$+N jets production is of significant interest in particle physics due to its role as a background process in searches for phenomena beyond the standard model and as a multi-scale QCD process. In the context of searches for new phenomena, $W^{\pm}$+N jets production is a major source of background in various channels involving supersymmetry and the decay of Higgs bosons. An interesting example is the search for the decay of a Higgs boson into a pair of c-quarks: despite the advanced techniques to properly identify jets produced by different quarks, $W+$ jets, and $Z+$ jets are dominant background sources. In addition, in processes with small cross-sections within the SM, such as single top quark production, WW production, and vector boson fusion, the $W^{\pm}$+N jets process can dominate over the small signal. This will be evident in Chapter 5, where indeed $W+$ jets is the most dominant background when measuring WW. The theoretical uncertainties in the production rates and kinematics of $W^{\pm}$+N jets processes are significant when going to a higher number of partons, which limits our ability to accurately identify and study new phenomena. Finally, $W+$ jets is an excellent candidate to constrain PDFs: the high p_{T} muon coming from the W boson decay can be effectively used to trigger events and to provide clean signatures with relatively low background, therefore allowing a precise study at higher orders of pQCD.

1.6 Double W vector bosons production

At hadron colliders, W bosons can also be produced through vector boson pair production. WW pairs can be produced through quark-quark annihilation (already at LO), and gluon-gluon fusion (at NLO). Typical Feynman diagrams for WW production are shown in Fig. 1.4.

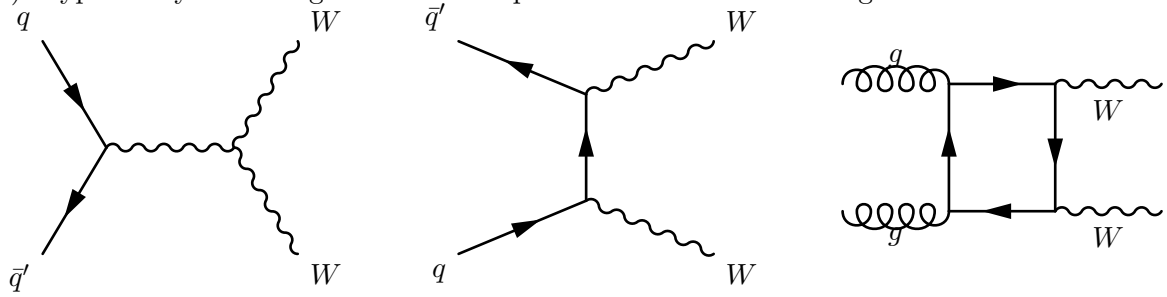


Figure 1.4: Feynman diagrams for WW production: quark-quark annihilation s-channel (left) and t-channel (center), gluon-gluon fusion (right).

WW production is an important process to test SM predictions. Assuming that there is physics beyond the SM, that it does not contain new low-mass particles, and that is consistent with the symmetries of the SM, its effects can be parametrized in terms of an Effective Field Theory (EFT). In

this framework, the new physics theory is obtained by expanding around the SM and integrating over degrees of freedom at higher energies. The main result of this procedure leads to additional terms in the Lagrangian, proportional to inverse powers of the mass scale of the new particles, up to numerical factors that depend on the new couplings. The new physics theory may contribute to the production of WW by acting on triple gauge-boson couplings (TGCs) such as $WW\gamma$ and WWZ. Indeed, these couplings already appear at LO in the s-channel, as shown in Fig. 1.4, and they could possibly deviate from their predicted SM values since up to date they are not yet highly constrained by measurements. In the EFT context, TGCs are parametrized by an effective Lagrangian that can be written in this form:

$$\mathcal{L}_{\text{eff}}^{WWX} = -ig_{WWX}\left[g_1^X X^\mu (W^-_{\mu\nu}W^{\nu+} - W^+_{\mu\nu}W^{\nu-}) - \kappa_X W^+_\mu W^-_\nu X^{\mu\nu} + \frac{\lambda_X}{M_W^2} X^{\mu\nu} W^{\rho+}_\nu W^-_{\rho\mu}\right] \quad (1.23)$$

where X^μ stands for the photon or the Z boson and new coupling parameters, g_1^X, κ_X and λ_X are introduced. In the SM, $g_1^X = \kappa_X = 1$ and $\lambda_X = 0$. These additional contributions induce TGCs beyond those present in the SM, and are referred to as anomalous TGCs (aTGCs). Experimentally, aTGCs would manifest themselves as an increased WW cross-section, especially as an increased yield of events at high transverse momentum of the W bosons and at a high invariant mass of the WW system. Therefore, a precise measurement of WW may rule out and new-physics effects. Finally, the pair production of WW bosons is one of the most important background processes for the study of one of the main Higgs boson decays ($H \to WW$). In Chapter 5, a study of WW production at the LHCb experiment is presented.

1.7 Experimental results on vector bosons

In this section, a limited review of relevant experimental results for $W+$ jets and WW production is presented, ranging from various experiments and analyses. Experimental results are compared to theoretical predictions using Monte Carlo (MC) simulations. MC simulations are obtained using event generators (such as MadGraph [26, 27] or Powheg [28]) interfaced with parton shower programs (typically Pythia8 [29] and Herwig [30]). Depending on the event generator used, theoretical predictions can be computed at LO or at NLO accuracy.

1.7.1 Results by CDF and D0

The CDF [31] experiment at the Tevatron proton anti-proton collider at Fermilab, performed a measurement of $W+$ jets cross-section at center-of-mass energy $\sqrt{s} = 1.96$ TeV, with an integrated luminosity of 320 ± 18 pb^{-1} [32]. The electrons coming from the W boson decay are required to have transverse energy $E_{\text{T}} > 20$ GeV and each event must have missing transverse energy $\not{E}_{\text{T}} > 30$ GeV, where $\not{E}_{\text{T}}$ is defined as the momentum imbalance in the plane perpendicular to the beam direction. The transverse mass of the W boson $m_{\text{T}}^W = \sqrt{2p_{\text{T}}^e p_{\text{T}}^\nu (1 - \cos\phi_{e\nu})}$ is required to be greater than 20 GeV, where p_{T}^e (p_{T}^ν) is the transverse momentum of the electron (neutrino) and $\phi_{e\nu}$ is the azimuthal angle between the electron and the neutrino. These requirements are necessary to remove background events that are unlikely to be true $W+$jets events but result instead from QCD or $t\bar{t}$ processes. Events are grouped according to the number of jets with transverse energy $E_{\text{T}} > 15$ GeV. The jet counting is done inclusively such that events with n jets are included in both the n-jet and the $n+1$-jet category. The $\not{E}_{\text{T}}$ distribution, distinct for events with W bosons, is used to extract the signal from the remaining background events. Figure 1.5 shows the results of a fit to $\not{E}_{\text{T}}$ for the $W+1$ jet category. The background fraction varies from 10% to 40% between 1-jet and 4-jet event categories, and up to 90% for the high jet E_{T} range.

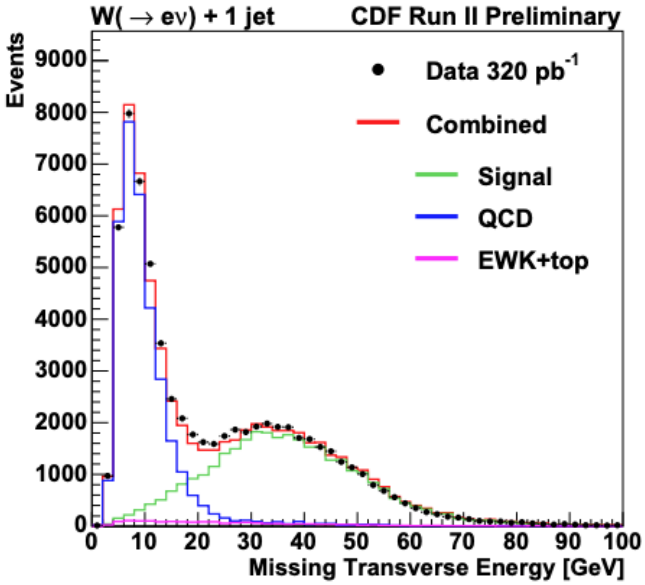


Figure 1.5: Fit to $\not\!\!E_{\mathrm{T}}$ for the $W+$ 1 jet category. Signal, QCD, and EWK+top contribution are shown, together with the combined result [32].

D0 [33], the second experiment at the Tevatron collider, performed a measurement of $W^{\pm}+$ jets production using data at center-of-mass energy $\sqrt{s} = 1.96$ TeV [34]. The D0 analysis for $W^{\pm}+$jets is based on an integrated luminosity of 4.2 fb^{-1}, and the electronic decay of the $W^{\pm}$ bosons is considered. Events are required to have an electron with $p_{\mathrm{T}} > 15$ GeV, missing transverse energy $\not\!\!E_{\mathrm{T}} > 20$ GeV, and transverse mass $M_{\mathrm{T}} > 40$ GeV. Jets are required to have minimum transverse momentum $p_{\mathrm{T}} > 20$ GeV. Differential cross-sections, normalized to the total $W^{\pm}$ boson cross-section and fully corrected for detector effects, are computed for several inclusive jet multiplicities and as a function of various kinematic quantities. Experimental data are compared with predictions from a variety of theoretical approaches. Results are shown in Fig. 1.6: the differential cross-section, the ratio between theory and data, and the ratio between the n and $n+1$ jet multiplicity cross-sections are shown as a function of the jet multiplicity, while the relative differential cross-section is shown as a function of the jet p_{T}. Over most of the phase space in which the measurements are presented, experimental uncertainties are smaller than the theoretical uncertainties at NLO. Results agree with SM predictions.

Both CDF [35] and D0 have measured WW cross-section in the all leptonic final state $l\nu l'\nu$, where $l = e, \mu$. Despite having the lowest branching ratio for WW decay ($\sim 4.6\%$), it has the cleanest signature, given that all other decays involve hadronic jets. Typical values of $\not\!\!E_{\mathrm{T}} > 25$ GeV are chosen for these analyses. Major backgrounds to the $ll'\not\!\!E_{\mathrm{T}}$ final state come from $W+$ jet or photon, where the jet or the photon are misidentified as a lepton, Drell-Yan (Z/γ^*) production with high $\not\!\!E_{\mathrm{T}}$ due to detector effects, $t\bar{t}$ and other dibosons processes such as WZ and ZZ. Both experiments have measured cross-sections consistent with the NLO calculation of the SM expectations.

CDF [36] and D0 have also probed the TGCs measuring WW and WZ events. CDF considered $l\nu jj$ final states, which have a greater background coming from $W+$ jets processes but increase the branching ratio of WW (WZ) by a factor 6.5 (10). Moreover, the $l\nu jj$ final can be sensitive to anomalous TGCs in the high W transverse momentum region, where the $W+$ jets contribution is rather low. D0 instead used the $ll'\nu\nu$ final state, therefore a cleaner final state but with a lower branching ratio. Results are shown in Fig. 1.7, where D0 performed a fit to the leading lepton p_{T} while CDF performed a fit to the p_{T} spectrum of the leptonically decaying W boson. Results are compatible with SM predictions and no evidence of anomalous TGCs has been found.

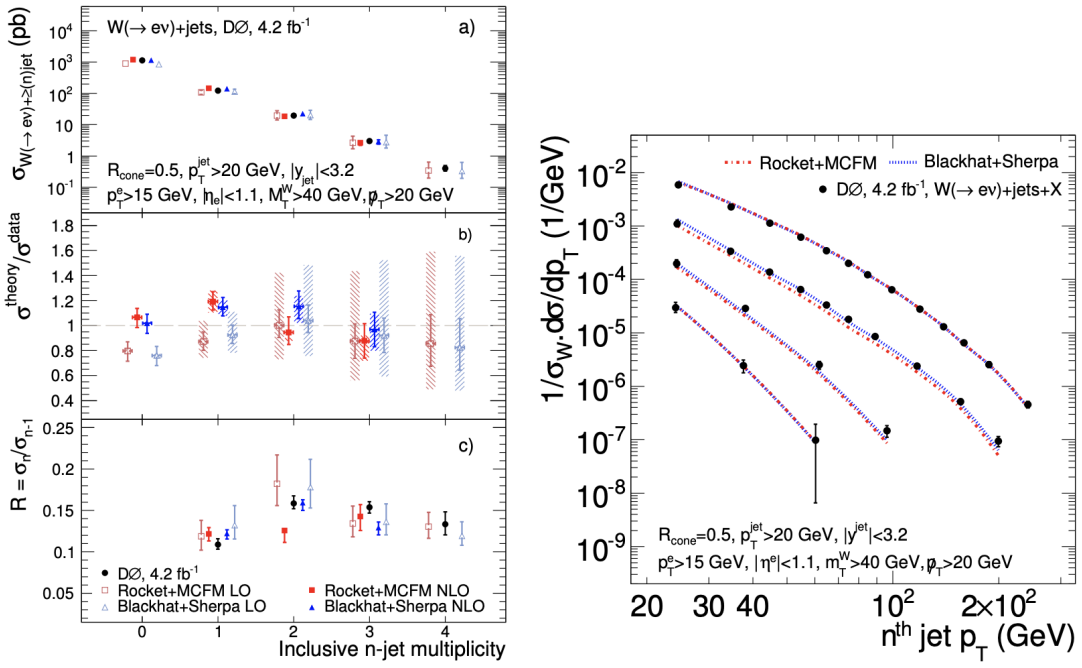


Figure 1.6: (left) Differential cross-section σ, ratio between theory and data $\sigma^{\text{theory}}/\sigma^{\text{data}}$, and the ratio between the n and $n+1$ jet multiplicity cross-sections $R = \sigma_n/\sigma_{n+1}$ are shown as a function of the jet multiplicity. (right) Relative differential cross-section $1/\sigma_W \cdot d\sigma/dp_{\text{T}}$ is shown as a function of the jet p_{T} [34].

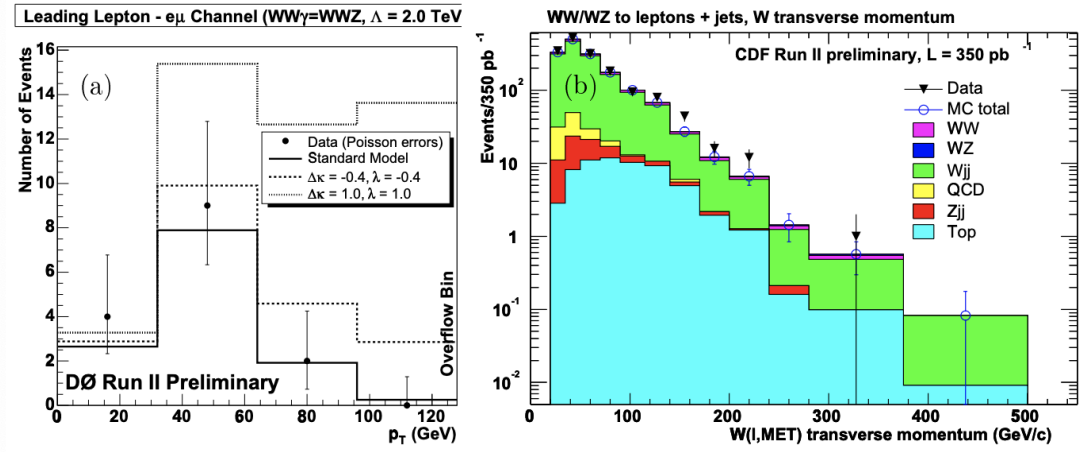


Figure 1.7: (left plot) Fit to the leading lepton p_{T} performed at the D0 experiment. (right plot) Fit to the p_{T} spectrum of the leptonically decaying W boson performed at the CDF experiment [34].

1.7.2 Results by ATLAS and CMS

ATLAS [37] and CMS [38] are two of the four experiments at the LHC proton-proton collider at CERN, defined in Section 2.1. At ATLAS, $W^{\pm}$+ jets production has been measured requiring an electron or muon with $p_{\text{T}} > 25$ GeV in the detector fiducial region [39]. To reconstruct the neutrino, $\not{E}_{\text{T}}$ is required to be greater than 30 GeV. Jets are reconstructed using the anti-k_{T} algorithm, with the distance parameter $R = 0.4$. The dominant backgrounds come from multijet production giving rise to a fake lepton signature in the detector, and, at higher jet multiplicities, top pair production. These backgrounds are estimated using data-driven techniques, and smaller background contributions (single-top, Z boson, and diboson production) are estimated using Monte Carlo (MC) simulations. The

background-subtracted data are corrected to the particle level using a Bayesian unfolding technique, and particle-level fiducial cross-sections are presented. The large sample is used to study numerous kinematic variables, including multiplicities up to 7 jets, and jet transverse momenta up to 1 TeV. The $W^{\pm}$ cross-section is measured differentially as a function of several kinematic quantities, such as the leading jet p_{T} and the scalar sum of transverse energies of all particles in the fiducial region in the event H_{T}. Figure 1.8 shows the results as a function of the leading jet p_{T}: data are compared with several event generators, showing good agreement within the experimental uncertainties.

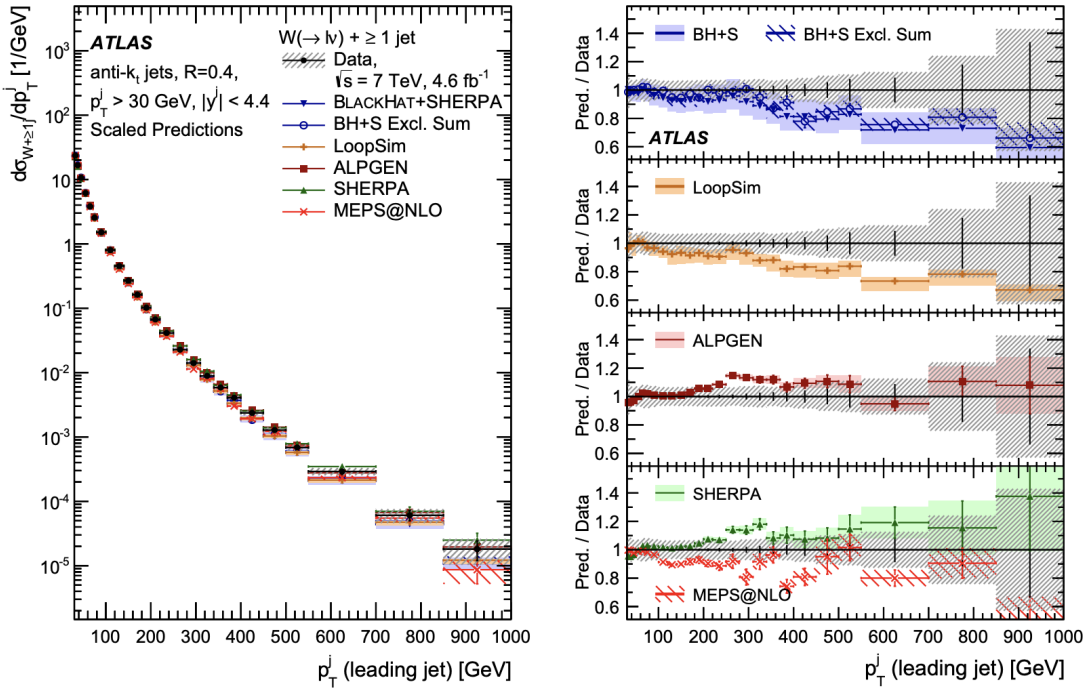


Figure 1.8: (left) Differential $W+$ jets cross-section measurement performed by the ATLAS experiment as a function of the leading jet p_{T}. (right) Comparison between data and several MC generators as a function of leading jet p_{T} [39].

WW production cross-section has been measured by ATLAS at a center-of-mass energy $\sqrt{s} = 13$ TeV using 3.15 fb^{-1} [40]. All leptonic final state have been considered, and WW event candidates are selected by requiring exactly one electron and one muon of opposite charge in the event, and significant missing transverse momentum. The selection of opposite sign leptons leaves background contributions from $t\bar{t}$ and single top with no jet vetoed, Drell-Yan production of τ lepton pairs, multi-jet production, and $W+$jets with jets misidentified as leptons. The measured fiducial cross-section is 529 ± 20 (stat.)$\pm$50 (syst.) $\pm$11 (lumi.) fb, and is found to be consistent with the most up-to-date SM predictions that include high-order QCD effects at NLO.

At CMS, the differential cross-sections for a $W^{\pm}$ boson produced in association with jets are measured from a data sample corresponding to an integrated luminosity of 2.5 fb^{-1} at a center-of-mass energy $\sqrt{s} = 13$ TeV [41, 42]. The analyzed $W+$ jets events are required to have a muon with $p_{\mathrm{T}} > 25$ GeV within the detector acceptance, and jets with $p_{\mathrm{T}} > 30$ GeV. The events are further required to be in the W boson transverse mass peak region, defined by $m_{\mathrm{T}}^{W} > 50$ GeV. The measured differential distributions are compared with predictions obtained with several event generators. The differential cross-section results of the measurement based on 13 TeV data are shown in Figure 1.9 as functions of the leading jet p_{T} and the H_{T} variable, defined as the scalar sum of the jets p_{T}, for one jet inclusive production. The MadGraph prediction shows remarkable agreement with data over the entire ranges of the leading jet p_{T} and the H_{T} distributions. The MadGraph prediction at LO

slightly underestimates data, generally at lower-p_{T} and -H_{T} regions.

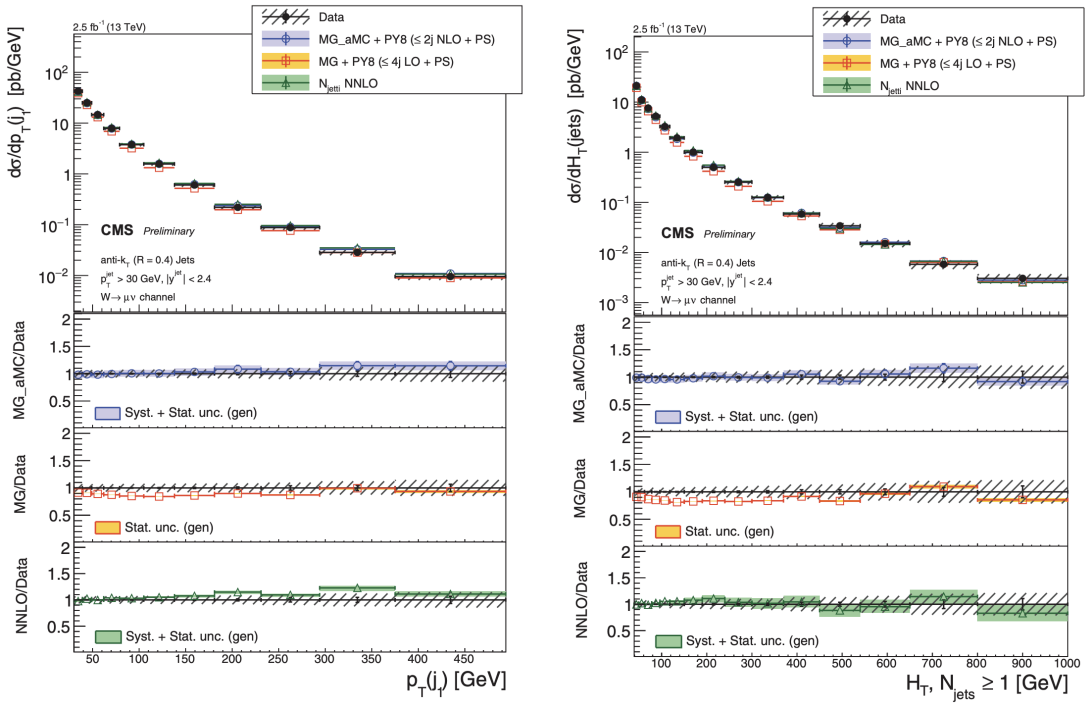


Figure 1.9: (left) $W+$ jets differential cross-section performed by the CMS experiment as a function of the leading jet p_{T}. (right) $W+$ jets differential cross-section as a function of H_{T} [41].

As ATLAS, also CMS have performed WW cross-sections measurements at a center-of-mass energy $\sqrt{s} = 13$ TeV, using 35.9 fb^{-1} of data [43]. The same full leptonic channel of the ATLAS analysis is used, with an electron and a muon of opposite charge in the final state with $p_{\mathrm{T}} > 25$ GeV for the leading lepton. Two methods for reducing background contributions are employed: a sequential cut approach on kinematic and topological quantities, and a multivariate approach using two random forest classifiers, one optimized to distinguish WW events from Drell-Yan, and the other one to separate WW from $t\bar{t}$ events. For both approaches, results are consistent with the SM expectations.

1.7.3 Results by LHCb

Despite its limited detector angular acceptance and lower integrated luminosity, at the LHCb experiment is possible to measure EW processes involving vector bosons [44–48]. The angular coverage is complementary to ATLAS and CMS, therefore at the LHC experiments, it is possible to test the SM predictions in the full phase space of proton-proton collisions.

A measurement of the differential cross-section of W+jets and Z+jets has been obtained by LHCb using data taken at $\sqrt{s} = 8$ TeV [49], by requiring a high p_{T} muon and at least one high p_{T} jet in the detector acceptance. Given that LHCb is not a hermetic experiment, this measurement cannot rely on measuring the missing energy due to the neutrino coming from the W boson decay. Instead, the muon isolation I_μ is used as a proxy for EW events. I_μ is defined as

$$I_\mu = \frac{p_{\mathrm{T}}^{\mu}}{p_{\mathrm{T}}^{\mu-\mathrm{jet}}} \tag{1.24}$$

where p_{T}^{μ} is the transverse momentum of the muon and $p_{\mathrm{T}}^{\mu-\mathrm{jet}}$ is the transverse momentum of a "jet" surrounding the muon with a given radius. I_μ is directly related to the number of particles surrounding

the muon: muons coming from the decay of a vector boson are isolated, and have an isolation value closer to one; whereas muons coming from background events (e.g. di-jets events) are not isolated. This behavior is evident in Fig. 1.10.

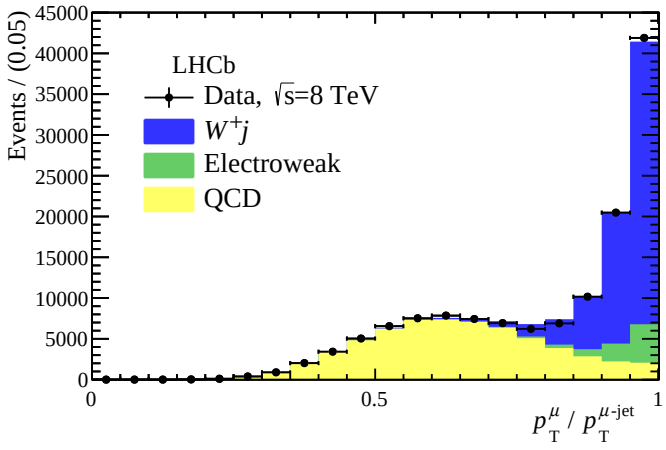


Figure 1.10: Muon isolation I_μ for W^++jet events. Dots represent LHCb data with superimposed distributions for signal, QCD events, and other Electroweak processes [49].

A template fit to I_μ is used to extract the number of $W+$ jets events, from which a measurement of the cross-section, both in the full fiducial region and differentially as a function of several kinematic variables, is calculated. cross-section ratios R_{WZ}, R_{W^+Z}, R_{W^-Z} between $W+$ jets and $Z+$ jets events, and $R_\pm$ between W^++ jets and W^-+ jets events are obtained. For W^++ jets and W^-+ jets events the cross-section asymmetry $A(Wj)$, defined as $A(Wj) = (\sigma(W^+j) - \sigma(W^-j)/(\sigma(W^+j) + \sigma(W^-j))$, is computed. Results for cross-section measurement, cross-section ratios, and cross-section asymmetry $A(Wj)$ obtained from the analysis are shown in Fig. 1.11: experimental results are compared with MadGraph and Powheg, showing good agreement between data and theoretical predictions.

At the moment of writing this thesis, no WW cross-section measurement has been performed at the LHCb: in the forward region, the WW cross-section is very low. Also, LHCb lacks missing momentum information used to identify WW events with semileptonic decay by ATLAS, CMS, and $p\bar{p}$ experiments. A feasibility study for measuring WW cross-section has been performed analyzing data corresponding to a total integrated luminosity of 5.4 fb^{-1} [50]. The full semileptonic channel is reconstructed, requiring one W boson to decay into an electron and a neutrino, and the other W boson to decay into a muon and a neutrino. The electron and the muon are required to have opposite electric charges, which helps in suppressing the contamination from $Z \to \mu\mu$ and $Z \to ee$ decays. The main backgrounds for this analysis are:

- $t\bar{t}$, where the electron and the muon come from the decays of the two W bosons, together with two b-jets;
- $Z \to \tau\tau$, where the τ leptons decay generating a $e\mu$ pair;
- $Z \to \mu\mu$ and $Z \to ee$, where one electron (muon) is misidentified as a muon (electron);
- $W(\to \mu\nu)+$ jets, where the jet is misidentified as an electron.

Both the signal and the backgrounds are modeled using MC simulations: the MC samples are produced at LO using Pythia8 [29] and the CT09MCS PDF set [51]. At the analysis level, the electron and the muon are required to have $p_T > 20$ GeV and be in the fiducial detector region. To further reduce the various sources of backgrounds, several requirements are considered:

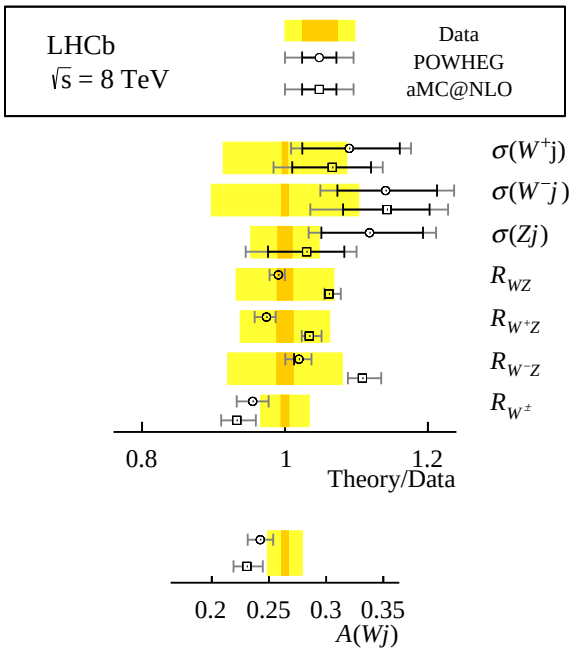


Figure 1.11: Comparisons between experimental results and theoretical predictions for $W+$ jets and $Z+$ jets cross-sections, cross-section ratios and cross-section asymmetry, in the full fiducial region [49].

- lepton isolation: the p_{T} of the jet surrounding the lepton is required to be less than 2 GeV, therefore requiring both leptons to be isolated;
- lepton impact parameter (IP): a cut on IP, defined as the distance of closest approach of the track to the collision point, is considered to reduce the background coming from $Z \to \tau\tau$ events since τ leptons have a higher IP with respect to prompt electrons and muons;
- jet multiplicity: WW events are less likely to include jets, therefore a cut on the number of jets is considered to reduce the contribution from $t\bar{t}$ events;

In Fig. 1.12, the $e\mu$ invariant mass is shown, together with the expected signal and background processes.

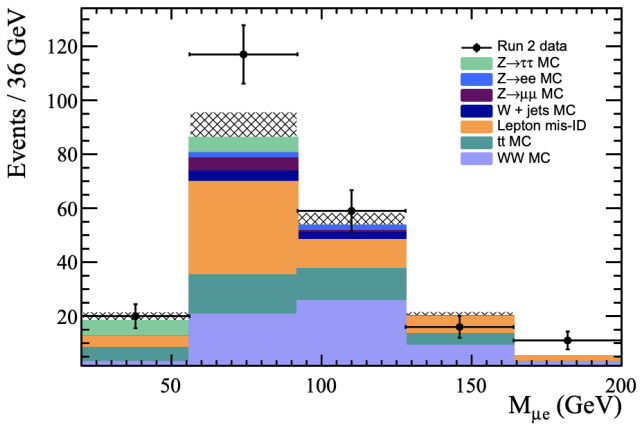


Figure 1.12: The $e\mu$ invariant mass $M_{\mu e}$ for WW and background events [50].

The feasibility of this measurement is assessed by measuring the significance $\mathcal{S}$, defined as

$$\mathcal{S} = \frac{N_{\mathrm{WW}}}{\sqrt{N_{\mathrm{WW}} + N_{\mathrm{background}}}} \tag{1.25}$$

where N_{signal} ($N_{\mathrm{background}}$) is the number of WW (background) events. In order to take into account also the precision of the model used to estimate the WW and the background contributions, an expanded definition of $\mathcal{S}$ is used, defined as

$$\mathcal{S}' = \frac{N_{\mathrm{WW}}}{\sqrt{N_{\mathrm{WW}} + N_{\mathrm{background}} + \sigma^2_{MC} + \sigma^2_{DD}}} \tag{1.26}$$

where σ^2_{MC} and σ^2_{DD} are respectively the statistical uncertainties on the simulated and data-driven portions of the background model. With this new definition, a value of $\mathcal{S}' = 2.8\sigma$, showing that a WW cross-section measurement using the $e\mu$ channel will be possible in the next LHCb data-taking campaign using ~ 16 fb^{-1}.

1.8 Analyses motivations

In this thesis, two analyses are presented:

- in Chapter 4, a measurement of the $W+$ jets process is performed using data collected at the LHCb experiment at center-of-mass energy $\sqrt{s} = 13$ TeV. The W boson is required to decay into a muon and a neutrino, and the measurement is performed both in the full fiducial region and differentially as a function of the leading jet p_{T}, leading jet and muon pseudorapidity η (defined in Section 2.2), and requiring at least one or two jets inside the LHCb acceptance. The analysis follows the approach of the previous analysis [49]: a fit to the muon isolation variable is performed to separate EW events from QCD, this time extending the measurement to the two jets case.
- in Chapter 5, a study of WW process is performed using the same Run 2 dataset as for $W+$ jets measurement, requiring one W boson to decay into a muon and a neutrino and the other W boson to decay hadronically into two jets. The signature for this analysis is a high p_{T} muon together with two high p_{T} jets. The higher branching ratio with respect to the full semileptonic decay is unfortunately limited by the higher backgrounds, particularly $W+$ jets. Therefore, identification techniques based on Machine Learning (ML) methods are required to increase the discrimination between WW and $W+$ jets events.

There are several motivations to perform these measurements:

- as explained in Sec.1.5, $W+$ jets is an excellent channel to study PDFs and pQCD. Particularly, performing this measurement at the LHCb experiment allows covering the full phase space of the proton-proton collisions, providing complementary results with respect to ATLAS and CMS;
- as already described in Sec. 1.7.3, WW process has never been measured in the forward region of proton-proton interactions. This study, even with low events expected, is paving the way to the measurement when more data will be available to test SM in the forward unexplored region.

Chapter 2

The LHCb experiment: description and performance

The analyses presented in this thesis introduced in Chapter 1 rely on data taken at the LHCb experiment during the Run 2 data-taking campaign (2016-2018). To properly understand how the final state objects, *i.e.* muons and jets, are reconstructed for the analyses, a description of the LHCb detector is necessary, particularly focusing on the sub-detectors involved in the muons and jets reconstruction. Then in Chapter 3, the reconstruction algorithm will be described. The structure of this Chapter is the following:

- Section 2.1 gives a brief description of the Large Hadron Collider (LHC) and its main experiments;
- Section 2.2 describes the LHCb experiment, its sub-detectors and their performance during Run 2, including also the triggering procedure and the typical data analysis flow used also in the analyses presented in this thesis. The description of the upgrades of the LHCb experiment is developed in Chapter 6.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [52] is a 24.6 km long circular collider, based at 100 m underground at CERN. It is made of two coaxial rings where two counter-rotating proton beams circulate. Beams run inside two beam pipes where hard vacuum has been obtained, and the two pipes intersect in four interaction points, where the main experiments are located. Protons are extracted from hydrogen ionization, and in order to reach the required center of mass energy, the acceleration of the beams is divided into several stages: the first acceleration stage is done by a Linear Accelerator (Linac), where protons are accelerated up to an energy of 50 MeV; then they are accelerated by the Proton Synchrotron Booster (PSB) up to 1.4 GeV, by the Proton Synchrotron (PS) up to 28 GeV, and by the Super Proton Synchrotron (SPS) up to 450 GeV, before entering in the LHC stage. A schematic view of the LHC complex is given in Figure 2.1. Inside the LHC, the beams are accelerated by means of 16 radio-frequency (RF) cavities working in a superconducting regime, reaching energy up to ~ 7 TeV, reaching a center of mass energy $\sqrt{s} \sim 14$ TeV.
Proton beams are not continuous, but they are made of 2808 bunches of $\sim 10^{11}$ protons, each bunch being 30 cm long and with a transverse dimension of 1 mm. Bunches are squeezed at the collision points reaching a transverse dimension of 16 μm, and they are spaced by ~ 25 ns each other, reaching

a nominal collision frequency of 40 MHz and a peak luminosity[1] of $\mathcal{L} = 2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$.

At the collision points, four experiments are located: ATLAS (A Toroidal LHC ApparatuS) [37] and CMS (Compact Muon Solenoid) [38] are General Purpose Detectors with a cylindrical structure, designed to study Higgs physics and physics beyond the Standard Model; ALICE (A Lead Ions Collision Experiment), dedicated to heavy ions collisions and the study of quark-gluon plasma; LHCb, which is designed to study the forward region of proton-proton collisions and the physics of b- and c- quarks.

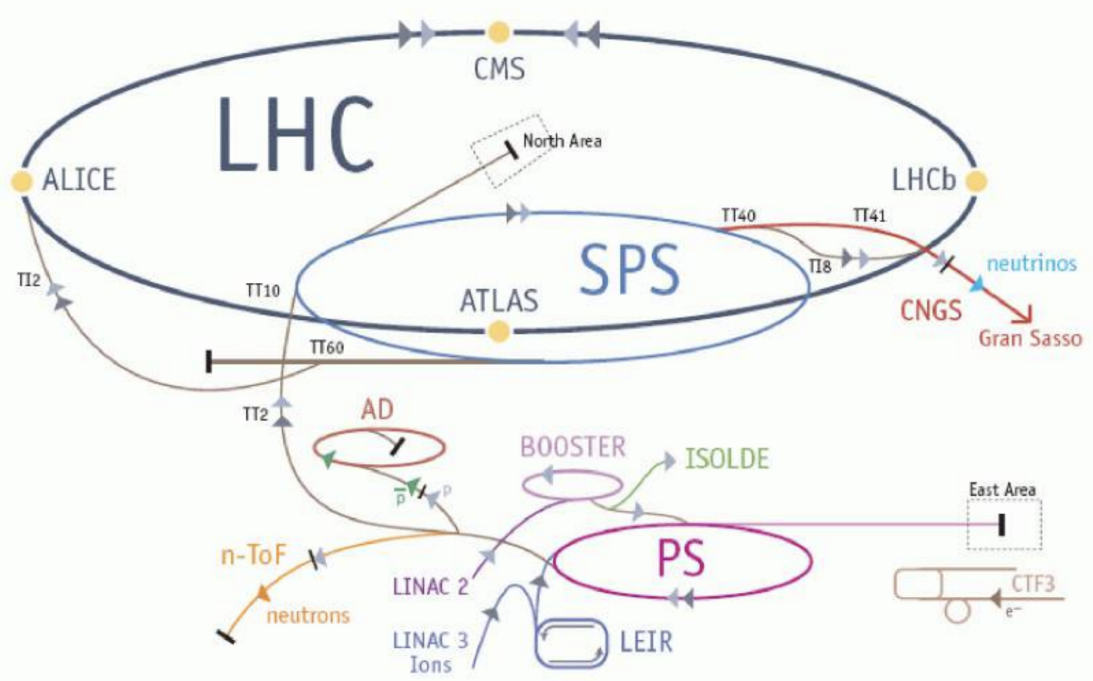


Figure 2.1: Schematic overview of CERN and the LHC complex.

2.2 The LHCb detector

The LHCb detector [53] is a single-arm spectrometer with an angular coverage in the forward region going from 10 mrad to 250 (300) mrad in the bending (non-bending) plane. The angular coverage of LHCb corresponds to a pseudo-rapidity[2] range between 1.8 and 4.9. The layout of the LHCb experiment is shown in Figure 2.2: its structure is divided into several sub-detectors, each devoted to a specific task; the LHCb coordinates system has the z-axis parallel to the beam direction, while the y- and x-axes form a right-handed system, with the y-axis opposite to the gravitational acceleration. Starting from the interaction point, these are the LHCb sub-detectors:

- the VErtex LOcator (VELO), which is the innermost sub-detector devoted to vertex reconstruction;
- the tracking system, made of four tracking stations, one located upstream (the Trigger Tracker, TT) and three downstream the magnet (T1, T2, T3);

[1]The instantaneous luminosity $\mathcal{L}$ is a measure of the number of collisions that take place in a detector per cm^2 per second, defined as

$$\mathcal{L} = f\frac{N_1 \cdot N_2}{4\pi \cdot \sigma_x \cdot \sigma_y}, \tag{2.1}$$

where f is the bunch crossing frequency, N_1 (N_2) is the number of protons in the bunches in the first (second) beam, and σ_x (σ_y) is the beam spread along the x (y) direction.

[2]For a particle, its pseudo-rapidity is defined as $\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$, where θ is the particle polar angle.

- two Ring Imaging Cherenkov (RICH) detectors, which are fundamental to achieving particle identification and separation between pions and kaons. They are placed respectively upstream and downstream of the magnet;
- a calorimeter system composed of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL), and a hadronic calorimeter (HCAL);
- a muon system, made of five stations, one upstream (M1) and four downstream (M2, M3, M4, M5) the calorimeter system.

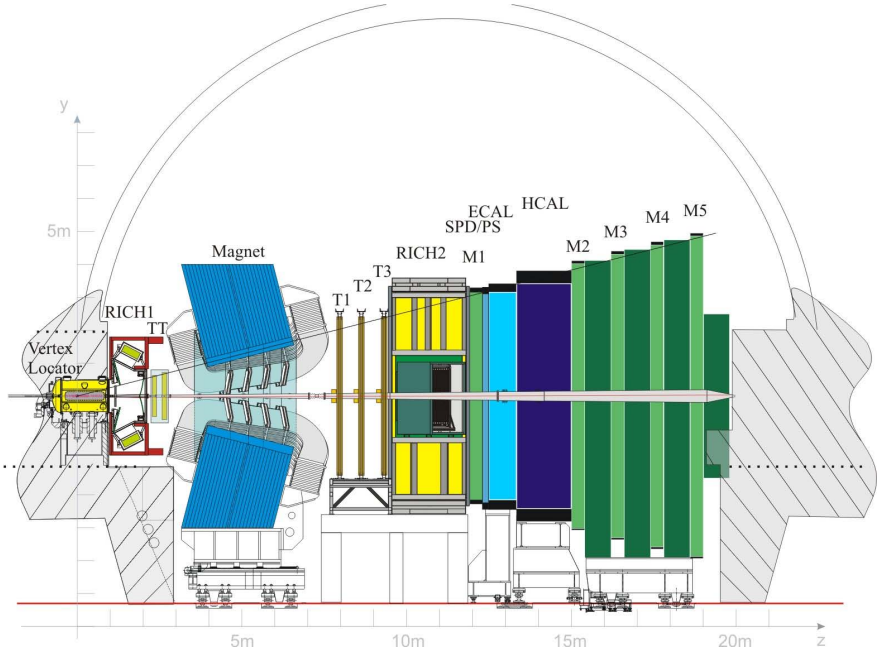


Figure 2.2: Schematic representation of the LHCb experiment, showing each sub-detector. The z-axis is parallel to the beam direction while the y-axis is parallel and opposite to the gravitational acceleration. [53]

A *luminosity leveling* technique [54] is applied to keep the luminosity delivered to the LHCb experiment at the constant value of $4.5 \times 10^{32}\text{cm}^{-2}\text{s}^{-1}$. This is achieved by slightly tilting the two proton beams increasing the interaction area. The LHCb experiment, therefore, operates at a luminosity that is lower than the nominal luminosity of other experiments such as ATLAS and CMS: this allows to have a lower pile-up (*i.e.* the average number of collisions per bunch crossing) and therefore cleaner collisions. The mean number of collisions is less than 2.5 per bunch crossing.

The LHCb center of mass energy has changed throughout the years, constantly approaching the nominal center of mass energy of 14 TeV. In particular, during the Run 2 data-taking campaign, the LHCb has operated at a center-of-mass energy $\sqrt{s} = 13$ TeV. The total integrated luminosity $\mathcal{L}_{\text{int}}$, the visible number of collisions per bunch crossing μ, and the peak luminosity $\mathcal{L}$ are reported for each year in Table 2.1.

2.2.1 Vertex Locator

The VELO [55] is the closest sub-detector to the interaction point. Its task is the reconstruction of tracks produced by charged particles coming from pp-collisions in order to precisely reconstruct the impact parameter (IP) of tracks and primary and secondary vertices, the latter fundamental to study b- and c-hadrons decays. The VELO has a cylindrical structure made of silicon modules: the passage of a charged particle creates electron-hole pairs producing electric current from which, the point of

	year	$\mathcal{L}_{\text{int}}$ (fb^{-1})	$\sqrt{s}$ (TeV)	μ	$\mathcal{L}(10^{32}\text{cm}^{-2}\text{s}^{-1})$
Run 1	2010	0.03	7	1.1	1.6
	2011	1.0	7	1.1	4.0
	2012	2.2	8	1.1	4.0
Run 2	2015	0.33	13	1.1	2.0
	2016	1.67	13	1.1	3.5
	2017	1.71	13	1.1	3.5
	2018	2.01	13	1.1	4.4

Table 2.1: Integrated recorded luminosity $\mathcal{L}_{\text{int}}$, center-of-mass energy $\sqrt{s}$, visible number of collisions per bunch crossing μ and peak luminosity $\mathcal{L}$ at LHCb for each year of the Run 1 and Run 2 campaigns. [53]

incidence of the particle is inferred. Each of the VELO modules provides a measure of the radius r (from 8 mm to 42 mm) and the angle ϕ (from 15 mrad to 390 mrad), that combined with the z coordinate of the module give the hitting position of the particle. The structure of a VELO module is shown in Figure 2.3: each module is 300 μm thick and has a semi-circular shape, extending from 42 mm to 8 mm, with an increase in strip density near the beamline as shown in Figure 2.3.

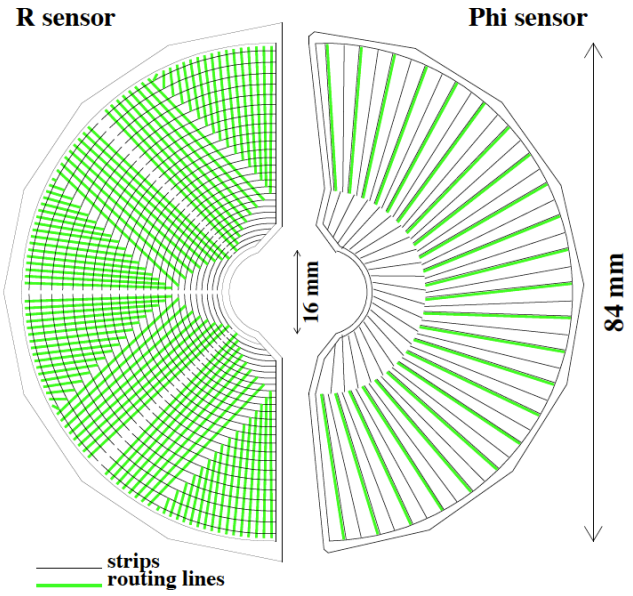


Figure 2.3: Structure of a VELO module: the two semi-circular modules composing the R and Phi sensors are represented [56].

The VELO is made of 21 pairs of modules, placed along a distance of 1 m parallel to the beam axis, as shown in Figure 2.4. The left and right array of modules are designed to slightly overlap, and they provide an angular coverage of $1.6 < \eta < 4.9$ for tracks produced in the range $\pm 2\sigma$ of the central interaction region. Given that during the injection phase, the beam is wider than 8 mm, the modules have a retractable system that allows the VELO to stay in an open configuration to a distance of 29 mm, until reaching beam stable conditions. In order to operate at the voltage required for $> 99\%$ hit efficiency, a CO_2 cooling system is used, which allows to keep a $20:1$ signal-to-noise ratio.

The best raw hit resolution for tracks is around 5 μm, depending on the width of the microstrip and the projected angle (*i.e.* the angle between the track and the perpendicular to the sensor) of the track, as shown in Figure 2.5.

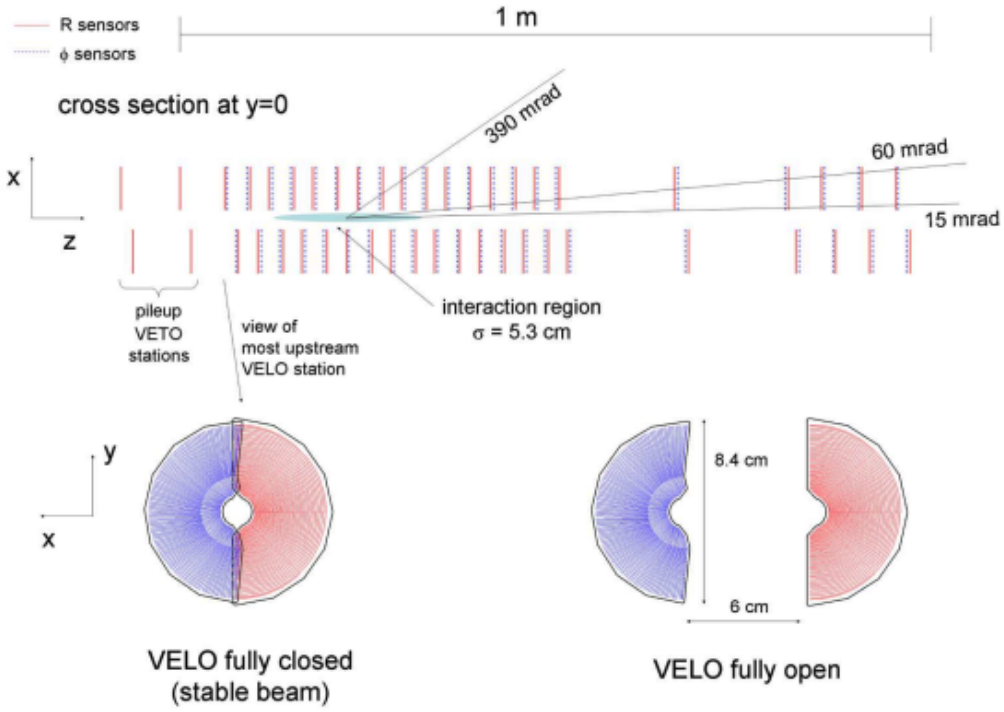


Figure 2.4: Placement of VELO modules along the z-axis (upper plot). VELO closed and open positions (lower plot) [53].

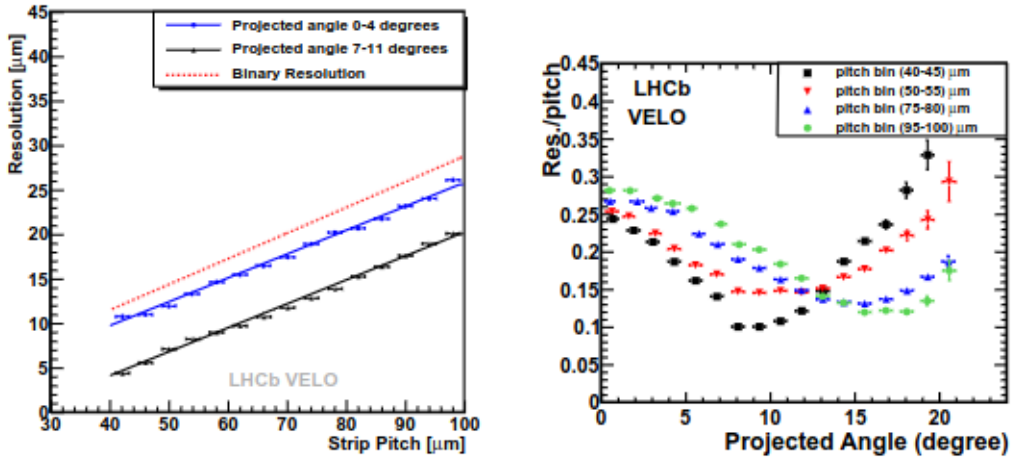


Figure 2.5: (left) Raw hit resolution as a function of the microstrip width for two projected angle bins for the R sensors. (right) Resolution divided by pitch as a function of the track projected angle for four different strip pitches [56].

2.2.2 Magnet

Between the TT and the T1 tracking stations, a room-temperature dipole magnet is located [57]. The magnet is made of two coils positioned respectively above and below the beamline. They provide a 4 Tm bending power, with an angular acceptance of ±250 (±300) mrad in the vertical (horizontal) plane. A schematic representation of the magnet is shown in Figure 2.6.

The magnetic field B_y is parallel to the y-axis, and its polarity can be inverted by reverting the current

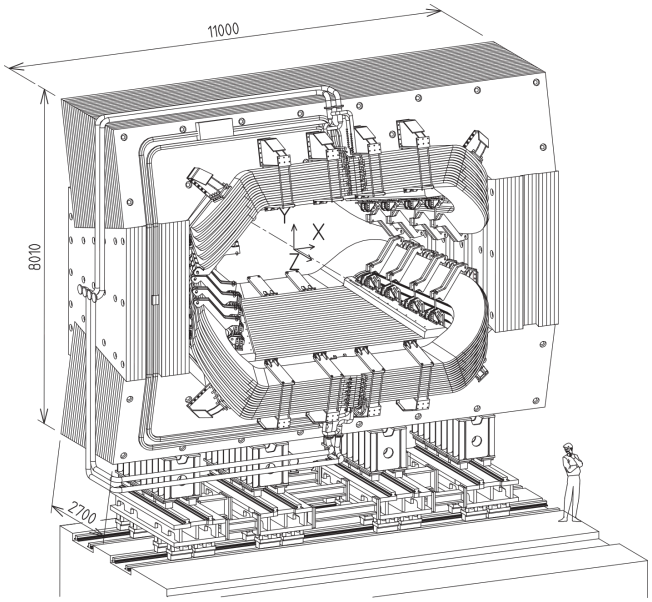


Figure 2.6: A schematic representation of the magnet inside LHCb (units expressed in mm) [53].

flowing inside the coils, therefore obtaining *up* and *down* configurations. The B_y value is mapped from the VELO to RICH2 with a precision of $< 4 \times 10^{-4}$T; as shown in Figure2.7, there is a residual field of 50 mT in the RICH system, with the peak field strength just between TT and T1.

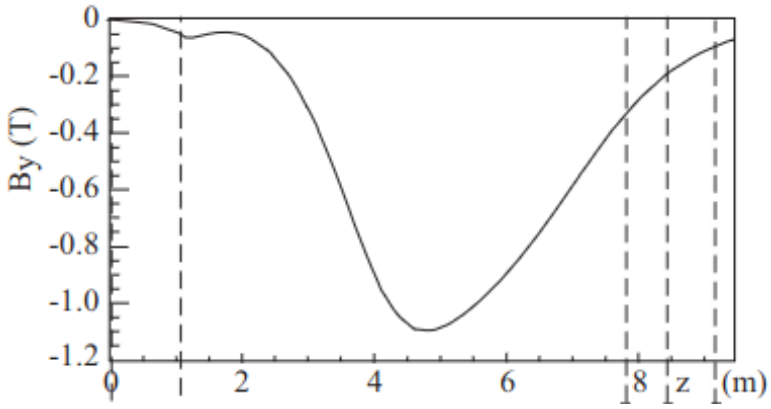


Figure 2.7: $B_y(T)$ value mapping as a function of z, from the interaction point up to the T1-3 tracking stations [56].

2.2.3 Tracking system

Together with the VELO, the tracking system provides information on tracks throughout the detector. It is made of four tracking stations on either side of the magnet: the Tracker Turicensis (TT) which is placed upstream of the magnet, and the T1, T2, and T3 stations, placed downstream of the magnet. The TT is 150 cm wide and 130 high, it covers the full geometrical acceptance of the detector and is devoted to the detection of low-momentum particles. It is made of two stations (TTa and TTb) separated by 0.3 m in the $z-$axis, each one with a pair of layers of 500 μm silicon sensors strips: the TTa station is made of $x-$ and $u-$layers, respectively tilted of 0° and $-5°$, while the TTb station is made of $v-$ and $c-$layers, with the $v-$layer tilted of 5° [58]. A schematic representation of the TT stations is shown in Figure 2.8. The single hit spatial resolution is around 50 μm for both the TT and

the IT.

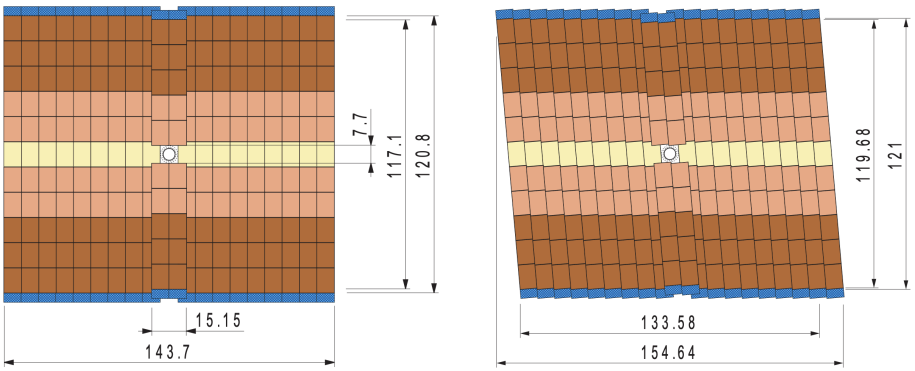


Figure 2.8: Representation of x- (left) and v-layer (right) of the TT station (units expressed in cm) [56].

The T1-3 stations are made of two distinct trackers: the inner tracker (IT) and the outer tracker (OT). The IT is 120 cm wide and 30 cm high, covering a cross-shaped region close to the beampipe, which is approximately 2% of the tracking acceptance. Given the high occupancy region (around 20% of tracks), the same silicon microstrip technology of the TT is used.

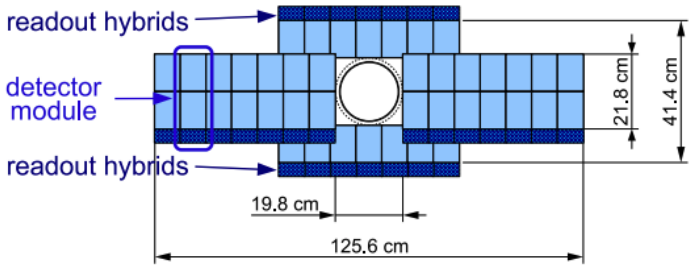


Figure 2.9: Representation of the x-layer of the IT station [56].

The OT [59] is a drift-time tracker which covers the external regions of the T1-3 stations. It is made of an array of 55000 straw-tube channels filled with a mixture of Argon, CO2, and O2, placed in two layers. The particles passing through the gas-filled tubes ionize the gas, and the resulting electrons are accelerated by a high potential difference; the delay in signal between the two layers is used to determine the hitting position. The OT has the same geometry of staggered layers with $\pm 5°$ tilting used in TT and IT. The OT defines a tracking acceptance of $2.0 < \eta < 5.0$ ($1.8 < \eta < 3.4$) in the vertical (horizontal) plane. The OT distance resolution is ~ 220 μm. A schematic view of the OT stations is shown in Figure 2.10.

2.2.4 Ring Imaging Cherenkov detectors

Particle identification (PID) is one of the fundamental tasks of the LHCb experiment. The good PID performance is reached thanks to the two Ring Imaging Cherenkov detectors [60, 61], RICH1 and RICH2 respectively placed upstream and downstream of the magnet, whose primary role is to perform identification of charged hadrons (such as π, K, and p). RICH1 is also able to help in the identification of charged leptons (e and μ), together with the calorimeters and the muon system. The functioning of a RICH detector is based on the Cherenkov effect. The cone angle of Cherenkov light

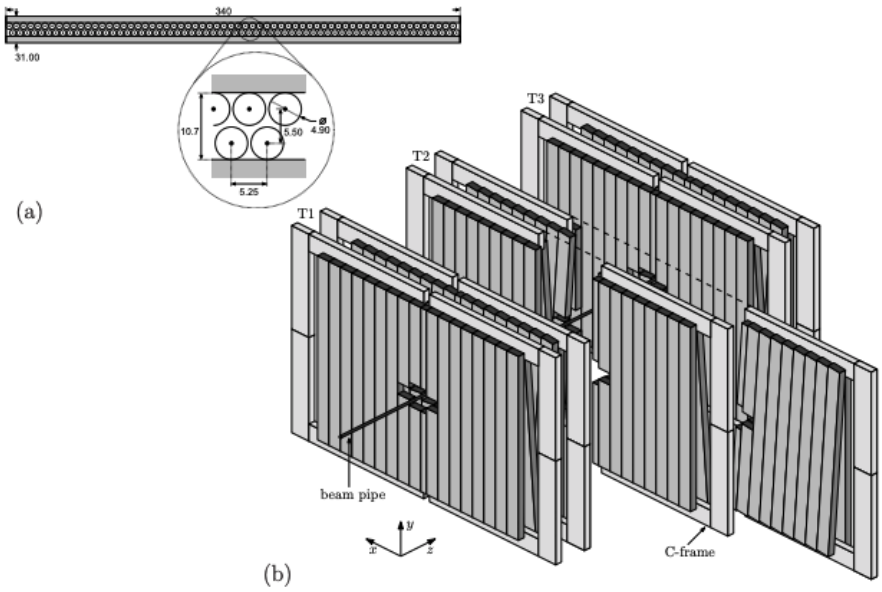


Figure 2.10: Schematic view of the OT stations, highlighting the section of the drift tube [56].

is directly related to the velocity of the particle by the relation

$$\cos\theta_c = \frac{c/n}{v}, \tag{2.2}$$

where θ_c is the cone angle, n is the refraction index of the medium, and c and v are respectively the speed of light and the velocity of the charged particle. By measuring the particle momentum p using the curvature of the trajectory, it is possible to compute the particle mass m and perform discrimination between charged hadrons:

$$m = \frac{\sqrt{n^2p^2\cos^2\theta_c - p^2}}{c}. \tag{2.3}$$

In order to have full coverage of the momentum range, two RICH detectors are used. The RICH1 is located upstream of the magnet at 1 m along the z-axis, between the VELO and the TT station, with an angular acceptance going from ±25 mrad to ±300 mrad (±25 mrad to ±250 mrad) in the horizontal (vertical) plane. Its active media are a solid aerogel radiator and a fluorobutane (C_4F_{10}) gas radiator, and it measures low momenta charged particles from 2 GeV/c to 40 GeV/c. The RICH2 is located downstream of the magnet at 9.5 m along the z-axis, between the T3 tracking station and the first muon station M1. Compared to the RICH1, it has a limited angular coverage, from ±15 mrad to ±120 mrad (±15 mrad to ±100 mrad) in the horizontal (vertical) plane. It is filled with carbon tetrafluoride (CF_4) gas radiator and it measures particle momenta from 15 GeV/c to 100 GeV/c. Each RICH detector uses spherical mirrors to focus the Cherenkov light emitted by charged particles, while flat mirrors are used to redirect light to photo-detectors. Hybrid photon detectors (HPD) made of 500×500 μm pixels collect Cherenkov light in the $200 - 600$ nm wavelength range and convert it into electrons to form an electric signal. A schematic view of RICH1 and RICH2 is shown in Figure 2.11.

2.2.5 Calorimeters system

The calorimeters system [62, 63] is made of four sub-detectors; these include the Scintillator Pad Detector (SPD), the Pre-Shower (PS) and two calorimeters, the Electromagnetic Calorimeter (ECAL) and

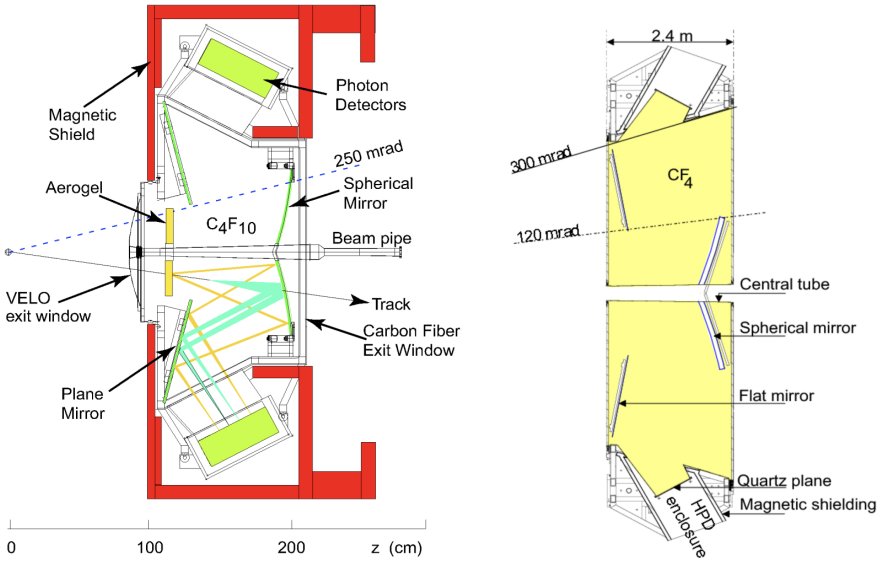


Figure 2.11: Schematic view of the RICH 1 (left) and RICH 2 (right) [56].

the Hadronic Calorimeter (HCAL), all of these placed right after the RICH2 and the M1 muon station, but before the M2-5 stations. A schematic view of the calorimeter system is shown in Figure 2.12. The calorimeters system has several functions:

- at the first level trigger (L0), it selects photons, electrons, and hadrons with significant transverse energies;
- it provides discrimination between electrons, photons, and hadrons by combining information coming from all its sub-detectors;
- it measures the energy of charged and neutral particles, particularly photons, π^0, and neutral hadrons.

The functioning of the calorimeter system is based on scintillation materials: particles interacting with scintillating materials produce showers of scintillation[3] light, which is transmitted by wavelength shifting (WLS) fibers and collected by photo-multiplier tubes (PMT). The scale of the scintillation response is directly related to the energy of the particle transversing the calorimeter system. Typical quantities that define the interaction between particles and calorimeters are the radiation length X_0 and the nuclear absorption length λ [5].

The calorimeters are sampling-calorimeters, therefore utilize alternate layers of scintillation materials and absorbers; this choice further reduces the radiation length and allows showers to be fully contained in the calorimeters, given that the shower penetration is directly related to the radiation length or, in the case of hadrons, to the nuclear absorption length.

Scintillator Pad Detector and Pre-shower

The SPD/PS system is responsible for electrons and photons identification. It consists of two planes of scintillator cells (the SPD is closest to the interaction region), crossed by WLS fibers and separated by a 15 mm lead converter. The cells are arranged with increasing granularity toward the beampipe. The angular acceptance is between ±30 mrad and ±300 mrad (±30 mrad and ±250 mrad) in the

[3]Bremsstrahlung and pair-production are responsible for electromagnetic showers, while the strong interaction is responsible for hadronic showers.

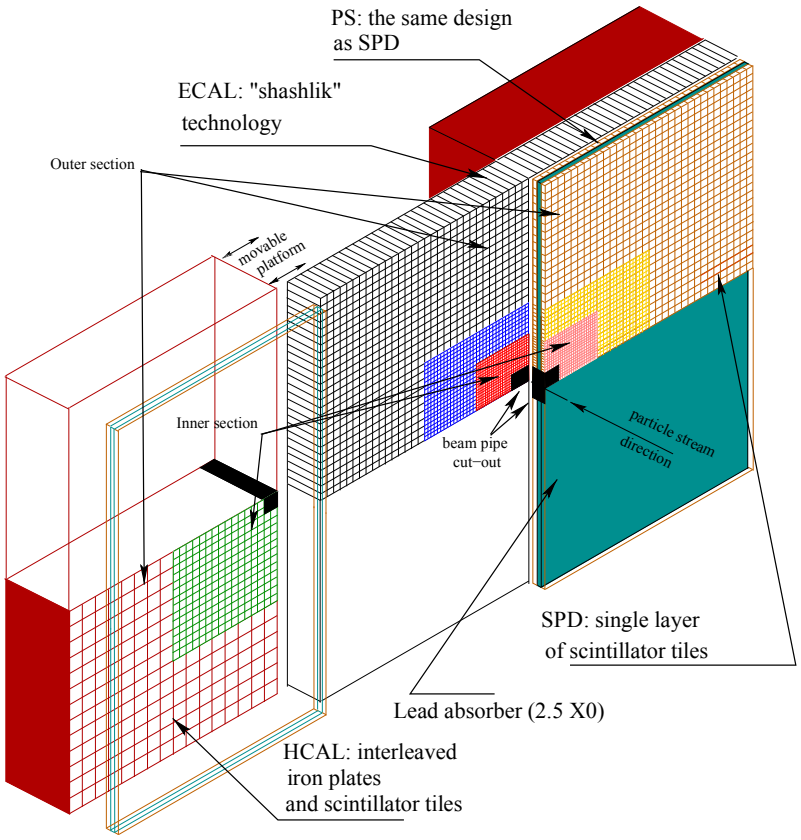


Figure 2.12: Schematic view of the calorimeter system, highlighting the different sub-detectors and their geometry [63].

horizontal (vertical) direction. The discrimination between electrons and photons is based on the fact that photons do not produce showers in the SPD but they produce showers in the PS; on the contrary, electrons produce showers in the SPD. Hadrons do not produce showers in the SPD/PS system, therefore allowing additional discrimination from electrons.

Electromagnetic calorimeter

The ECAL [64] is placed at ~ 12.5 m along the $z-$axis from the interaction point. It has an acceptance from ± 25 mrad to ± 300 mrad (± 250 mrad) in the horizontal (vertical) direction. The ECAL is composed of 66 layers of 2 mm lead tiles and 4 mm scintillator tiles, separated by 120 μm of a reflective paper, for a total depth of 42 cm, which corresponds to 25 radiation lengths. The granularity of ECAL increases toward the beamline, where the occupancy is greater; in this way, the ECAL surface is divided into 3 regions, inner, middle, and outer, with cells of respectively 4×4, 6×6 and 12×12 cm^2, as shown in Figure 2.13. The ECAL provides energy measurement with an energy resolution

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \tag{2.4}$$

where E is the energy measured in GeV and $\oplus$ indicates the sum in quadrature. The PMT gain is optimized for b-physics, typically in an energy range of few GeV. Due to this choice, cells experience ADC saturation for transverse energies E_T greater than 10 GeV: measurement of very high energetic

particles (*e.g.* electrons coming from $Z \to e^+e^-$, $W \to e\nu_e$ or $t\bar{t}$ processes) is therefore limited and performance is degraded. This particular aspect of ADC saturation is further studied in Chapter 6 in the context of the ECAL upgrade.

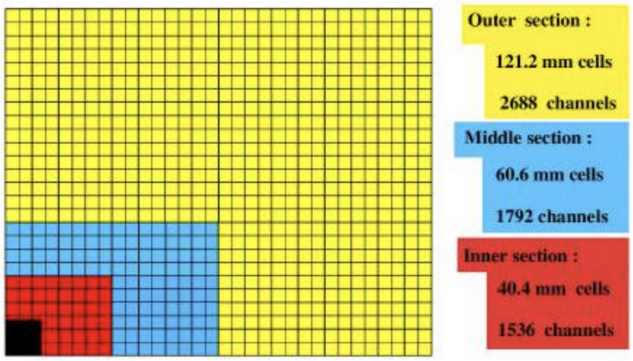


Figure 2.13: Segmentation of SPD/PS and ECAL. The different ECAL regions are highlighted in different colours [63].

Hadronic calorimeter

The HCAL is downstream of ECAL at $\sim$ 13.3 m from the interaction point, and it measures the remaining energy of neutral and charged hadrons. It is made of alternating iron and polystyrene tiles placed perpendicular to the x-axis. A single module extends downstream with three scintillating tiles and three absorber tiles, up to a depth of 128.3 cm for a total of 5.6 nuclear interaction lengths. As for ECAL, scintillation light is transmitted to PMTs through WLS fibers, which run along the edge of the layers of the tiles. The HCAL has the same ECAL angular acceptance, but its granularity is reduced to two regions, outer and inner with cells respectively of 26×26 and 13×13 cm^2, as shown in Figure 2.14. This is mainly to account for the larger size of hadronic showers. The HCAL energy resolution is extracted from a fit to the data at several energies

$$\frac{\sigma_E}{E} = \frac{(69 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%, \tag{2.5}$$

where E is the energy measured in GeV and the uncertainties come from the fit result; this energy resolution is sufficient for trigger purposes. The HCAL not only measures energy for neutral and charged hadrons, but it also provides stopping power containing hadronic showers before entering the muon stations.

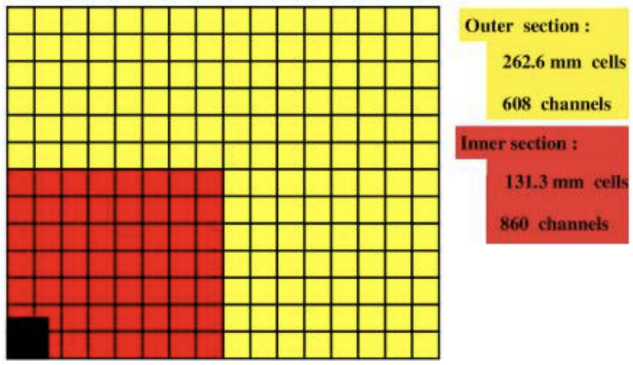


Figure 2.14: Segmentation of HCAL. The inner and outer sections are highlighted in different colours [63].

2.2.6 Particle identification performance (PID)

The PID is performed mainly using the RICH detectors and the calorimeter system [65]. The RICH detectors are built to discriminate between pions π and kaons K. The variable used for the discrimination is the difference of logarithmic likelihood

$$\mathrm{DLL}(\pi - K) = \ln \mathcal{L}_\pi - \ln \mathcal{L}_K, \tag{2.6}$$

where $\mathcal{L}_\pi$ ($\mathcal{L}_K$) is the likelihood for the π (K) mass hypothesis. If $\mathrm{DLL}(\pi - K)$ is measured over a specific threshold, the particle is identified as a π, otherwise as a K.
Given the PID task between two kinds of particles A and B, two important figures of merit are measured to quantify the PID performance:

- the identification efficiency for A, defined as the number of particles correctly identified as A over the number of total detected A particles;
- the B misidenfitication rate, defined as the number of B particles wrongly identified as A over the total detected B particles.

Identification efficiencies and misidentification rates have been measured by collecting data with a trigger configuration selecting hadrons; K, π and protons p have been selected from $K^0_S \to \pi^+\pi^-$, $\Lambda \to p\pi^-$ and $D^{*+} \to D^0(K^-\pi^+)\pi^+$ decays, and identification efficiencies and misidentification rates have been measured as a function of particles momentum, as shown in Figure2.15, where kaon identification efficiency and pion misidentification rate are shown. Depending on the chosen threshold, different behaviors are presented, generally showing better (*i.e.* higher identification efficiency and lower misidentification rate) performance for low momentum particles. The average kaon efficiency for $\mathrm{DLL}_{\mathrm{rich}}(K - \pi) > 0$ is 95% with a pion misidentification of 10%. As part of the PID system,

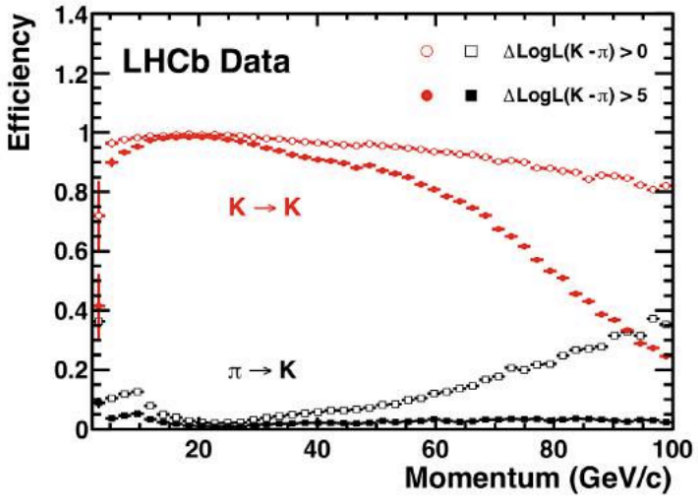


Figure 2.15: Kaon identification efficiency and pion misidentification rate as a function of particles momenta [56].

the calorimeter system can perform discrimination between electrons and hadrons; one of the most discriminating variables is the ratio of the ECAL cluster energy over the particle momentum p: as shown in Figure 2.16 this provides good separation between electrons and hadrons. As done for RICH detectors, a similar difference of logarithmic likelihood $\mathrm{DLL}_{\mathrm{calo}}(e - h)$ is defined. Without using RICH detectors information, an average electron identification efficiency of 91.9% is reached for $\mathrm{DLL}_{\mathrm{calo}}(e - h) > 2$, with a hadron misidentification rate of 4.5%. Including also RICH detectors in the likelihood evaluations, the electron identification efficiency increase to 97%, while the hadron misidentification rate decrease to 2%.

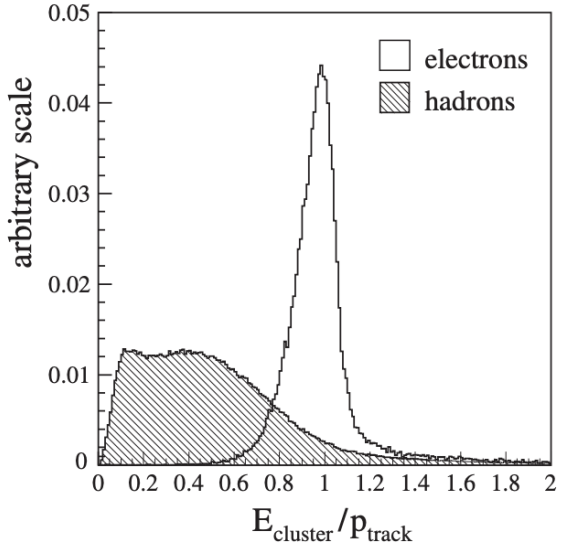


Figure 2.16: Ratio between the energy of a ECAL cluster E_{cluster} and the particle momentum measured by the tracking system p_{track} for electrons and hadrons [56].

2.2.7 The muon system

The muon system [66] is the detector most distant to the interaction point, and it provides muon identification. The muon system is made of five stations: the first M1 station is placed between RICH2 and the calorimeter system, while the remaining M2-5 stations are placed downstream of the calorimeter system. The angular acceptance of the muon system lies between ±20 mrad and ±306 mrad (±16 mrad and ±258 mrad) in the horizontal (vertical) direction. The detection technology of the muon system relies on multi-wire proportional chambers (MWPCs), analogous to straw drift chambers, that collect the electron cascades generated by gas ionization from interaction with a muon. The only exception is the inner part of the M1 station, which uses triple-GEM detectors in order to cope with high radiation and to provide comparable tracking performance: the rate of charged particles in this region is approximately 500 kHz/cm^2. All the stations are divided into chambers of different sizes and granularity (projective with respect to the interaction point), which are themselves divided into logical pads, which define the x and y spatial resolutions. The M2-5 stations are interleaved with iron absorbers, giving a total thickness of $\sim$ 20 interaction lengths, necessary to select muons with $p > 6$ GeV/c. A particle is identified as a muon if it crosses all five stations in a straight line (there is no residual magnetic field); the muon identification efficiency is 98% on average, with a pion and kaon misidentification $< 1\%$. The first three chambers provide high-precision measurements of the muon p_{T}, with a resolution of $\sim 20\%$. The M4-5 stations are mainly used for particle ID. A schematic view of the geometry of muon chambers is shown in Figure 2.17.

2.2.8 Trigger

The LHCb experiment collects data at an interaction rate of 40 MHz, which is completely off the bandwidth and storage capabilities of the detector. Therefore, the data acquisition rate has to be lowered at a rate of $\sim$ 12.5 kHz, for data to be stored and analyzed. This happens through a real-time selection, provided by event-by-event analysis. The trigger is composed of two parts: a hardware-based Level-0 (L0) trigger, synchronous with the bunch crossing frequency, that selects events using information from the calorimeter and the muon system, and a two-part software-based High Level Trigger (HLT), which provides selections based on full event reconstruction, which is executed asynchronously on a processors farm. A scheme of the trigger stages performed during Run

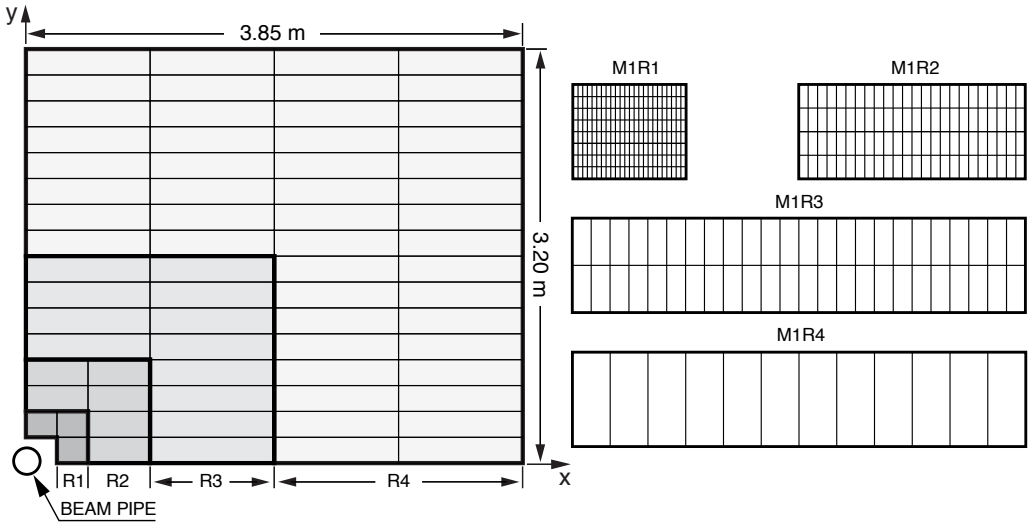


Figure 2.17: Front view of a quadrant of a muon station, where each rectangle represents a muon chamber (left). Different muon chambers of the M1 station (right) [66].

2 is shown in Figure 2.18 [67].

Level-0 Trigger

The L0 trigger uses information from the calorimeter and the muon systems to select events with electrons, photons, hadrons and muons with high transverse momentum p_{T} or high transverse energy E_{T}. The calorimeter and muon systems are connected to an L0 Decision Unit (L0-DU) that decides if the event is accepted, reducing the rate from 40 MHz to 1 MHz, where events with electrons or photons, hadrons, and muons add up 150, 450 and 400 kHz respectively [68]. Additionally, a pile-up system gives information on the number of measured primary vertices (PVs) in an event. For ECAL and HCAL, the decision to trigger an event is based on the transverse energy deposited in clusters made of 2×2 cells, defined as

$$E_{\mathrm{T}} = \sum_{i=1}^{4} E_i \sin\theta_i \tag{2.7}$$

where E_i is the energy deposited in the i-cell and θ_i is the angle between the z-axis and a line from the interaction point to the center of the $i-$cell. An event is accepted if one of these conditions is satisfied:

- `L0Muon`: a straight-line track traversing all five stations with a p_{T} exceeding a `L0Muon` threshold;
- `L0DiMuon`: two tracks identified as muons with the product of their p_{T} above a given threshold, only if the largest p_{T} track has failed the `L0Muon` requirement;
- `L0Electron`: an ECAL cluster (as defined in Eq. 2.7) with E_{T} over a given threshold, with hits in the PS and in the SPD;
- `L0Photon`: a ECAL cluster with E_{T} over a given threshold and no hits in the SPD;
- `L0Hadron`: a HCAL cluster with E_{T} over a given threshold.

L0 thresholds for E_{T}, p_{T} and SPD hits have changed during the years of data taking; The selected thresholds for this thesis are shown in Table 2.2.

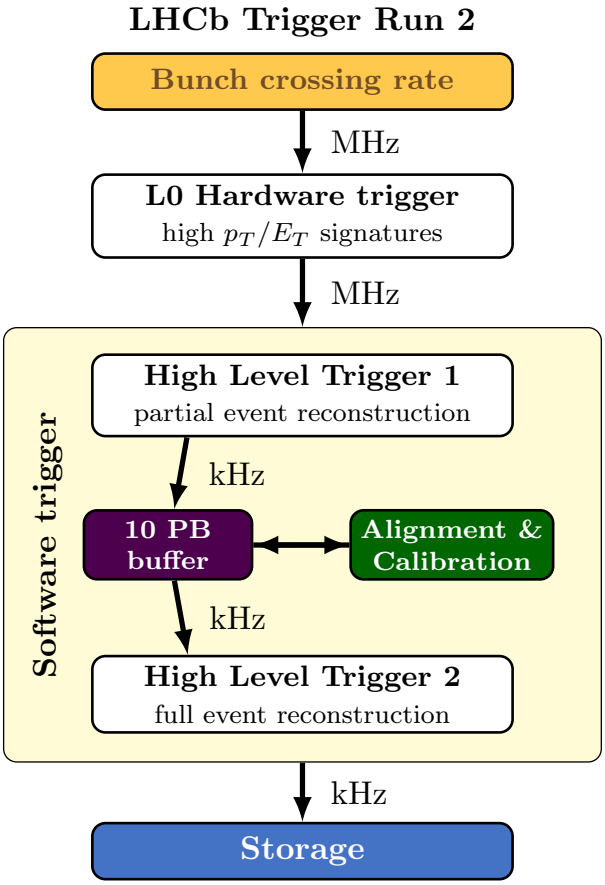


Figure 2.18: General scheme of the LHCb trigger used during Run 2 [68].

High Level Trigger

Events selected by the L0 trigger are fed into the HLT at a rate of 1 MHz. The HLT is divided in two stages: the HLT1, which further reduces the L0 output rate to 110 kHz, and the HLT2, which send events to permanent storage at a rate of 12.5 kHz [68]. It runs on a large computing infrastructure called Event Filter Farm, which allows the HLT1 output to be written in a local buffer to perform quality offline reconstruction. The HLT1 performs track reconstruction, using information from the VELO and the T1-3 stations to produce long tracks which are fitted and evaluated to reject fake tracks. The VELO tracks are used to reconstruct PVs. HLT1 selects events containing muon candidates displaced from the PVs and di-muon candidates, performing muon ID with tracks with at least 2 hits in the muon stations and $p > 3$ GeV. The HLT2 performs full event reconstruction on top of the HLT1 output. Several algorithms are put in place, to select candidates with sufficient p_{T} and significant displacement from the PVs. Examples of HLT2 algorithms are [68]:

- inclusive topological algorithms, reconstructing b hadrons' decays by selecting 2-,3-, or 4-body vertices. They account for $\sim 40\%$ of the HLT2 output rate;
- exclusive and inclusive c-hadrons algorithms, reconstructing D mesons from specific nodes (exclusive triggers) or by combining all charged hadrons in the event (inclusive triggers). They account for $\sim 40\%$ of the HLT2 output rate;
- muon and di-muon algorithms, reconstructing high p_{T} muons or di-muons depending on the physics goal;

line	p_T [GeV/c]	$p_T^1 \cdot p_T^2$ [GeV^2/c^2]	E_T [GeV]	nSPD
L0-Muon	> 1.76	-	-	< 600
L0-DiMuon	-	$> (1.6)^2$	-	< 900
L0-Hadron	-	-	> 3.7	< 600
L0-Photon	-	-	> 3	< 600
L0-Electron	-	-	> 3	< 600

Table 2.2: Level 0 trigger thresholds considered in this thesis [68].

- electroweak physics and exotic searches algorithms, accounting for the remaining HLT2 output rate, select specific physics channels.

In the analyses presented in this thesis, muon and di-muon algorithms are used to reconstruct muons and di-muons coming from the decay of $W^{\pm}$ and Z bosons.

2.2.9 LHCb framework and data flow

In a typical analysis, the main ingredients are the data and the tools used to analyze this data. One of the most important tools consists in MC simulations of the physics processes under study and how they interact with the detector. The generation of physics processes has been already described in Chapter 1. Both data and MC simulations need to go through a data flow that prepares them for analysis.

The LHCb framework and data flow consist of several steps, from the measurement (or the simulation) of particle-matter interactions up to the analysis level. A schematic view of the LHCb data flow during Run 2 operations is shown in Figure 2.19. The whole software framework is based on the Gaudi [69] framework.

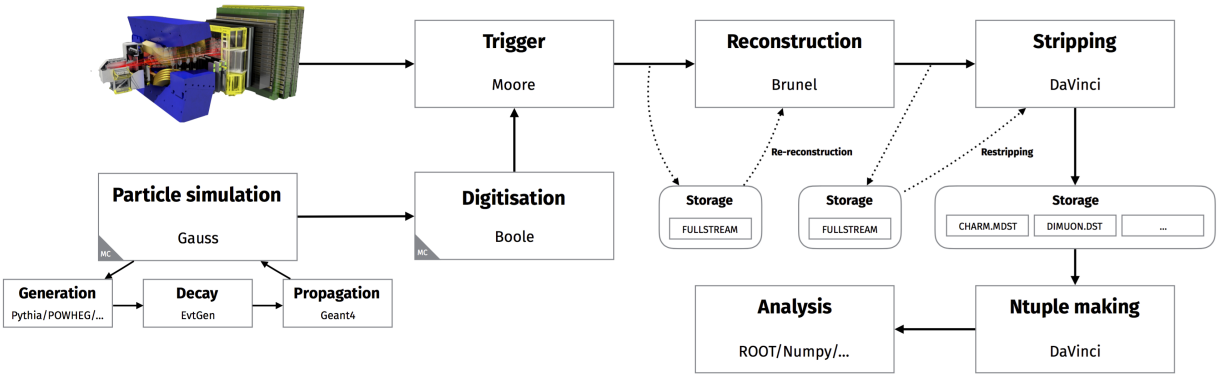


Figure 2.19: Chartflow highlighting the LHCb software framework for simulation, reconstruction and data analysis [56].

The steps (and the relative software packages) can be summarized in the following way:

1. data acquisition/simulation: this step depends on whether we are dealing with data or simulations. When dealing with data, this accounts for the detection and measurement of proton-proton collisions in the LHCb experiment. In the case of simulations, proton-proton collisions are generated and particle-detector interactions are simulated. to mimic the detector response. The simulation step is performed by the software Gayss [70] and consists of several sub-steps:

   (a) the hard proton-proton scattering process is simulated by a generator program (e.g. Mad-

GRAPH [26, 27] or PYTHIA [29], the ones considered in this thesis), while the quarks fragmentation and hadronization are simulated by a parton shower (PYTHIA [29]);

(b) particles are propagated to the detector and decays of hadronic particles are simulated using EVTGEN;

(c) the interactions of stable particles reaching the detector and beam pipe materials are simulated using the GEANT4 [71, 72] software.

2. digitization: once the simulations of the physics processes and their interaction with the detector have been performed, particle-matter interactions are processed to create hits in the detectors, by reproducing the digitization of the front-end electronics. The software devolved to this task is called BRUNEL [73], which includes all the instrumental effects (e.g. noise, inefficiencies).

3. trigger: at this stage, data and simulations are treated in the same way. Data or simulated events that pass specific cuts are selected by the trigger system, as explained in Section 2.2.8. The software MOORE [73] operates both online, applying HLT algorithms to events accepted by the L0 stage, and also offline.

4. reconstruction: the software BRUNEL [73] performs the full reconstruction of events that pass the trigger step. The muon and jet reconstruction algorithms used in the analyses presented in this thesis are described in Chapter 3.

5. stripping: the software DAVINCI [73] in different *streams*, depending on the objects reconstructed and their features. In this thesis, the `WMuLine` stripping line is used, grouping together all the events where a high p_{T} muon is present.

6. analysis: the events of a specific stream are finally processed to reconstruct a specific decay channel. The output is a ROOT [74] n-tuple that is afterward analyzed with the preferred analysis tools.

Chapter 3

Objects reconstruction at LHCb

While the previous Chapter described the various sub-detectors of the LHCb experiment, this Chapter is devoted to the event reconstruction. Particular care is given to the reconstruction of muons and jets, which are the main final objects of the analyses presented in this thesis. This Chapter has the following structure:

- Section 3.1 describes the reconstruction of tracks and vertices;
- Section 3.2 is devoted to the description of muon reconstruction and identification, dealing in particular with the performance of reconstructing and identifying high p_{T} muons;
- Section 3.3 describes the algorithms for jet reconstruction and identification;
- finally, Section 3.4 deals with the state-of-the-art algorithms used at the LHCb experiment to identify the flavor of the quarks generating the jets. These algorithms go by the name of "jet tagging", and possible improvements of these algorithms are presented in Chapter 7.

3.1 Pattern recognition

The first task in event reconstruction is the reconstruction of tracks and vertices. Tracks are built by employing pattern recognition techniques, that associate information coming from hits in the sub-detectors into trajectories that are afterwards fitted. In order to avoid any double counting or detector mismatch that generates *fake tracks*, tracks undergo several selections. By intersecting tracks, vertices are found and finally, particles are identified by the presence of signals in other sub-detectors, such as the calorimeters or the muon system.

3.1.1 Tracks

The tracking system includes VELO, TT, IT and OT. At first, seeding algorithms are used to reconstruct track segments. This happens in the VELO, where from hits in the (r,ϕ,z) coordinates system, straight lines are extrapolated to the interaction point, forming VELO seeds. In the T-stations, where the magnetic field is present, seeds are formed using lines in the (x,z) plane. Once seeds are found, several algorithms combine them in a precise sequence:

- forward tracking: to provide trajectories, VELO seeds are combined with remaining single hits in the tracking stations, including hits that are consistent with the reconstructed path;
- track matching: VELO and T-stations seeds not used in the forward tracking are considered, extrapolating VELO seeds through the magnetic field and combining them with T-stations

seeds, if compatible. Further hits in TT are afterwards added if they are consistent with the track trajectory;

- upstream and downstream tracking: remaining unpaired seeds from VELO and TT are extrapolated to the T-stations, where individual hits are used to compute momenta.

Once hits have been grouped together, they are used as input for the track fitting procedure: besides providing the final track, this procedure gives also the momentum of the particle and the sign of its charge. t The best estimate of the track parameters is obtained with a Kalman fit [75]. Several classes of tracks are defined, depending on which sub-detector information is used to create tracks:

- long tracks, which leave hits throughout all the tracking system;
- downstream tracks, which hit the TT and the T1-3 stations only;
- upstream tracks, hitting only the VELO and the TT;
- VELO tracks, which leave hits only in the VELO;
- T tracks, with information coming only from the T1-3 stations.

A plot representing all the types of tracks is shown in Figure 3.1.

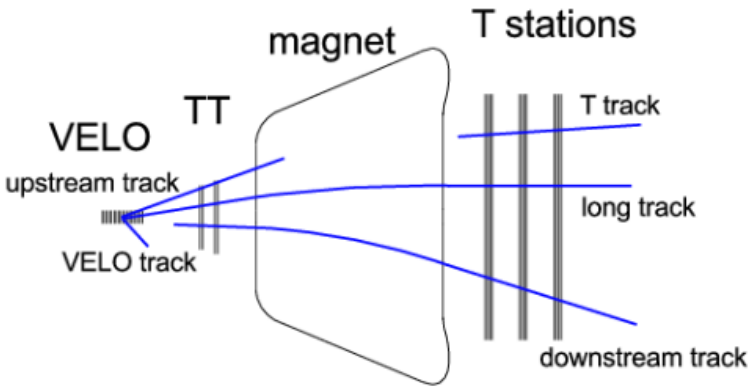


Figure 3.1: Representation of all the types of reconstructed tracks in the LHCb detector.

In order to determine the momentum resolution ($\delta p/p$) of long tracks, benchmark processes are considered such as $J/\psi \to \mu^+\mu^-$ decays, collected from a data sample with a particular trigger configuration to select high energy di-muons; for momentum particles below 20 GeV/c resolution is about 0.5% while it is 0.8% for momentum particles around 100 GeV/c. For the invariant mass resolution (σ_m/m where σ_m is the peak width of the resonance at a mass m) six different resonances have been used, again observed in di-muons samples: J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and Z boson; for particle mass below 10^4 MeV/c^2 resolution is 0.45% while it is around 1.9% for masses around 10^5 MeV/c^2 [76]. Results for both momentum and invariant mass resolution are shown in Figure 3.2.

3.1.2 Vertices

From the detected and reconstructed tracks, it is also possible to define the Primary Vertex (PV) [77], which is the proton-proton interaction point where almost all the particles are generated. First of all, tracks are clustered in seeds; if the cluster has less than six tracks, it is discarded. After this, for each

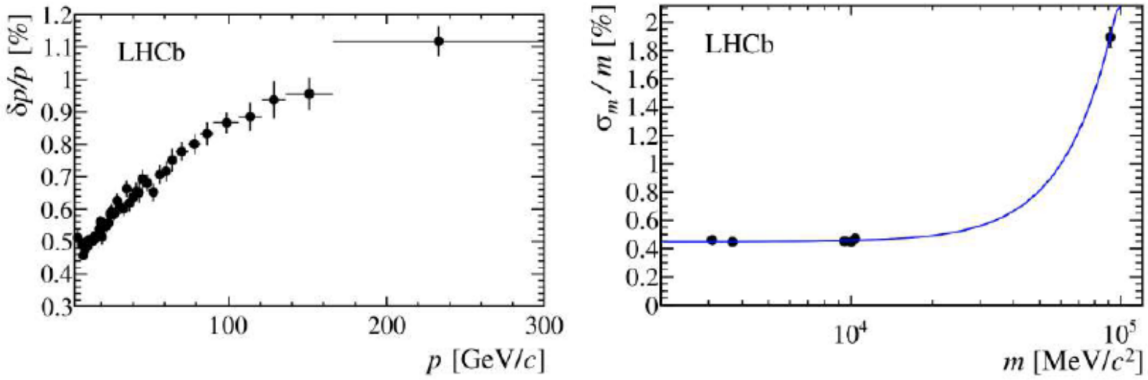


Figure 3.2: (left) Momentum resolution of LHCb tracking system in Run 1. (right) Relative mass resolution as a function of di-muons mass resonances, measured in Run 1.

seed the PV position is calculated minimizing the following quantity:

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \frac{d^2_{0i}}{\sigma^2_{d_{0i}}} \tag{3.1}$$

where d_{0i} is the impact parameter of the $i-$track and $\sigma_{d_{0i}}$ its error. If one or more tracks (inside a seed) have $\frac{d_0}{\sigma_{d_0}} > 4$ then the one with the highest $\frac{d_0}{\sigma_{d_0}}$ is removed from the cluster, and a new PV position is computed by minimizing the new χ^2_{PV}. This procedure goes on till there are no more tracks to discard. The PV reconstruction efficiency and resolution are obtained from simulations: the average efficiency goes from 90.0% to 97.5% while the probability of reconstructing a false PV is about 1%; as for PV resolution, typical values are about 8 μm, 10 μm and 50 μm respectively for the x, y, and z coordinates.

3.2 Muon reconstruction and identification

The identification of muons in LHCb is mostly based on the Muon detector: only muons are able to penetrate the calorimeters and muon filters and thus leave a signal on the dedicated muon chambers. The first step in muon reconstruction is to define a boolean variable, called `IsMuon` [78], from the extrapolation of a long or a downstream track through the muon stations, making a statement about whether a track is consistent with a muon hypothesis. Only tracks within the acceptance of the muon stations are extrapolated, obtaining the expected track coordinates in the muon stations M2, M3, M4, and M5. For each of the stations, a search for hits within an elliptic, momentum dependent, field of interest (FoI) around the extrapolated track is performed, defined as

$$\mathrm{FoI}_a(p) = \rho_{a,1} + \rho_{a,2} \cdot \exp\left(\frac{-\rho_{a,3} \cdot p}{GeV}\right) \tag{3.2}$$

where $a \in (x, y)$ and $\rho_{a,i}$ with $i = 1, 2, 3$ are parameters defining the FoI [78]. If the track has at least one hit within the FoI in a minimum number of stations, it is given the attribute `IsMuon`. Two other boolean variables are constructed similarly to `IsMuon`: `IsMuonLoose`, which requires a fewer amount of hits with respect to `IsMuon`, and `IsMuonTight`, which requires the same amount of hits but using only crossed hits. An overview of the requirements for the `IsMuon` and `IsMuonLoose` variables is reported in Table 3.1.

p [GeV]	Required stations `IsMuon`	 `IsMuonLoose`
$p < 3$	Always false	Always false
$p < 6$	M2 and M3	At least two of M2-M4
$6 < p < 10$	M2 and M3 and (M4 or M5)	At least three of M2-M5
$p > 10$	M2 and M3 and M4 and M5	At least three of M2-M5

Table 3.1: `IsMuon` and `IsMuonLoose` requirements as a function of muon momentum p [78].

3.2.1 Muon reconstruction efficiency

The efficiency $\varepsilon^{\mu}_{\text{reco}}$ of properly reconstructing a muon is factorized into three components as

$$\varepsilon^{\mu}_{\text{reco}} = \varepsilon^{\mu}_{\text{trg}} \times \varepsilon^{\mu}_{\text{trk}} \times \varepsilon^{\mu}_{\text{id}}, \tag{3.3}$$

where $\varepsilon^{\mu}_{\text{trg}}$ is the muon trigger efficiency, $\varepsilon^{\mu}_{\text{trk}}$ is the long track reconstruction efficiency, and $\varepsilon^{\mu}_{\text{id}}$ is the muon-identification efficiency [79]. MC simulations cannot be used directly to measure $\varepsilon_{\text{reco}}$, given that they may not properly model some crucial detector effects. Therefore, real data are used with a *tag-and-probe* technique: data with two muons are reconstructed and selected, and while one of the two muons is required to pass an additional specific selection (the *tag* muon), the efficiency is measured by counting how many times the other muon (*probe*) passes the chosen selection. The probe muon is used to determine the corresponding efficiency for that specific selection. For high-p_{T} muons as the ones analyzed in this thesis, the leptonic decay of the Z boson into two muons is chosen, given its high signal purity and the proper p_{T} range typical of these analyses. Data coming from the Run 2 data-taking campaign are used to compute the efficiencies.

Trigger efficiency

To compute the trigger efficiency, muons from $Z \to \mu\mu$ events are required to trigger dedicated lines at the three trigger level stages: `L0MuonEW`, `Hlt1SingleHighPTMuon` and `Hlt2EWSingleMuonVHighPT` respectively for the L0, HLT1, and HLT2 stage. Selection requirements placed on the $Z \to \mu\mu$ events are summarised in Table 3.2, where $I_{\text{T}}^{\text{charged}}$ is a variable which ensures that muons are isolated particles, defined as

$$I_{\text{T}}^{\text{charged}} = \frac{p^{\mu}_{\text{T}}}{p^{\mu}_{\text{T}} + p^{\mu-cone}_{\text{T}}} \tag{3.4}$$

where p^{μ}_{T} is the muon transverse momentum and $p^{\mu-cone}_{\text{T}}$ is the total transverse momentum of the particles contained in a cone of radius $R = 0.5$ around the muon.

Tag	Probe	Event
Long track	Long track	$60 < M_{\mu\mu} < 120$ GeV
isMuon	isMuon	
Triggered	-	$\chi^2_{vtx}/\text{ndf} < 5$
$\sigma_P/P < 0.1$		
$p_{\text{T}} > 20$ GeV	$p_{\text{T}} > 20$ GeV	
$2 < \eta < 4.5$	$2 < \eta < 4.5$	
$I_{\text{T}}^{\text{charged}} > 0.85$	$I_{\text{T}}^{\text{charged}} > 0.85$	

Table 3.2: Selection requirements for the tag muon, the probe muon and the whole event for the measurement of the trigger efficiency $\varepsilon^{\mu}_{\text{trg}}$ [79].

To check the validity of the tag-and-probe method, a comparison by using the truth muon information in simulated $Z \to \mu\mu$ samples is done, showing that the tag-and-probe method does not introduce any

bias. The muon trigger efficiency might be affected by background events, particularly two sources of background are found:

- misidentification background, which is studied by applying a tag-and-probe approach on data events containing two muons with the same charge. The background is found to be negligible;
- heavy flavor (HF) background coming from heavy flavor processes that could mimic the muon signal. Given that here the selection requirement $\chi^2_{vtx}/\text{ndf} < 5$ has been applied, the HF background is found to be negligible.

The only uncertainty for the trigger efficiency arises from the statistical uncertainty of the $Z \to \mu\mu$ data sample, which is found to be $\sim 0.002\%$.

Figure 3.3 shows muon trigger efficiencies for 2016, 2017, and 2018 data as a function of muon p_{T} and η.

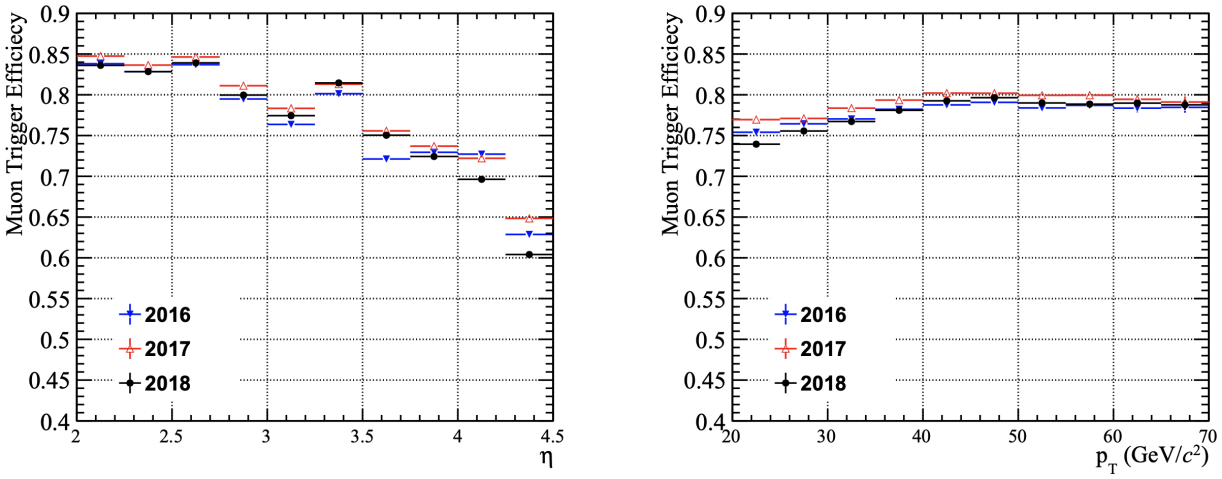


Figure 3.3: Muon trigger efficiency $\varepsilon^{\mu}_{\mathrm{trg}}$ for 2016, 2017, and 2018 datasets as a function of (left) muon η and (right) muon p_{T} [79].

Tracking efficiency

To compute the tracking efficiency $\varepsilon^{\mu}_{\mathrm{trk}}$, the tag muon coming from $Z \to \mu\mu$ events must be a long track that passes both the muon trigger and the identification requirements. The probe is a MuonTT track, a track reconstructed by combining hits in the TT and the muon stations: this choice is related to the fact that long tracks leave hits in the VELO and in the T-stations, but not necessarily in the TT. Therefore, by choosing a MuonTT track, the efficiency is not expected to have a dependence on hits from these particular sub-detectors. A list of the requirements placed on the tag and the probe muon is presented in Table 3.3.

Differently from the evaluation of trigger and identification efficiencies, the tracking efficiency determined using the tag-and-probe method is biased by different effects, that need to be corrected:

- bias correction: the tag-and-probe method use a MuonTT track as a probe particle, however not all long tracks have an associated MuonTT track. This is corrected using simulated signal events and relying on the truth particle information;
- track matching efficiency: a MuonTT track is matched to a long track by imposing some matching conditions, which might introduce inefficiencies in the track matching. These are studied

Tag	Probe	Event
Long track	Muon TT track	$70 < M_{\mu\mu} < 110$ GeV
isMuon	-	$\lvert\Delta\phi\rvert > 0.1$ rad
Triggered	-	$\chi^2_{vtx}/\text{ndf} < 5$
$\sigma_P/P < 0.1$		
$\chi^2_{\text{IP}} < 6$	$\chi^2_{\text{IP}} < 6$	
$p_{\text{T}} > 20$ GeV	$p_{\text{T}} > 20$ GeV	
$2 < \eta < 4.5$	$2 < \eta < 4.5$	

Table 3.3: Selection requirements for the tag muon, the probe muon and the whole event for the measurement of the trigger efficiency $\varepsilon^{\mu}_{\text{trk}}$ [79].

in data and simulations using an "inverted" tag-and-probe method, which shows a difference of 0.28% in the track matching efficiency.

Results for the corrected muon tracking efficiency $\varepsilon^{\mu}_{\text{trk}}$ for 2016, 2017, and 2018 datasets are shown in Figure 3.4, as a function of the muon p_{T} and η.

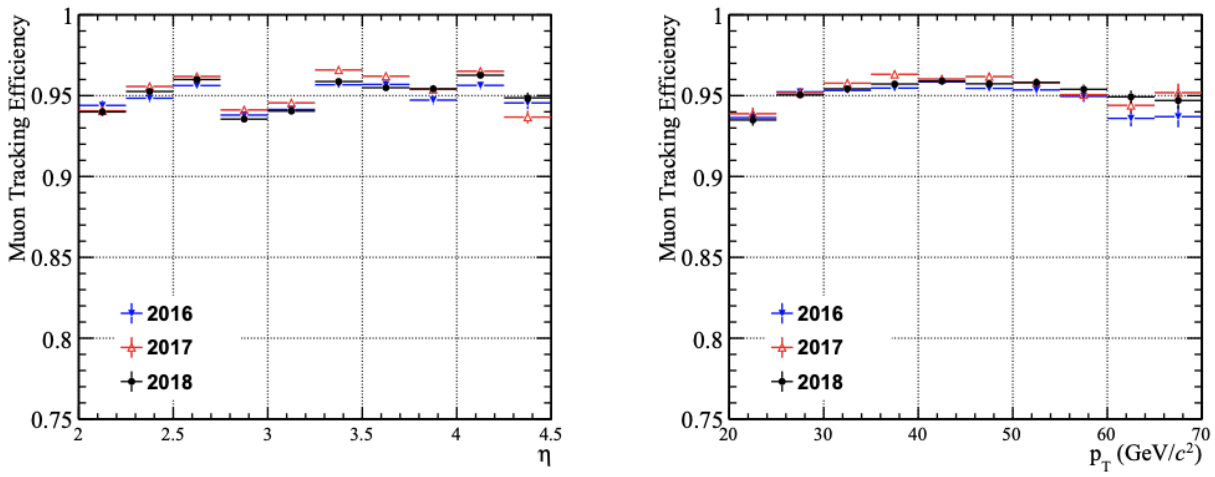


Figure 3.4: Muon tracking efficiency $\varepsilon^{\mu}_{\text{trk}}$ for 2016, 2017, and 2018 datasets as a function of (left) muon η and (right) muon p_{T} [79].

Identification efficiency

Finally; the muon identification efficiency $\varepsilon^{\mu}_{\text{id}}$ is the probability that a real muon passes the muon quality requirements. The tag muon coming from $Z \to \mu\mu$ events is required to be fully reconstructed, identified, and triggered, while the probe muon is a long track without any requirement on muon identification. The muon identification efficiency is calculated by counting the number of probe muons satisfying the `isMuon` requirement. An overview of the selection requirements is shown in Table 3.4.

The tag-and-probe approach is compared with the *Truth method* using simulated $Z \to \mu\mu$ events, which uses particle true information from MC simulations: good agreement in all regions is found, confirming that no bias is introduced by the tag-and-probe approach. The main background for $Z \to \mu\mu$ selection comes from Heavy flavor (HF) processes, where a muon with high p_{T} coming from the decay of a heavy flavor quark is produced. To account for this background source, and to consider also detector resolution effects, the invariant mass shape in data $m_{\text{inv}}(data)$ is fitted by the following

Tag	Probe	Event
Long track	Long track	$60 < M_{\mu\mu} < 120$ GeV
isMuon	-	$\lvert\Delta\phi\rvert > 2.7$
Triggered	-	$\chi^2_{vtx}/\text{ndf} < 5$
$\sigma_P/P < 0.1$		
$p_\text{T} > 20$ GeV	$p_\text{T} > 20$ GeV	
$2 < \eta < 4.5$	$2 < \eta < 4.5$	

Table 3.4: Summary of the requirements applied to identify a muon [79].

function

$$m_\text{inv}(data) = m_\text{inv}(Z \to \mu\mu) \times Gausssian + Exp(\text{HF}) \tag{3.5}$$

where $m_\text{inv}(Z \to \mu\mu)$ is the invariant mass shape taken from simulations, *Gaussian* is a Gaussian function to account for detector effects and Exp(HF) accounts for the HF background. Systematic uncertainties are found to be much smaller than statistics uncertainties, therefore no systematic uncertainty is assigned. Results for muon identification efficiency for 2016, 2017, and 2018 datasets are shown in Figure 3.5, as a function of the muon p_T and η.

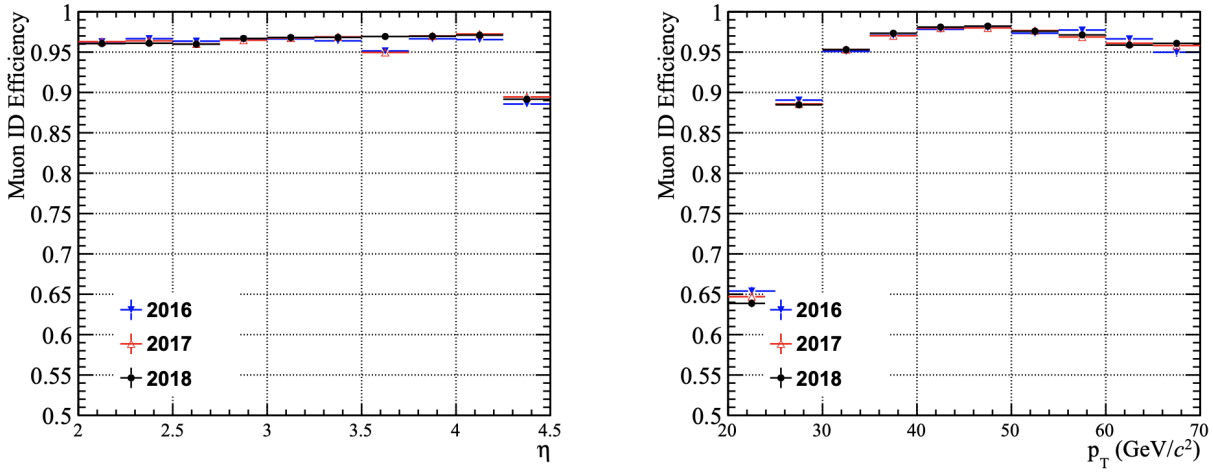


Figure 3.5: Muon identification efficiency $\varepsilon^\mu_\text{id}$ for 2016, 2017, and 2018 datasets as a function of (left) muon η and (right) muon p_T [79].

3.3 Jet reconstruction and identification

The analyses presented in Chapters 4 and 5 rely on the ability to properly reconstruct jets. In proton-proton collisions, jets are identified as collimated streams of particles resulting from the fragmentation and hadronization processes of quarks and gluons. Due to QCD properties, bare quarks cannot exist and through fragmentation, they produce showers of particles that subsequently hadronize into stable hadrons, that when interacting with the detector produce the typical signature of a jet. Being a collimated stream of particles, a jet leaves information both in the tracking system (by means of its charged particles) and in the calorimeter system (both charged and neutral particles leaving energy). Given that the LHCb experiment was originally designed to study b physics, the calorimeter system is not optimized to study high p_T objects such as jets. As described in Section 2.2.5, ECAL has a very low ADC saturation and HCAL is mainly used for L0 trigger. On the other side, the LHCb tracking system has excellent performance, therefore a hybrid algorithm that uses information coming

both from tracking and calorimeter systems is necessary to reconstruct jets: in this way, the tracking system selects charged particles while the calorimeters system selects neutral energy deposits. The LHCb jet reconstruction algorithm consists of several steps:

- Particle Flow algorithm [80]: it selects tracks from the tracking system and energy clusters in the calorimeter system to be used as input for the clustering step;
- clustering: particles selected by the Particle Flow step are clustered into jets using the anti-k_{T} algorithm;
- jet 4-momentum computation: kinematics of the jets is computed from the 4-momenta of jet constituents;
- Jet Energy Correction: the jet 4-momentum is corrected with a correction factor that depends on the jet kinematics.

3.3.1 Particle Flow

The Particle Flow (PF) [80] algorithm selects tracks and energy clusters in the calorimeters to define inputs (and therefore particles) to be used by the clustering algorithm to form a jet. In this way, it provides a sequence to event reconstruction and provides neutral energy recovery. It consists of several steps, shown in Figure 3.6 and summarised here:

- reconstructed tracks and clusters are paired with PID information and matched to each other to form *isolated charged particles*;
- energy released by tracks in the calorimeter systems is estimated and subtracted from the energy of the associated cluster. The resulting clusters are defined as *non-isolated particles* and they are used for neutral energy recovery;
- the remaining clusters that are not matched to tracks are selected as *isolated neutral particles*.

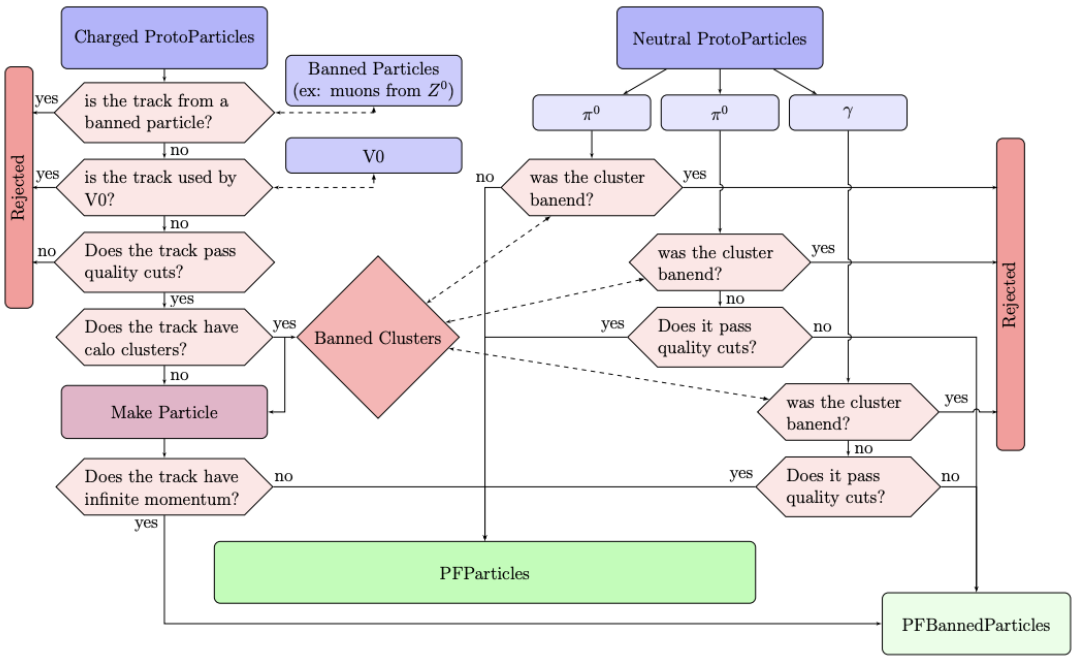


Figure 3.6: Flowchart of the Particle Flow algorithm [81].

In the following, a brief overview of the requirements for isolated charged particles, non-isolated, and isolated neutral particles is given.

Isolated charged particles

In order to improve the quality of jet reconstruction, quality requirements on tracks are applied, based on the following variables:

- the track transverse momentum p_{T};
- the χ^2 coming from the Kalman filter fit of the track;
- the probability P_{ghost} for a track to be an artifact of the pattern recognition and not a real particle;
- the momentum resolution $\frac{\sigma(q/p)}{q/p}$, where q is the particle charge and p is the particle momentum.

Depending on the category of the considered tracks (long, VELO, upstream or downstream), different requirements are used, as shown in Table 3.5. From the PID system, a particle mass and a particle type (between p, e^-, μ^-, π^- and respective antiparticles) are associated to tracks. VELO tracks are included in the jet clustering step as they provide information on the PV of the jets, whereas they do not give information on the particles' momenta.

	long	downstream	upstream	VELO
p_{T}(trk) [MeV/c]	-	-	> 100	-
χ^2	< 5	< 1.5	< 1.5	< 10
P_{ghost}	< 0.5	-	-	-
$\frac{\sigma(q/p)}{q/p}$	> 10	> 10	> 2	-

Table 3.5: List of requirements applied to tracks by the Particle Flow algorithm [81].

Non-isolated neutral particles

The selection of clusters not isolated from tracks follows several steps:

1. requirements on ECAL and HCAL $\chi^2_{track-cluster}$ are applied, where $\chi^2_{track-cluster}$ is the χ^2 of the track-cluster matching: $\chi^2_{track-cluster,ECAL} < 25$ and $\chi^2_{track-cluster,HCAL} < 25$ (< 15) for energies below (above) 10 GeV;
2. clusters are grouped such that different groups do not share tracks, both in the ECAL and HCAL;
3. an empirical parametrization of E/p as a function of p (where E is the cluster energy released by charged particles and p the track momentum) is used in order to compute the expected released energy E^{exp} in the calorimeters;
4. if $E^{exp} > 1.8E^{meas}$ where E^{meas} is the measured energy of the cluster then the cluster is discarded;
5. if otherwise $E^{meas} > 1.8E^{exp}$ then E^{exp} is subtracted from E^{meas} obtaining $E_{subtracted}$;
6. finally if $E_T > 2$ GeV then $E_{subtracted}$ is selected as a non-isolated neutral particle and used as input in the following anti-k_T jet clustering algorithm.

To estimate the energy released by tracks in the calorimeters, a parametrization of E/p is necessary. The procedure to get this parametrization goes by the name of E/p calibration. The calibration has been performed on a 2015 data sample of pp collisions at $\sqrt{s} = 13$ TeV, collected with a minimum bias trigger configuration. Tracks matched to calorimeter clusters are selected, and some requirements are applied to remove background from minimum ionizing particles like muons:

- only on PV in the event;
- long or downstream track;
- given a track, no other tracks within $\Delta R < 0.5$, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where η is the pseudorapidity and ϕ is the azimuthal angle;
- track $p_{\mathrm{T}} > 50$ MeV/c;
- $\chi^2_{\mathrm{track-cluster}} < 1$ for one ECAl or HCAL cluster;
- cluster transverse energy $E_{\mathrm{T}}(\mathrm{cluster}) > 200$ MeV.

Depending on how the tracks and the clusters have been associated with each other, several "track-cluster" categories are identified:

- tracks associated with a cluster in ECAL but not in HCAL;
- tracks associated with a cluster in HCAL but not in ECAL;
- tracks associated with a cluster in ECAL, independently of the presence of an HCAL cluster;
- tracks associated with a cluster in HCAL, independently of the presence of an ECAL cluster;
- tracks associated with an HCAL cluster;
- tracks associated with both an ECAL and an HCAL cluster.

A separate E/p calibration for tracks identified as hadrons or electrons, and for each of the "track-cluster" categories defined above, is performed. The E/p distribution is fitted with empirical functions as a function of p. For all the categories where only one cluster from ECAL or HCAL is considered, the function used is

$$\frac{E}{p}(p) = a_1 e^{-a_2 p} + a_3 \tag{3.6}$$

while for the last category, E is computed as the number of ECAL and HCAL cluster energies, and the function used for the fit is

$$\frac{E}{p}(p) = (a_1 + a_2 p + a_3 p^2) e^{-a_4 p} + a_5 \tag{3.7}$$

Figure 3.7 shows examples of E/p calibration for some categories obtained in data, compared with MC samples of pp collisions at $\sqrt{s} = 13$ TeV with a minimum bias trigger configuration.

Isolated neutral particles

Finally, calorimeters clusters not associated with tracks are stored as inputs for jet clustering. Typical neutral particles identified by the ECAL are photons and π^0 which decay into two photons. π^0 are divided into two categories, depending on how they decay: *merged* π^0, when the two photons are almost collinear and produce a single cluster, and *resolved* π^0, when the two photons are detected as two separate clusters. Some requirements are applied to select ECAL clusters isolated from the tracks and to ensure the quality of the input object before entering the clustering step. The following observables are considered:

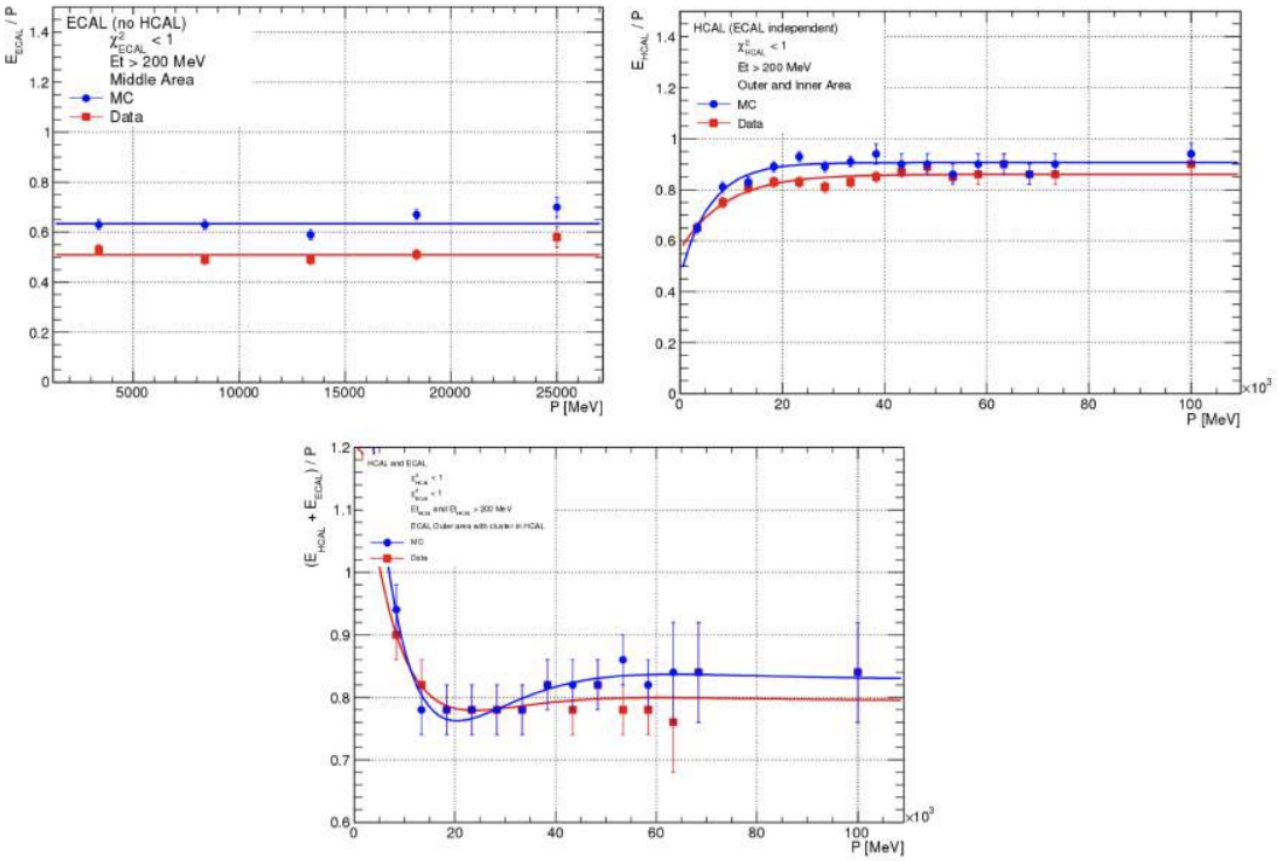


Figure 3.7: E/p calibration for tracks associated with ECAL cluster but not in HCAL (upper left plot), tracks associated with HCAL cluster independently of HCAL (upper right plot) and tracks associated with both ECAL and HCAL cluster (lower plot), for data (red dots) and MC simulations (blue dots) [81].

- the likelihood for the photon hypothesis, `PhotonID`;
- the cluster transverse energy E_{T};
- the $\chi^2_{\mathrm{track-cluster}}$ for each track-cluster combination.

The list of the requirements applied in the ECAL clusters selection is reported in Table 3.6. For HCAL clusters, no particle identification requirements are applied and different $\chi^2_{\mathrm{track-cluster}}$ cuts are applied depending on the cluster energy: $\chi^2_{\mathrm{track-cluster}} > 25$ for HCAL cluster with energy below 10 GeV, and $\chi^2_{\mathrm{track-cluster}} > 15$ for energies above 10 GeV.

	merged π^0	resolved π^0	photons
E_{T} [MeV/c]	-	-	> 200
`PhotonID`	-	> -4	> -1 (> -2 with T track)
`PhotonID` for 1 γ	-	> -2	-
$\chi^2_{\mathrm{track-cluster}} > 15$	> 25	> 25	> 25 (> 16 with T track)

Table 3.6: List of requirements applied to ECAL clusters identified as photons or π^0 [81].

3.3.2 Clustering

Particles selected by the PF algorithms are used as inputs for the jet clustering, the step that indeed creates the jet object. A good jet clustering algorithm has to fulfill two main requirements:

- infrared (IR) safety, meaning that results are finite at all orders of perturbative QCD (due to soft gluon emissions);
- collinear (UV) safety, meaning that jet definition is invariant to soft collinear gluon emissions.

Sequential algorithms that preserve IR and UV safety use ordered clustering. The main steps are the following:

1. for each i and j particles in the listed PF inputs, the following distance d_{ij} is computed

$$d_{ij} = \min(p_{\mathrm{T},i}^{2c}, p_{\mathrm{T},j}^{2c}) \frac{\Delta R_{ij}^2}{R^2}, \tag{3.8}$$

   where $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the distance in the (η, ϕ) plane, and $p_{\mathrm{T},i}$, η_i and ϕ_i are respectively the transverse momentum, the pseudorapidity and the azimuthal angle of the i particle. R is the parameter radius of the jet.

2. for each particle i, the distance to the beam B is calculated as

$$d_{iB} = p_{\mathrm{T},i}^{2c} \tag{3.9}$$

   and the smallest distance between d_{ij} and d_{iB} is found. At this point, there are two possible outcomes:

   (a) if d_{ij} is the smallest distance, the two particles i and j are combined by summing their four-momenta to form another particle l, which substitutes the initial particles and the algorithm goes back to the beginning at the distance evaluation using the new particle l;

   (b) if d_{iB} is the smallest distance, then the particle i is removed from the list, it is defined as a jet and the algorithm starts back;

3. the algorithm finishes when there are no more particles in the list.

Depending on the c exponent in Equation 3.8, different jet clustering algorithms are defined:

- if $c = 0$, the algorithm is the Cambridge/Aachen, that provides ordering by only using distance in the (η, ϕ) plane;
- if $c = 1$, the algorithm is the k_{T}, which orders particles in ascending order. The main consequence of this algorithm is that soft particles tend to cluster among themselves before clustering with a hard particle;
- if $c = -1$, the algorithm is the anti-k_{T} [82], where inputs are ordered in descending p_{T}. Therefore, soft particles will tend to cluster with a hard particle before clustering among themselves.

The differences in the way the three algorithms clusters the same event are shown in Figure 3.8. At LHCb, the anti-k_{T} [82] algorithm is used, given that among the IR and UV safe algorithms is the one not producing irregularities due to soft emissions in the boundaries of jets, as highlighted in Figure 3.8. Besides the choice of the algorithm, the only hyper-parameter that has to be carefully chosen is the clustering radius R, given that this will affect the jet reconstruction performance; indeed, the radius R should be large enough to contain all the QCD radiation, but not so large to avoid contamination from the underlying event. Performance studies in Run 1 for the $Z \to b\bar{b}$ process have shown that the best performance is achieved for $R = 0.5$. This value has been chosen also for Run 2 analyses.

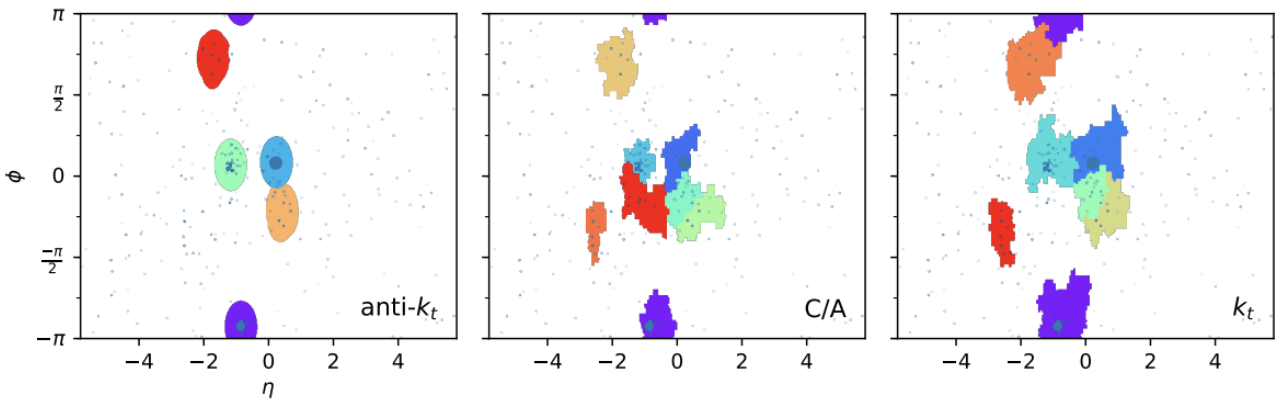


Figure 3.8: Jet clustering with $R = 0.5$ for the same data event using (left) anti-k_T, (center) Cambridge/Aachen and (right) k_T algorithms [83].

3.3.3 Jet 4-momentum computation and MC validation

After the clustering step, the jet four-momentum p_{jet} is determined using the "E-recombination scheme", by computing $p_{\text{jet}} = (E_{\text{jet}}, \vec{p}_{\text{jet}})$ as the sum of the N jet-constituents particles momenta $p_i = (E_i, \vec{p}_i)$:

$$E_{\text{jet}} = \sum_i^N E_i \qquad \vec{p}_{\text{jet}} = \sum_i^N \vec{p}_i \tag{3.10}$$

To validate jet reconstruction, MC simulations are used. In simulated events, the true information of particles generated inside the detector (`MCParticles`) is present, therefore true MC jets (jet_{MC}) are defined as jets clustered with the anti-k_T algorithm, but using all stable `MCParticle` instead of the reconstructed ones. Neutrinos are not considered in the clustering process. In this way, it is possible to evaluate the contribution of detector resolution effects to jet reconstruction. To perform a comparison between jet_{MC} and reconstructed jet (jet_{reco}), they need to be associated to one another; to do so, jet_{MC} are selected by requiring to have $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ with respect to jet_{reco}. If more than one jet_{MC} satisfies this criteria, the jet_{MC} with the smallest ΔR is chosen. The jet particle composition of jet_{MC} and jet_{reco} as measured during Run 1 is shown in Figure 3.9 as a function of jet η. It can be seen that around 60% of jet constituents are charged particles (detected by the tracking system), 30% are π^0 and photons, and 10% are neutral hadrons detected by HCAL.

3.3.4 Jet Energy Correction and Jet Energy Scale

When simulating proton-proton collisions, the energy of reconstructed jets $E(jet_{\text{reco}})$ might differ from the energy of associated MC jets $E(jet_{\text{MC}})$. MC jets are built using as input MC particles, therefore they represent the true result of the jet clustering. To correct the jet reconstructed energy to its true value, a multiplicative correction factor k_{MC} is evaluated in simulation, defined as

$$k_{\text{MC}} = \frac{E(\text{jet}_{\text{MC}})}{E(\text{jet}_{\text{reco}})} \tag{3.11}$$

k_{MC} is the so-called "Jet Energy Correction" factor. Given that the jet_{MC} direction is found to be the same as the jet_{reco}, the same correction factor is applied to each component of the jet four-momentum vector. k_{MC} takes into account the effects of the pile-up, the noise, and the non-uniformity of the detector. To evaluate k_{MC}, simulated events of jets generated by $b-$, $c-$, light (u,d,s) quarks, and gluons at a center-of-mass energy $\sqrt{s} = 13$ TeV have been used. A parametrization of k_{MC} is obtained

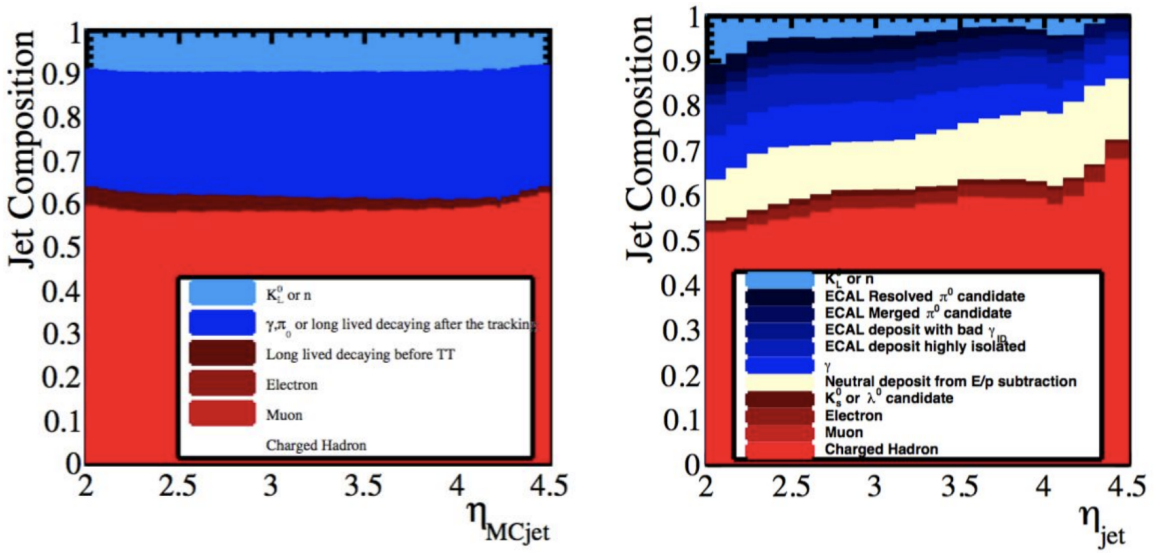


Figure 3.9: Jet particle composition as a function of (left) η of jet_{MC} and (right) η of jet_{reco} as measuring during Run 1 [81].

as a function of jet p_{T}, η, ϕ, the fraction of charged particles in the jet (cpf), and the number of primary vertices (nPV), using a cubic function model. Results obtained using Run 1 simulations are shown in Figure 3.10

3.3.5 Jet identification

Different sources such as detector noise or high energetic lepton might create signatures similar to jets. Therefore, to reduce the impact of so-called *fake jets*, some requirements are applied:

- number of tracks inside the jet pointing to the primary vertex, $\text{nPV}_{\text{trks}} \geq 2$;
- maximum fraction of transverse momentum carried by a single ParticleFlow particle, mpf < 0.8;
- maximum transverse momentum carried by a track, mpt ≥ 1.2 GeV;
- fraction of charged particles on the jet, cpf > 0.1

To measure the quality of the jet identification requirements applied, the jet identification efficiency is evaluated using MC simulations. The jet identification efficiency is defined as the number of correctly reconstructed and identified jets with respect to the true jets. $Z \to \mu\mu$+jet MC events are used to measure the jet identification efficiency. To properly obtain jet identification efficiency in real data, MC events are reweighted such that the reconstructed jet multiplicity distribution in $Z \to \mu\mu$+jet MC events resembles the distribution in data. Weights are obtained as a function of the jet p_{T}. Figure 3.11 shows the jet identification efficiency as a function of jet p_{T} as obtained from Run 1 simulations.

3.4 Jet tagging

"Jet tagging" usually refers to the method used to identify the flavor of the quark (typically a heavy quark such b- or c-quark) that originates the jet. This task is achieved by analyzing the jet properties to take a decision on the jet flavor. At the LHCb experiment, there are different methods to identify the jet flavor [84]:

- Secondary Vertex (SV) tagging: b and c hadrons have a certain lifetime τ and they travel

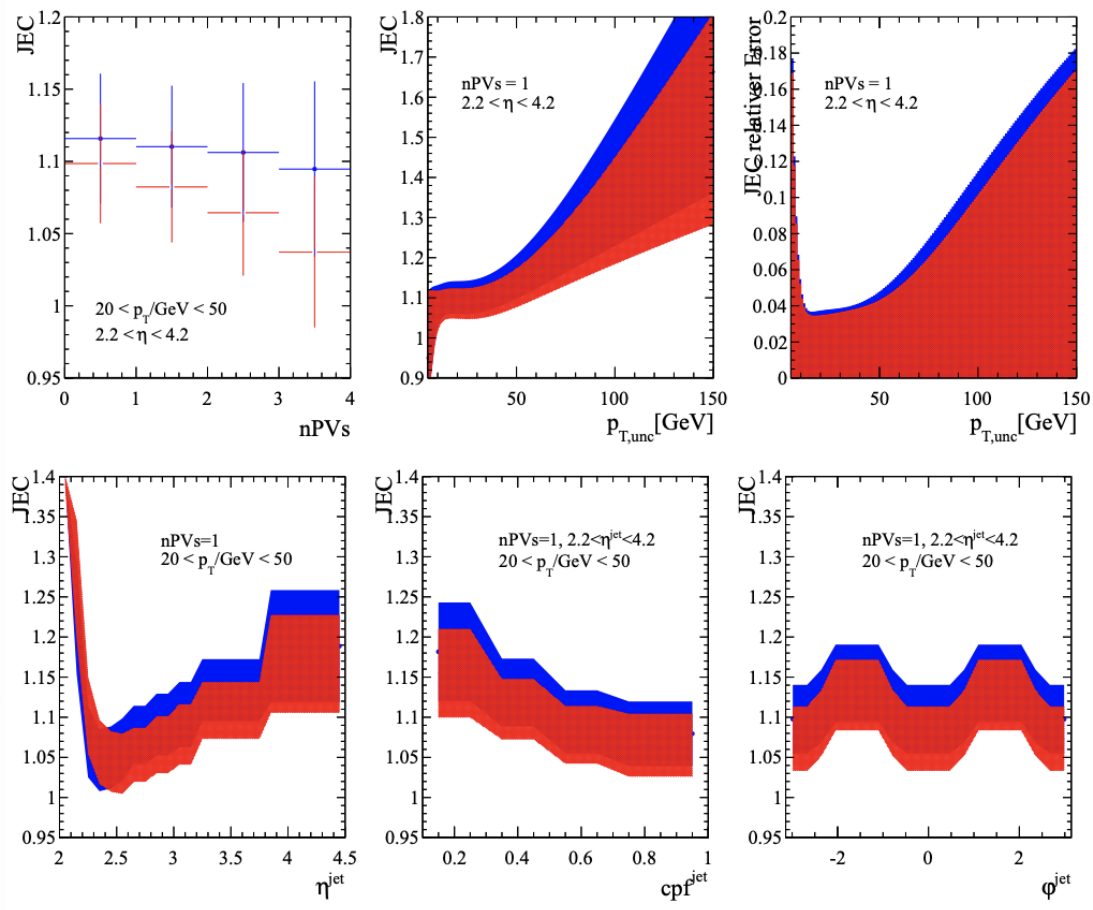


Figure 3.10: k_{MC} correction factor (here denoted as "JEC") as a function of several jet observables, as obtained from Run 1 simulations. Results are shown for different values of the jet radius R, with $R = 0.5$ ($R = 0.7$) for blue (red) curves [81].

a finite distance d inside the detector before decaying. The average distance is computed as $< d >= \gamma c\tau\beta$ where c is the speed of light, $\beta = \frac{v}{c}$, v is the particle velocity and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ is the Lorentz factor. In this way, an SV is generated by the decay of hadrons detached from the PV where the main interaction has occurred. If an SV is found inside a jet, the jet itself can be identified as being originated by a b- or a c-quark, allowing the separation between *heavy flavor* jets (those originated by b and c quarks) from *light* jets (those generated by u, d s quarks, and gluons). The separation between heavy flavor and light jets, and between b- and c-jets is obtained by combining the outputs of two Boosted Decision Trees (BDTs), that uses SV-related observables as inputs. The tagging efficiency (*i.e.* the number of tagged jets over the number of reconstructed jets) for b- and c-jets as a function of jet p_{T} is shown in Figure 3.12; This is the standard tool used to separate between b-, c- and light jets;

- jet-charge tagging: an effective jet charge can be defined as

$$Q = \frac{\sum_i (p_{\mathrm{T}}^{\alpha})_i q_i}{\sum_i (p_{\mathrm{T}}^{\alpha})_i} \ , \tag{3.12}$$

where p_{T} and q_i are respectively the transverse momentum and the charge of the particles inside the jet, and α is a parameter that can be tuned. Q can be used as an estimator of the charge of the quark generating the jet;

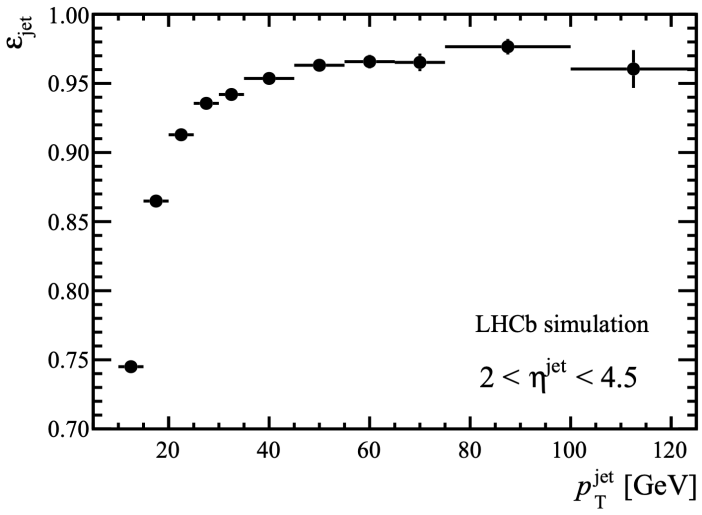


Figure 3.11: Jet identification efficiency as a function of the true jet p_T as obtained from Run 1 simulations [81].

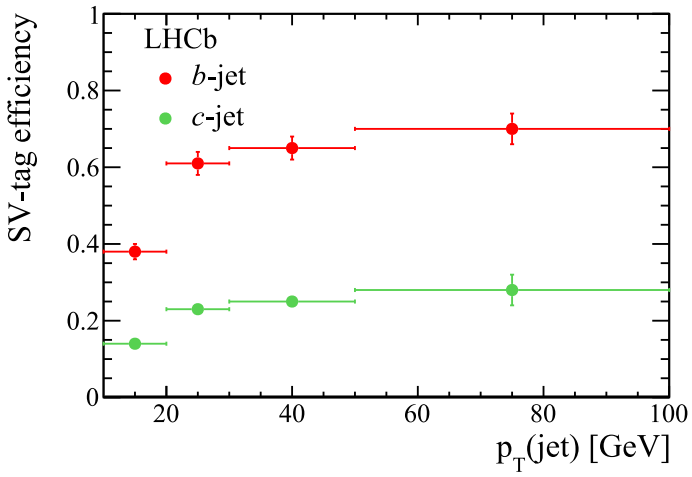


Figure 3.12: Jet tagging efficiency for b- and c-jets as a function of the true jet p_T as obtained from Run 1 simulations [84].

- lepton-charge tagging: B and D hadrons may decay semileptonically, generating highly energetic leptons. This algorithm searches for a lepton inside the jet, satisfying some kinematical requirements. If a lepton is found, the lepton charge can be used to infer the charge of the quark generating the jet. This tagging method has been used at the LHCb experiment to measure the $b\bar{b}$ asymmetry using Run 1 data [85], by identifying muons coming from the semileptonic decay of B meson. Possible improvements coming from new tagging techniques are studied in Chapter 7.

Chapter 4

Measurement of W+jets

In this chapter, measurements of W bosons cross-section produced in association with jets and with W decaying to a muon and a neutrino are presented. Particularly, differential cross-section measurements as a function of the muon pseudorapidity, and the leading jet p_{T} and pseudorapidity are performed. The full Run 2 dataset is considered, corresponding to an integrated luminosity of approximately 5.4 fb^{-1}. The muon channel is chosen given the optimal performance in reconstructing muons, as described in Chapter 3, and to reduce physics backgrounds. The electron channel is not considered due to the saturation effects of the ECAL, already mentioned in Chapter 2 and better assessed in Chapter 6. The Chapter is structured in this way:

- Section 4.1 briefly presents the analysis strategy;
- Section 4.2 presents the requirements to select events for the analysis;
- Sections 4.3 and 4.4 deal with the signal section and the determination of the backgrounds;
- Section 4.5 describes the fit procedure to extract W+ jets yields;
- Sections 4.6, 4.7, 4.8 and 4.9 deal respectively with the evaluation of reconstruction efficiencies, the acceptance factors, the unfolding procedure and the systematic uncertainties related to the measurements;
- finally, Sections 4.10 and 4.11 present respectively the results of the analysis and some final considerations for the future prospects.

4.1 Analysis strategy

This Chapter reports the first measurement of the W+ jets process at a center-of-mass energy $\sqrt{s} = 13$ TeV in the forward region. The measurement is performed in the fiducial region defined by the kinematic requirements listed in Table 4.1. These requirements are applied to truth-level particles. A truth-level jet is defined as a jet composed of all stable particles generated by MC simulations.

The analysis strategy can be summarised in the following steps:

- signal sample is generated using MC simulations and analysis cuts are applied;
- background samples describing EW processes are generated using MC simulations, while the QCD background is evaluated with a data-driven technique;

Object	Fiducial Cut
Muons	$p_{\mathrm{T}} > 20$ GeV $2.0 < \eta < 4.5$
Jets	$p_{\mathrm{T}} > 20$ GeV $2.2 < \eta < 4.2$
Event	$p_{\mathrm{T}}(\mu + \mathrm{jets}) > 20$ GeV

Table 4.1: Fiducial cuts applied to the final state objects in the $W+$ jets analysis.

- a fit to the muon isolation I_μ is performed, with

$$I_\mu = \frac{p_{\mathrm{T}}(\mu)}{p_{\mathrm{T}}(\mu^{\mathrm{jet}})} \tag{4.1}$$

  where $p_{\mathrm{T}}(\mu)$ is the transverse momentum of the muon and $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ is the transverse momentum of a jet of radius $R = 0.4$ surrounding the muon. From the I_μ fit, the $W+$ jets yield is extracted.

- reconstruction and selection efficiencies, together with acceptance factors, are evaluated using MC samples, and they are applied to the $W+$ jets yield to correct for detector effects and unfolding;

- finally, the cross-section for $W+$ jets is evaluated in the full fiducial region and differentially as a function of the muon pseudorapidity η^μ, the leading jet transverse momentum $p_{\mathrm{T}}^{\mathrm{jet}}$ and the leading jet pseudorapidity η^{jet}. Experimental results are compared with theoretical predictions.

A flowchart of the analysis strategy followed in this Chapter is shown in Figure 4.1.

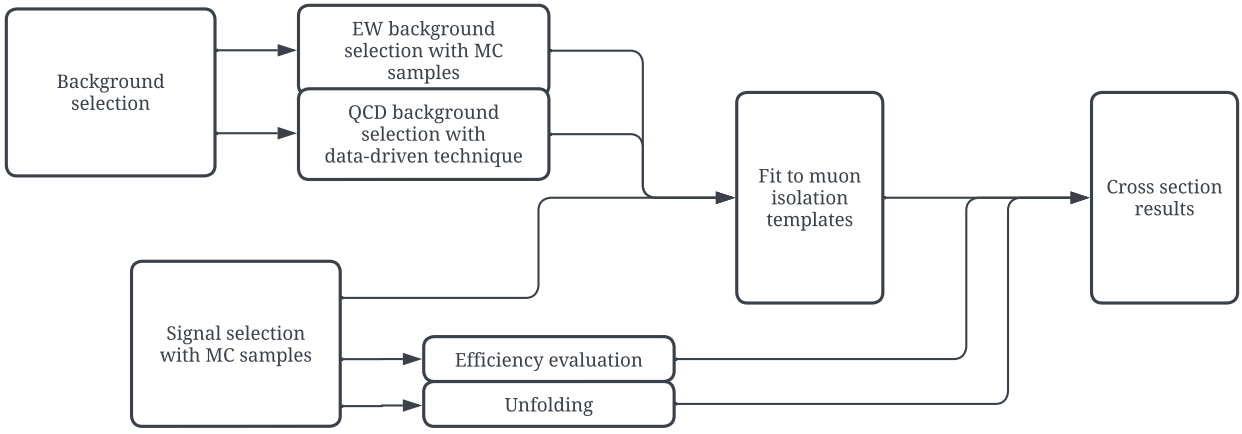


Figure 4.1: Flowchart of the $W+$ jets analysis, highlighting the main steps of the analysis.

4.2 Dataset selection

The signature for this analysis is a highly energetic muon with at least N jets associated, with $N = 1, 2$. At the trigger level, a set of minimum requirements are considered for the muon. Standard trigger requirements used in several EW analyses [44, 45, 47, 48, 86, 87] at the LHCb are used, and they are listed in Table 4.2, for both hardware and software trigger stages.

There are a few motivations for the application of these requirements:

Trigger line	Selection
`L0MuonEW`	nSPD<10000
	$p_{\mathrm{T}} > 6.0$ GeV
`Hlt1SingleMuonHighPT`	$p > 8.0$ GeV
	$p_{\mathrm{T}} > 6.0$ GeV
	Track $\chi^2 < 4.0$
`Hlt2SingleMuonHighPT`	IP< 0.25 mm
	IP $\chi^2 < 100$
	$p_{\mathrm{T}} > 15.0$ GeV

Table 4.2: Hardware and software trigger requirements on muons for the $W+$ jets analysis, with the list of the selection cuts applied.

- the muon is required to have a minimum p_{T} and p to avoid contamination from background low energetic muons. Muons coming from EW backgrounds have typically high p_{T} with respect to muons produced from semileptonic decays;
- the number of hits in the SPD (nSPD) is lower than a certain threshold to discriminate between muons and hadrons;
- the impact parameter of the muon has to be close to 0, in order to select a muon coming from the primary vertex. This allows the rejection of muons coming from τ leptons or hadrons decay, which tend to have higher IP;
- both the IP and the track have to be well reconstructed, therefore the cuts on the χ^2.

Events that pass the requirements at the trigger level are afterward passed through some selection cuts, to further enhance the signal contribution with respect to background ones. Kinematic selection requirements on p_{T} and η are applied to the reconstructed muon and jets, as listed in Table 4.3. These requirements are denoted as "analysis cuts", and closely match the fiducial region cuts defined in Table 4.1.

Object	Selection
Muon	$p_{\mathrm{T}} > 20$ GeV
	$2.0 < \eta < 4.5$
	$(E_{\mathrm{ECAL}} + E_{\mathrm{HCAL}})/p < 0.04$
	IP< 0.04 mm
Jets	$p_{\mathrm{T}} > 20$ GeV
	$2.2 < \eta < 4.2$
Event	$\Delta R(\mu, \mathrm{jets}) > 0.5$

Table 4.3: Kinematic requirements at the analysis level to further increase (decrease) the signal (background) contribution.

These analysis cuts are considered for the following reasons:

- muons coming from background processes and semileptonic decays are removed by requiring a cut on the p_{T}, while the cut on the muon pseudorapidity is required to ensure that the muon is reconstructed inside the LHCb acceptance;
- the cut on $(E_{\mathrm{ECAL}} + E_{\mathrm{HCAL}})/p$, where E_{ECAL} (E_{HCAL}) is the energy released by the muon in the ECAL (HCAL) and p is its momentum, is required to reject hadrons that pass through the calorimeter system (leaving there an energy deposit) and produce a signal in the muon chambers;

- muons are required to have a lower IP (*i.e.* to be *prompt* muons), in order to reject muons coming from decay of hadrons or τ leptons, which tend to have higher IP as they are more displaced from the PV;
- jets are required to be in the instrumented part of the calorimeters, where the reconstruction efficiency as a function of η^{jet} is almost flat. The cut of the jet η ensure this, while the cut on the jet p_{T} allows to have higher jet identification and reconstruction efficiencies;
- finally, muons are required to be spatially separated from jets, to ensure that selected muons do not come from processes happening inside the jet. The distance is defined as $\Delta R = \sqrt{\Delta\eta^2(\mu, j) + \Delta\phi^2(\mu, j)}$, where $\Delta\eta(\mu, j)$ ($\Delta\phi(\mu, j)$) is the difference in pseudorapidity (azimuthal angle) between the muon and the jets.

In addition to these kinematic requirements, two more requirements are considered at the analysis stage:

- multiple candidates removal: once the μ+jet(s) signature is considered, several event candidates are selected, depending on how many muons and jets pass the reconstruction and selection requirements. In order to properly select one candidate per event, the candidate with the highest $p_{\text{T}}(\mu) + p_{\text{T}}(\text{jets})$ is selected, to ensure the selection of muons and jets coming from the signal process;
- $Z \to \mu\mu$ veto: among the EW processes selected by this analysis, $Z(\to \mu\mu)$ + jet(s) is the highest background to W+jet(s). In order to reduce events where just a muon from $Z \to \mu\mu$ is reconstructed, events containing another high-p_{T} muon with $p_{\text{T}} > 20$ GeV and such that the invariant mass of the di-muon system is in the range $60 < M(\mu\mu) < 120$ GeV are rejected.

In Table 4.4, the cumulative selection efficiencies for the $W + 1$ jet and $W + 2$ jets selection are shown, for 2016, 2017, and 2018 datasets. The cumulative selection efficiencies are evaluated starting from the reconstructed events, and for a given requirement x the efficiency is evaluated as

$$\varepsilon_x = \frac{N_x}{N_{x-1}} \tag{4.2}$$

where N_x is the number of events passing the x requirement and N_{x-1} is the number of events before the x requirement. The expected number of events N_{expected} is evaluated as

$$N_{\text{expected}} = N_{\text{reconstructed}} \cdot \varepsilon_{\text{sel}} \tag{4.3}$$

where $N_{\text{reconstructed}}$ is the number of reconstructed events and ε_{sel} is the selection efficiency.

4.3 Signal selection

Signal samples for $W+$ 1 jets and $W+$ 2 jets are obtained from MC simulations. The following processes are simulated, requiring some generator level cuts (*i.e.* cuts applied to objects before the detector simulation):

- $W(\to \mu^{\pm}\nu_{\mu})$+1 jet, simulated at LO with Pythia8, requiring one leading parton coming from the proton-proton interaction, and with the muon from the W boson to have $p_{\text{T}} > 17$ GeV and $\theta_{\mu} < 400$ mrad;
- $W(\to \mu^{\pm}\nu_{\mu})$+2 jets, simulated at LO with MadGraph [26, 27], requiring the muon to have $p_{\text{T}} > 10$ GeV and $1.5 < \eta < 5$ and the two jets with $p_{\text{T}} > 10$ GeV and $\eta < 5$;

		2016	2017	2018
$W+1$ jet	L0	63.47 ± 0.01	66.09 ± 0.01	65.66 ± 0.01
	Hlt1	51.24 ± 0.01	52.90 ± 0.01	52.34 ± 0.01
	Hlt2	51.24 ± 0.01	52.90 ± 0.01	52.34 ± 0.01
	IP< 0.04 mm	31.51 ± 0.01	33.67 ± 0.01	33.40 ± 0.01
	$(E_{\text{ECAL}} + E_{\text{HCAL}})/p < 0.04$	13.44 ± 0.01	14.157 ± 0.009	13.964 ± 0.008
	$\chi^2/\text{ndof} < 2.5$	13.34 ± 0.01	14.065 ± 0.009	13.879 ± 0.008
	$P(\chi^2) > 0.01$	12.51 ± 0.01	13.239 ± 0.009	13.126 ± 0.007
	$p_{\text{T}}(j) > 20$ GeV	7.509 ± 0.007	7.974 ± 0.007	7.915 ± 0.006
	$2.2 < \eta(j) < 4.2$	5.958 ± 0.006	6.335 ± 0.006	6.273 ± 0.005
	$p_{\text{T}}(\mu) > 20$ GeV	5.958 ± 0.006	6.335 ± 0.006	6.273 ± 0.005
	$2.0 < \eta(\mu) < 4.5$	5.369 ± 0.006	5.679 ± 0.006	5.588 ± 0.005
	multiple candidates	4.936 ± 0.006	5.205 ± 0.006	5.161 ± 0.005
	Z veto	4.857 ± 0.006	5.124 ± 0.006	5.088 ± 0.005
	$\Delta R(\mu, j) > 0.5$	4.783 ± 0.006	5.049 ± 0.006	5.015 ± 0.005
	total selected events	1406688	1455366	1975770
$W+2$ jets	L0	61.36 ± 0.02	64.46 ± 0.02	63.48 ± 0.02
	Hlt1	45.58 ± 0.02	48.22 ± 0.02	47.21 ± 0.02
	Hlt2	45.58 ± 0.02	48.22 ± 0.02	47.20 ± 0.02
	IP< 0.04 mm	26.73 ± 0.02	29.27 ± 0.02	28.75 ± 0.01
	$(E_{\text{ECAL}} + E_{\text{HCAL}})/p < 0.04$	10.01 ± 0.01	10.77 ± 0.01	10.49 ± 0.01
	$\chi^2/\text{ndof} < 2.5$	9.92 ± 0.01	10.69 ± 0.01	10.42 ± 0.01
	$P(\chi^2) > 0.01$	9.25 ± 0.01	9.98 ± 0.01	9.78 ± 0.01
	$p_{\text{T}}(j_0) > 20$ GeV	5.47 ± 0.01	5.88 ± 0.01	5.78 ± 0.01
	$2.2 < \eta(j_0) < 4.2$	4.37 ± 0.01	4.71 ± 0.01	4.62 ± 0.01
	$p_{\text{T}}(j_1) > 20$ GeV	2.33 ± 0.01	2.55 ± 0.01	2.48 ± 0.01
	$2.2 < \eta(j_1) < 4.2$	1.84 ± 0.01	2.02 ± 0.01	1.97 ± 0.01
	$p_{\text{T}}(\mu) > 20$ GeV	1.84 ± 0.01	2.02 ± 0.01	1.97 ± 0.01
	$2.0 < \eta(\mu) < 4.5$	1.65 ± 0.01	1.80 ± 0.01	1.75 ± 0.01
	$\Delta R(\mu, j_{0,1}) > 0.5$	1.46 ± 0.01	1.57 ± 0.01	1.54 ± 0.01
	multiple candidates	1.44 ± 0.01	1.48 ± 0.01	1.45 ± 0.01
	Z veto	1.38 ± 0.01	1.42 ± 0.01	1.39 ± 0.01
	total selected events	142214	142782	168014

Table 4.4: Cumulative selection efficiencies and total selected events in 2016, 2017, and 2018 datasets.

Pythia8 is used for parton showering, and the CT09MCS PDF set is used [51]. Particle-matter interactions in the detector are simulated using Geant4 [71, 72]. The full simulation framework of the LHCb experiment is used, as described in Section 2.2.9.

Typical quantities to assess the number of expected events for a specific physics processes are the generator efficiency ε_{gen} and the reconstruction efficiency $\varepsilon_{\text{reco}}$, defined as

$$\varepsilon_{\text{gen}} = \frac{N_{acc}}{N_{gen}} \qquad \varepsilon_{\text{reco}} = \frac{N_{reco}}{N_{acc}} \tag{4.4}$$

where N_{gen} is the number of generated events in the MC simulation, N_{acc} is the number of accepted events that pass the generator level cuts and N_{reco} is the number of events that have been reconstructed after the interaction with the detector. Table 4.5 shows the expected cross-section, ε_{gen}, and $\varepsilon_{\text{reco}}$ for $W+1$ jet and $W+2$ jets events.

Process	σ (pb)	ε_{gen}	$\varepsilon_{\text{reco}}(\%)$
$W(\to \mu\nu)+$ 1 jet	22640	0.26	15.97 ± 0.02
$W(\to \mu\nu)+$ 2 jets	1231.3	0.44	21.83 ± 0.01

Table 4.5: Expected cross-section, ε_{gen}, and $\varepsilon_{\text{reco}}$ for $W + 1$ jet and $W + 2$ jets events

Finally, the selection requirements described in Section 4.2 are applied to the $W+1$ jet and $W+2$ jets reconstructed events. The cumulative efficiencies are shown in Table 4.6, together with the expected $W + 1$ jet and $W + 2$ jets events for an integrated luminosity $\mathcal{L} = 5.4$ fb^{-1}.

	$W(\to \mu\nu)+$ 1 jet	$W(\to \mu\nu)+$ 2 jets
`L0`	87.53 ± 0.05	87.57 ± 0.05
`Hlt1`	79.24 ± 0.06	79.22 ± 0.06
`Hlt2`	79.24 ± 0.06	79.22 ± 0.07
IP<0.04 mm	75.58 ± 0.07	77.85 ± 0.07
$E/p < 0.04$	69.93 ± 0.07	70.78 ± 0.07
$\chi^2/\text{ndof} < 2.5$	69.90 ± 0.07	70.75 ± 0.07
$P(\chi^2) > 0.01$	69.17 ± 0.07	70.10 ± 0.07
$p_{\text{T}}(j_0) > 20$	33.39 ± 0.07	47.38 ± 0.06
$2.2 < \eta(j_0) < 4.2$	26.36 ± 0.07	37.32 ± 0.06
$p_{\text{T}}(j_1) > 20$	/	22.99 ± 0.06
$2.2 < \eta(j_1) < 4.2$	/	17.62 ± 0.06
$p_{\text{T}}(\mu) > 20$ GeV	26.36 ± 0.06	17.62 ± 0.06
$2.0 < \eta < 4.5$	24.92 ± 0.06	16.70 ± 0.06
multiple candidates	23.28 ± 0.06	14.01 ± 0.06
Z veto	23.28 ± 0.06	14.01 ± 0.06
$\Delta R(\mu, j_{0,1}) > 0.5$	22.91 ± 0.06	13.42 ± 0.06
expected events	1162983 ± 3045	85685 ± 383

Table 4.6: Cumulative selection efficiencies for $W + 1$ jet and $W + 2$ jets events.

4.4 Background determination

The main backgrounds for this analysis are divided into two categories, treated in two different ways: events coming from EW processes are obtained from MC simulations, while events coming from QCD processes are obtained from a data-driven approach.

4.4.1 EW background determination

The muon isolation I_μ distributions (usually referred to as "template") for EW background processes have been obtained using MC simulations. EW background contributions are mainly coming from Z boson production in association with jets, $t\bar{t}$ and single top production, and vector pair boson production, namely WZ and WW. ZZ production is found to be negligible and therefore it is not included in the analysis. For the $W+$ 1 jet selection, the following processes are simulated:

- $Z(\to \mu\mu)+1$ jet, simulated with Pythia8, requiring one leading parton coming from the proton-proton interaction, and with two muons from the Z boson to have $p_{\text{T}} > 17$ GeV and $\theta_\mu < 400$ mrad;
- $W(\to \tau^{\pm}\nu_\tau)+1$ jet, simulated with Pythia8, requiring one leading parton coming from the

proton-proton interaction, and with the tau lepton from the W boson to have $p_{\mathrm{T}} > 4$ GeV and $\theta_\mu < 400$ mrad. The τ lepton decay is performed by the TAUOLA [88] package;

- $Z(\to \tau\tau)$+1 jet, simulated with PYTHIA8, requiring one leading parton coming from the proton-proton interaction, and with two tau leptons from the Z boson to have $p_{\mathrm{T}} > 4$ GeV and $\theta_\mu < 400$ mrad. The τ lepton decay is performed by the TAUOLA package;
- $gg \to t\bar{t}$, simulated with PYTHIA8, requiring the two top quarks to decay into b-quarks and W bosons, and with the muon from the W boson to have $p_{\mathrm{T}} > 17$ GeV and $\theta_\mu < 400$ mrad;
- $q\bar{q} \to t\bar{t}$, simulated with PYTHIA8, requiring the two top quarks to decay into b-quarks and W bosons, and with the muon from the W boson to have $p_{\mathrm{T}} > 17$ GeV and $\theta_\mu < 400$ mrad;
- single top, simulated with PYTHIA8, requiring the top quark to decay into b-quark and W boson, and with the muon from the W boson to have $p_{\mathrm{T}} > 17$ GeV and $\theta_\mu < 400$ mrad;
- WZ, simulated with PYTHIA8, requiring one muon coming from W or Z bosons to have $p_{\mathrm{T}} > 15$ GeV and $\theta_\mu < 400$ mrad;
- WW, simulated with PYTHIA8, requiring one muon coming from W bosons to have $p_{\mathrm{T}} > 15$ GeV and $\theta_\mu < 400$ mrad.

For the W+ 2 jets selection, $W(\to \tau\tau)$ and Z boson production in association with at least two jets are simulated using MADGRAPH. The following processes are simulated:

- $Z(\to \mu\mu)$+2 jets, requiring the two muons to have $p_{\mathrm{T}} > 10$ GeV and $1.5 < \eta < 5$ and the two jets with $p_{\mathrm{T}} > 10$ GeV and $\eta < 5$;
- $W(\to \tau^{\pm}\nu_\tau)$+2 jets, requiring the τ lepton to have a $p_{\mathrm{T}} > 10$ GeV and $1.5 < \eta < 5$ and the two jets with $p_{\mathrm{T}} > 10$ GeV and $\eta < 5$. The τ lepton decay is performed by the TAUOLA package;
- $Z(\to \tau\tau)$+2 jets, requiring the two τ leptons to have a $p_{\mathrm{T}} > 10$ GeV and $1.5 < \eta < 5$ and the two jets with $p_{\mathrm{T}} > 10$ GeV and $\eta < 5$. The τ lepton decay is performed by the TAUOLA package;

As for the signal simulation, PYTHIA8 is used for parton showering, and the CT09MCS PDF set is used. Particle-matter interactions in the detector are simulated using GEANT4[71, 72]. The full simulation framework of the LHCb experiment is used, as described in Section 2.2.9.

Typical LO Feynman diagrams for EW background processes are shown in Figure 4.2.

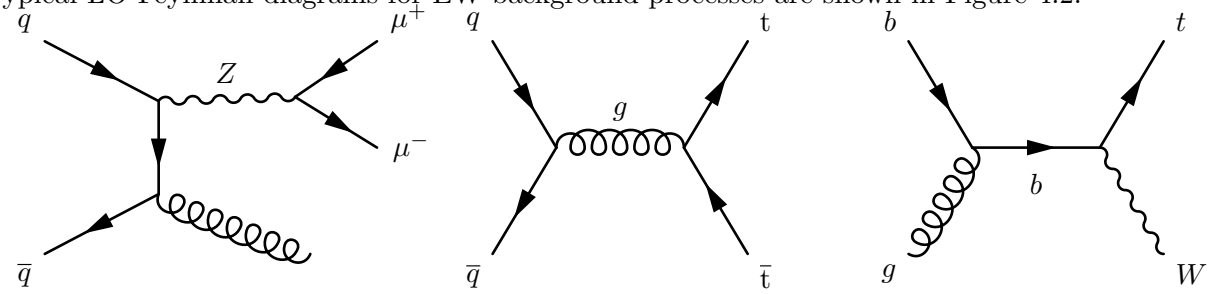


Figure 4.2: LO Feynman diagrams for several EW backgrounds: (left) Z boson production in association with one jet, (center) $t\bar{t}$ production, (right) single top production through s-channel.

Table 4.7 shows the cross-sections obtained by PYTHIA8 and MADGRAPH, the generator level efficiency $\varepsilon_{\mathrm{gen}}$ and the reconstruction efficiency $\varepsilon_{\mathrm{reco}}$ for the EW backgrounds.

In Tables 4.8 and 4.9 the cumulative selection efficiencies and the number of expected events N_{expected} are presented for all the physics processes considered, both for the W+ 1 jets and W+ 2 jets selection.

Process	σ (pb)	ε_{gen}	$\varepsilon_{\text{reco}}(\%)$	Process	σ (pb)	ε_{gen}	$\varepsilon_{\text{reco}}(\%)$
$Z+$ 1 jet	2240	0.32	24.76 ± 0.02	$Z+$ 2 jets	142.1	0.47	49.79 ± 0.06
$gg \to t\bar{t}$	451.9	0.07	30.57 ± 0.02	$gg \to t\bar{t}$	451.9	0.07	22.74 ± 0.02
$q\bar{q} \to t\bar{t}$	77.3	0.09	45.45 ± 0.02	$q\bar{q} \to t\bar{t}$	77.3	0.09	38.21 ± 0.02
$W(\to \tau\nu)$	22750	0.29	0.714 ± 0.006	$W(\to \tau\nu)$	1231.3	0.42	1.41 ± 0.03
$Z(\to \tau\tau)$	2232	0.37	2.450 ± 0.009	$Z(\to \tau\tau)$	142.1	0.49	6.33 ± 0.03
WZ	24.6	0.06	30.01 ± 0.03	WZ	24.6	0.06	13.42 ± 0.03
WW	67.2	0.08	32.04 ± 0.04	WW	67.2	0.08	14.73 ± 0.05
single top	153.18	0.04	29.35 ± 0.04	single top	153.18	0.04	14.82 ± 0.03

Table 4.7: cross-section computed at LO with PYTHIA8 [29], generator level efficiency ε_{gen}, and reconstruction efficiency $\varepsilon_{\text{reco}}$ for the EW backgrounds for the $W+$ 1 jet selection (left) and for the $W+$ 2 jets selection (right).

	$Z(\to \mu\mu)$	$t\bar{t}$	$W(\to \tau\nu)$	$Z(\to \tau\tau)$	WZ	WW	single top
`L0`	87.29 ± 0.04	88.55 ± 0.03	85.7 ± 0.4	86.1 ± 0.2	88.09 ± 0.05	88.1 ± 0.07	88.74 ± 0.07
`Hlt1`	78.69 ± 0.02	77.79 ± 0.04	78.1 ± 0.5	78.4 ± 0.2	78.92 ± 0.07	79.1 ± 0.09	78.19 ± 0.09
`Hlt2`	78.69 ± 0.02	77.79 ± 0.04	78.1 ± 0.5	78.4 ± 0.2	78.91 ± 0.07	79.1 ± 0.09	78.19 ± 0.09
IP<0.04 mm	75.03 ± 0.06	70.24 ± 0.04	46.9 ± 0.6	44.2 ± 0.3	76.34 ± 0.07	76.6 ± 0.09	71.9 ± 0.1
$E/p < 0.04$	69.99 ± 0.06	63.51 ± 0.04	43.1 ± 0.6	40.2 ± 0.3	70.67 ± 0.08	70.8 ± 0.1	65.6 ± 0.1
$\chi^2/\text{ndof} < 2.5$	69.99 ± 0.06	63.48 ± 0.04	43.1 ± 0.6	40.2 ± 0.3	70.64 ± 0.08	70.8 ± 0.1	65.6 ± 0.1
$P(\chi^2) > 0.01$	69.99 ± 0.06	62.88 ± 0.04	42.6 ± 0.6	39.8 ± 0.3	69.98 ± 0.08	70.1 ± 0.1	64.9 ± 0.1
$p_{\text{T}}(j) > 20$ GeV	37.52 ± 0.06	53.04 ± 0.04	22.3 ± 0.5	28.8 ± 0.2	50.99 ± 0.09	51.8 ± 0.1	53.5 ± 0.1
$2.2 < \eta(j) < 4.2$	29.84 ± 0.06	38.27 ± 0.04	17.6 ± 0.4	23.3 ± 0.2	39.22 ± 0.08	40.9 ± 0.1	41.8 ± 0.1
$p_{\text{T}}(\mu) > 20$ GeV	29.84 ± 0.06	38.27 ± 0.04	17.6 ± 0.4	23.3 ± 0.2	39.22 ± 0.08	40.9 ± 0.1	41.8 ± 0.1
$2.0 < \eta < 4.5$	28.11 ± 0.06	34.67 ± 0.04	16.8 ± 0.4	22.3 ± 0.2	36.10 ± 0.08	38.2 ± 0.1	37.5 ± 0.1
multi. cand.	21.50 ± 0.05	20.55 ± 0.03	15.7 ± 0.4	20.3 ± 0.2	25.37 ± 0.07	28.5 ± 0.1	27.7 ± 0.1
Z veto	12.29 ± 0.04	20.06 ± 0.03	15.7 ± 0.4	20.3 ± 0.2	21.51 ± 0.07	28.4 ± 0.1	27.6 ± 0.1
$\Delta R(\mu, j) > 0.5$	12.19 ± 0.04	19.60 ± 0.03	15.6 ± 0.4	20.2 ± 0.2	21.16 ± 0.07	27.8 ± 0.1	27.2 ± 0.1

Table 4.8: Cumulative efficiencies for EW backgrounds for the $W+$ 1 jet selection.

	$Z(\to \mu\mu)$	$t\bar{t}$	$W(\to \tau\nu)$	$Z(\to \tau\tau)$	WZ	WW	single top
`L0`	87.80 ± 0.03	88.34 ± 0.02	83.7 ± 0.1	85.6 ± 0.1	88.1 ± 0.1	87.9 ± 0.1	88.5 ± 0.1
`Hlt1`	79.61 ± 0.03	77.69 ± 0.03	75.1 ± 0.1	77.3 ± 0.1	78.6 ± 0.1	78.7 ± 0.1	77.9 ± 0.1
`Hlt2`	79.61 ± 0.03	77.69 ± 0.03	75.1 ± 0.1	77.3 ± 0.1	78.6 ± 0.1	78.7 ± 0.1	77.9 ± 0.1
IP<0.04 mm	77.76 ± 0.03	70.25 ± 0.04	53.1 ± 0.1	43.5 ± 0.1	76.5 ± 0.1	76.7 ± 0.1	71.9 ± 0.1
$E/p < 0.04$	72.17 ± 0.03	63.49 ± 0.04	48.7 ± 0.1	39.6 ± 0.1	70.8 ± 0.1	70.9 ± 0.1	65.5 ± 0.1
$\chi^2/\text{ndof} < 2.5$	72.14 ± 0.03	63.45 ± 0.04	48.7 ± 0.1	39.3 ± 0.1	70.7 ± 0.1	70.8 ± 0.1	65.5 ± 0.1
$P(\chi^2) > 0.01$	71.43 ± 0.03	62.82 ± 0.04	48.1 ± 0.1	38.9 ± 0.1	70.1 ± 0.1	70.1 ± 0.1	64.8 ± 0.1
$p_{\text{T}}(j_0) > 20$	50.81 ± 0.03	54.84 ± 0.04	34.7 ± 0.1	28.4 ± 0.1	54.6 ± 0.1	54.5 ± 0.1	53.7 ± 0.1
$2.2 < \eta(j_0) < 4.2$	40.43 ± 0.03	40.77 ± 0.04	25.6 ± 0.1	23.2 ± 0.1	43.5 ± 0.1	44.2 ± 0.1	42.5 ± 0.1
$p_{\text{T}}(j_1) > 20$	26.30 ± 0.02	32.38 ± 0.04	16.1 ± 0.1	14.9 ± 0.1	29.6 ± 0.1	29.9 ± 0.1	30.5 ± 0.1
$2.2 < \eta(j_1) < 4.2$	20.59 ± 0.02	24.13 ± 0.03	10.5 ± 0.1	11.6 ± 0.1	23.7 ± 0.1	24.5 ± 0.1	23.7 ± 0.1
$p_{\text{T}}(\mu) > 20$ GeV	20.59 ± 0.02	24.13 ± 0.03	10.5 ± 0.1	11.6 ± 0.1	23.7 ± 0.1	24.5 ± 0.1	23.7 ± 0.1
$2.0 < \eta < 4.5$	19.48 ± 0.02	22.04 ± 0.03	10.1 ± 0.1	11.1 ± 0.1	21.6 ± 0.1	22.9 ± 0.1	21.5 ± 0.1
mult. cand.	12.26 ± 0.02	11.47 ± 0.03	8.2 ± 0.1	8.6 ± 0.1	16.2 ± 0.1	18.8 ± 0.1	17.3 ± 0.1
Z veto	4.90 ± 0.01	11.23 ± 0.02	8.2 ± 0.1	8.58 ± 0.1	13.9 ± 0.1	18.7 ± 0.1	17.2 ± 0.1
$\Delta R(\mu, j_{0,1}) > 0.5$	4.73 ± 0.01	10.80 ± 0.03	8.2 ± 0.1	8.4 ± 0.1	13.4 ± 0.1	17.9 ± 0.1	16.7 ± 0.1

Table 4.9: Cumulative efficiencies for EW backgrounds for the $W+$ 2 jets selection.

MC simulations are used to extract relevant distributions and reconstruction and selection efficiencies. The evaluation of the expected number of events for the EW background processes is done in

two different ways, depending on the considered background: with a data-driven technique, and by normalizing to NLO cross-sections.

$Z(\to \mu\mu)+$ jets estimation

The main background coming from EW processes is $Z+$ jets. The evaluation of this background is done with a data-driven technique: in this way, the normalization does not rely on the LO estimation used to obtain the templates, and therefore additional contributions at NLO are automatically included. $Z+$ jets events are selected in data requiring two muons and at least one jet (or two jets) in the LHCb acceptance. At least one muon is required to pass the trigger requirements listed in Table 4.2, while both muons and jets are required to have $p_{\mathrm{T}} > 20$ GeV and to be spatially separated requiring $\Delta R(\mu, \mathrm{jets}) > 0.5$. The standard pseudorapidity requirement for the muon (jets) is $2.0 < \eta(\mu) < 4.5$ ($2.2 < \eta(\mathrm{jets}) < 4.2$). In order to properly select muons coming from the decay of the Z boson, the invariant mass of the di-muon system $M_{\mu\mu}$ is required to be in the range $60 < M_{\mu\mu} < 120$ GeV. The same selection requirements are applied to simulated $Z+$ jets events, using MC samples described in Section 4.4. The full Run 2 dataset is considered, corresponding to an integrated luminosity of 5.4 fb^{-1}. The list of all the selection requirements is presented in Table 4.10.

Object	Selection requirement
Muons	$p_{\mathrm{T}} > 20$ GeV $2.0 < \eta < 4.5$
Jet	$p_{\mathrm{T}} > 20$ GeV $2.2 < \eta < 4.2$
Event	$60 < M_{\mu\mu} < 120$ GeV

Table 4.10: Selection requirements to select $Z(\to \mu\mu)+$ jets events in data and $Z+$ jets simulated samples.

Relevant distributions for typical kinematic quantities for data and $Z+$ jets simulated events are shown in Figure 4.3.

The yield of events selected in data is a direct measurement of how many $Z(\to \mu\mu)+$ jets events are reconstructed with the cuts listed in Table 4.10. In the full Run 2 dataset, 62569 events for $Z(\to \mu\mu)+$ 1 jet and 3921 events for $Z(\to \mu\mu)+$ 2 jets are selected. To properly measure the amount of $Z+$ jets events reconstructed and selected in this analysis, the $Z+$ jets yield has to be corrected for the difference in reconstruction performance when reconstructing only one muon instead of two. The reconstruction $\varepsilon_{\mathrm{reco}}$ and selection $\varepsilon_{\mathrm{sel}}$ efficiencies are evaluated in $Z+$ jets MC samples as

$$\varepsilon_{\mathrm{reco}} = \frac{N_{reco}}{N_{acc}} \qquad \varepsilon_{\mathrm{sel}} = \frac{N_{sel}}{N_{reco}} \tag{4.5}$$

where N_{acc} is the number of simulated events that passed the generator level cuts, N_{reco} is the number of reconstructed events, and N_{sel} is the number of selected events. Results for the $Z(\to \mu\mu)+$ 1 jet and $Z(\to \mu\mu)+$ 2 jets selection where both muons are reconstructed are listed in Table 4.11.

Process	$\varepsilon_{\mathrm{reco}}$	$\varepsilon_{\mathrm{sel}}$
$Z(\to \mu\mu)+$ 1 jet	$(25.65 \pm 0.02)\%$	$(6.03 \pm 0.02)\%$
$Z(\to \mu\mu)+$ 2 jets	$(49.61 \pm 0.02)\%$	$(1.61 \pm 0.02)\%$

Table 4.11: Reconstruction efficiency $\varepsilon_{\mathrm{reco}}$ and selection efficiency $\varepsilon_{\mathrm{sel}}$ for $Z(\to \mu\mu)+$ jets events where both muons are reconstructed.

The measured $Z+$ jets yield is therefore corrected for ε_{reco} and ε_{sel}, and finally, the reconstruction and selection efficiencies listed in Tables 4.8 and 4.9 are applied to get the expected number of $Z(\to$

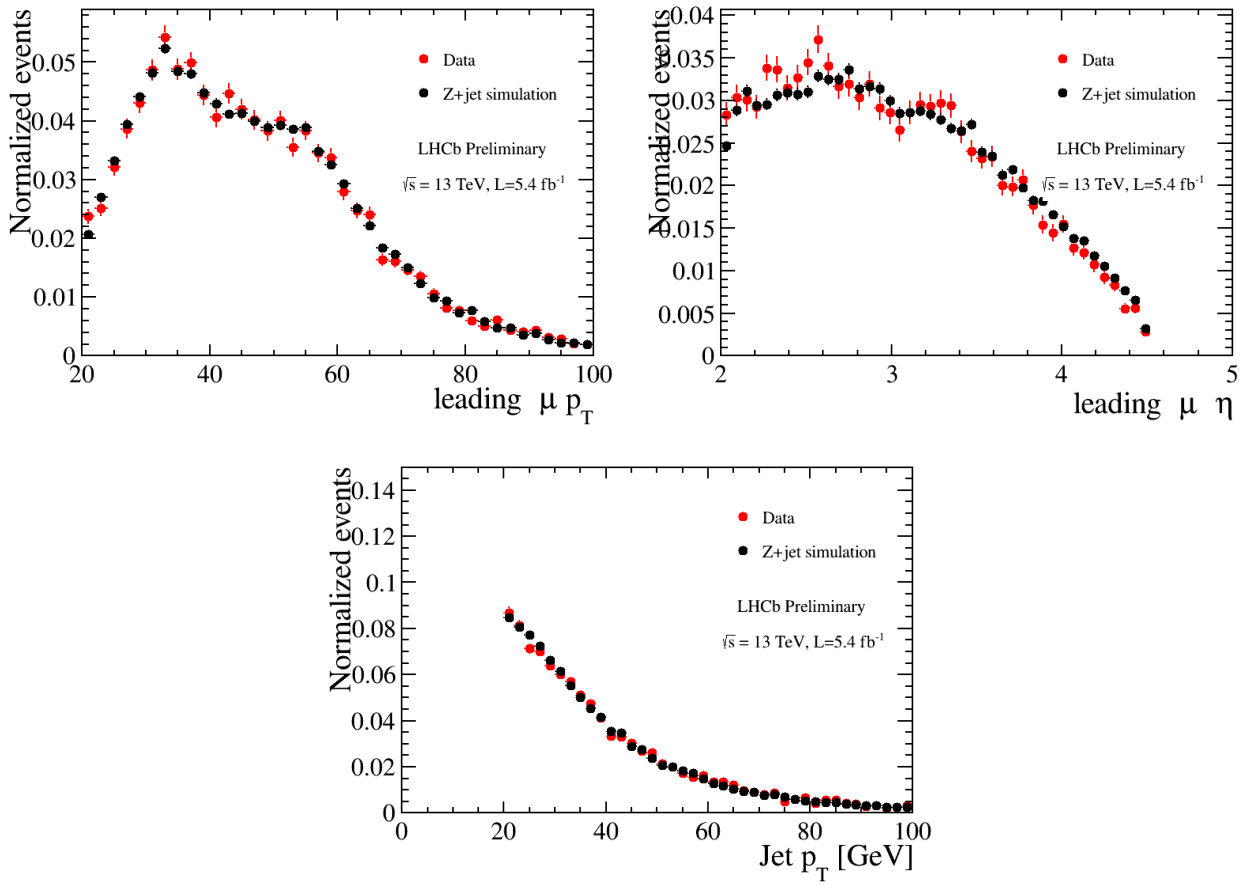


Figure 4.3: Normalized distributions of leading muon p_{T} (upper left plot), leading muon η (upper right plot) and leading jet p_{T} (lower plot) for data and $Z+$ jets events selected using cuts listed in Table. 4.10.

$\mu\mu)+$ jets events where only one muon is reconstructed. In this way, 122098 $Z(\rightarrow \mu\mu)+$ 1 jet events and 11521 $Z(\rightarrow \mu\mu)+$ 2 jets events are expected in the fiducial region of the analysis.

$W(\rightarrow \tau\nu_\tau)+$ jets and $Z(\rightarrow \tau\tau)+$ jets estimation

A second source of background comes from the decay of W and Z bosons into τ leptons, where the τ leptons subsequently decay into muons. For the $Z(\rightarrow \tau\tau)+$ jets case, this is computed from the $Z(\rightarrow \mu\mu)+$ jets data-driven estimate, corrected for the probability of reconstructing the τ leptons, assuming lepton flavor universality (*i.e.* the Z boson has the same probability of decaying into a $\mu\mu$ or $\tau\tau$ final state). For the $Z(\rightarrow \tau\tau)+1$ jet selection, a total amount of 23047 events is selected, while 2601 events are selected for the $Z(\rightarrow \tau\tau)+2$ jets selection. For the $W(\rightarrow \tau\nu_\tau)+$ jets case, this contribution is included in the fit to the muon isolation described in Section 4.5, and it is fitted together with the $\mu\nu_\mu$ plus jets final state. Once the fit is performed, the $W(\rightarrow \tau\nu_\tau)+$ jets is estimated from the probability of reconstructing the τ leptons already presented in Tables 4.8 and 4.9, and subsequently removed from the $W(\rightarrow \mu\nu_\mu)+$ jets contribution. The relative contribution of $W(\rightarrow \tau\nu_\tau)+$ jets to $W(\rightarrow \mu\nu_\nu)+$ jets is estimated from simulations, and it is found to be 3.04%.

$t\bar{t}$, single top, WW and WZ estimation

Finally, the normalizations of $t\bar{t}$, single top and vector boson pair production are obtained by rescaling the LO cross-section expectations listed in Table 4.7 to NLO predictions. The LO σ_{LO} and NLO σ_{NLO}

cross-sections and the expected number of events after having applied reconstruction and selection efficiencies listed in Tables 4.7, 4.8 and 4.9 are shown in Table 4.12.

Process	σ_{LO} (pb)	σ_{NLO} (pb)	events for W+ 1 jet selection	events for W+ 2 jets selection
$t\bar{t}$	529.2	833.9	21469	9088
single top	153.2	214.2	3693	1145
WW	67.2	106.0	4078	1207
WZ	24.6	50.6	1041	295

Table 4.12: LO σ_{LO} and NLO σ_{NLO} cross-sections and the expected number of events at NLO after having applied reconstruction and selection efficiencies for $t\bar{t}$, single top, WW, and WZ processes.

4.4.2 QCD background estimation

Another relevant background contribution to W+ jets events comes from QCD background. This QCD background mainly arises from jets produced in the proton-proton collision, together with a very high-energetic muon inside one of the jets. These very energetic muons may come from the decay of heavy-flavor quarks or in-flight hadrons, and they might mimic the signal signature. The impact of the QCD background is evident when studying the muon isolation distributions for data and MC simulated events. Muon isolation is defined as

$$I_\mu = \frac{p_{\mathrm{T}}(\mu)}{p_{\mathrm{T}}(\mu^{\mathrm{jet}})} \tag{4.6}$$

where $p_{\mathrm{T}}(\mu)$ is the transverse momentum of the muon and $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ is the transverse momentum of a jet of radius $R = 0.4$ surrounding the muon. It measures the activity (*i.e.* the number of particles) surrounding the signal muon: more isolated muons have I_μ values close to one. A pictorial representation of I_μ is shown in Figure 4.4.

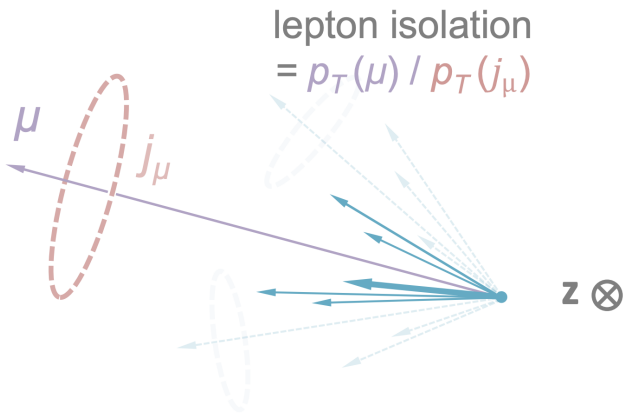


Figure 4.4: Pictorial representation of the muon isolation I_μ.

Figure 4.5 compares the I_μ distributions (normalized to one) for data and W+ 1 jet and Z+ 1 jet events. Two main considerations can be made from this comparison:

- I_μ is not able to differentiate between different EW processes, given that they show almost the same behavior for I_μ, *i.e.* a very well-isolated muon. This is expected, given that the muons coming from EW processes are rather similar and therefore it is their activity around them. This poses the question of how to separate EW backgrounds from the W+ jets signal candidates: as described in Section 4.4.1, expectations for EW backgrounds are simply removed before obtaining the W+ jets yield;

- in the region $I_\mu < 0.8$, almost no EW processes are present, but a relevant amount of events populate this region. This is where the QCD contribution comes in, with muons that indeed present lower I_μ as they tend to have higher activity around them.

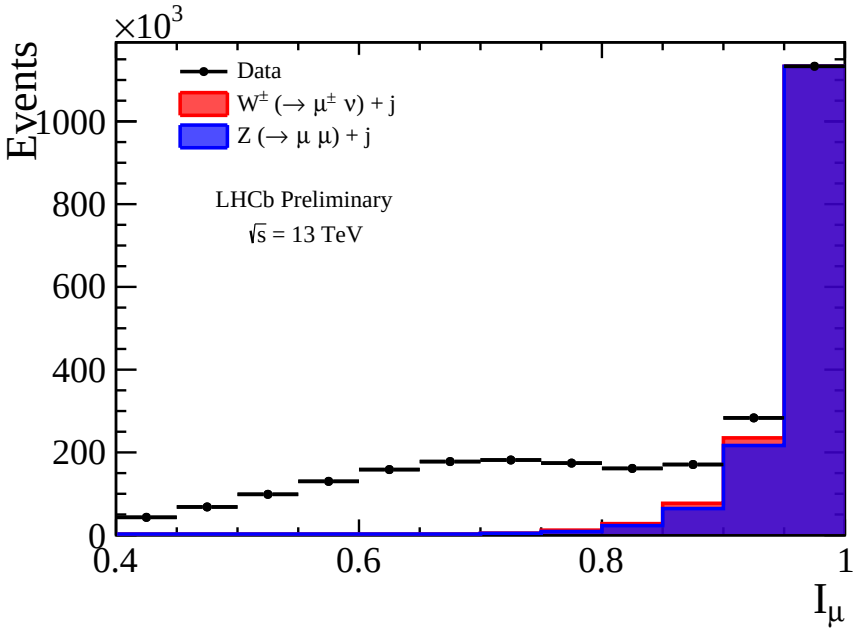


Figure 4.5: I_μ distributions in data, $W+$ 1 jet and $Z+$ 1 jet events. Distributions are normalized to match the bin of highest I_μ, in order to demonstrate that the $I_\mu < 0.8$ range cannot be described by EW processes.

Muons produced by QCD might be very isolated, and therefore also the high isolation range (*i.e.* for $I_\mu > 0.8$) might be populated by these processes. As a quick check of this statement, samples of $b\bar{b}$, $c\bar{c}$ and $q\bar{q}$ (where q stands for u,d,s quarks, and gluons) di-jets have been generated with PYTHIA8 and passed through the LHCb simulation framework; the events are reconstructed and selected with the same analysis cuts described in Section 4.2. I_μ templates for these di-jets samples are shown in Figure 4.6: clearly, the vast majority of these events are in the $I_\mu < 0.8$ region, where indeed muons show higher activity around them, but a not negligible contribution is also present in the $I_\mu > 0.8$ range, where all the EW processes (and particularly the $W+$ jets signal) are present.

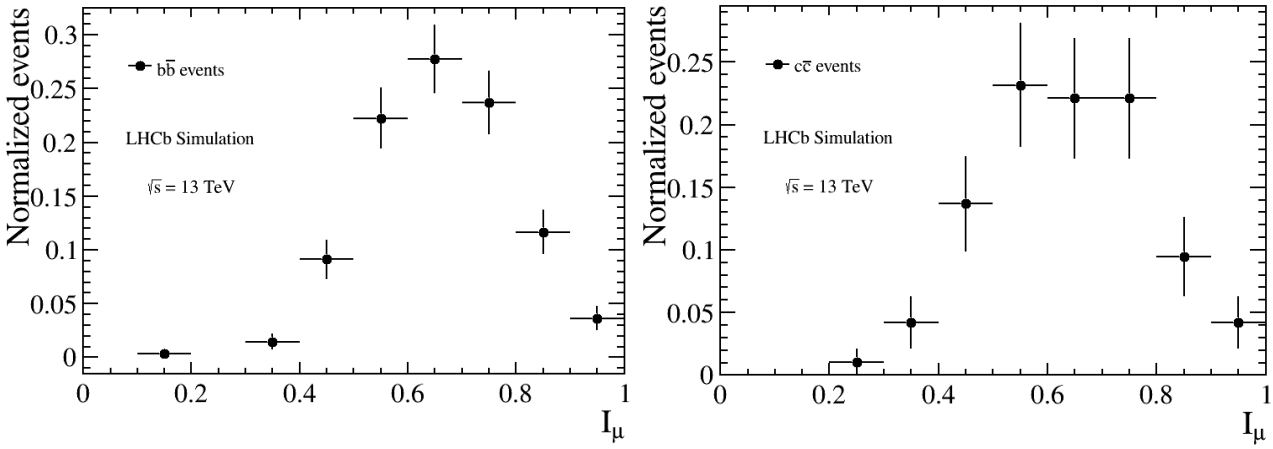


Figure 4.6: I_μ distributions for $b\bar{b}$ (left) and $c\bar{c}$ (right) di-jets events. Distributions are normalized to one.

It is therefore necessary to properly describe the QCD background in the $I_\mu > 0.8$ region. One

possible approach will require simulating this background with MC generators, but this poses two main issues: it is not easy to simulate all the QCD processes happening, and the number of events to be simulated will be huge given that the reconstruction and selection requirements are optimized for the signal selection. To avoid these two issues, a data-driven approach is used. This on the other hand poses the problem that QCD events in the $I_\mu > 0.8$ region cannot be disentangled from the EW processes. Therefore, a technique based on the definition of a signal (SR) and a control (CR) region is considered. The discriminating variable used to define SR and CR is the total transverse momentum of the muon and the jets $p_{\mathrm{T}}^{\mathrm{total}} = p_{\mathrm{T}}(\mu^{\mathrm{jet}} + \mathrm{jets})$. The reason for the choice of this variable for defining SR and CR is the following: events, where a W boson is present, are characterized by a higher $p_{\mathrm{T}}^{\mathrm{total}}$ due to the emission of a neutrino, which contributes to missing transverse momentum; on the other hand, QCD events do not have this behavior, given that typically no neutrino is produced in those events, resulting in a lower $p_{\mathrm{T}}^{\mathrm{total}}$. This is quite evident in Fig **??**, where W+jets events tend to have a $p_{\mathrm{T}}^{\mathrm{total}}$ more shifted towards higher values, due to neutrino emission. As a cross-check of this choice, also $Z(\rightarrow \mu\mu)$+ jets events are shown: since no neutrino is present in the $Z(\rightarrow \mu\mu)$+ jets final state objects, the peak of $p_{\mathrm{T}}^{\mathrm{total}}$ is shifted to lower values.

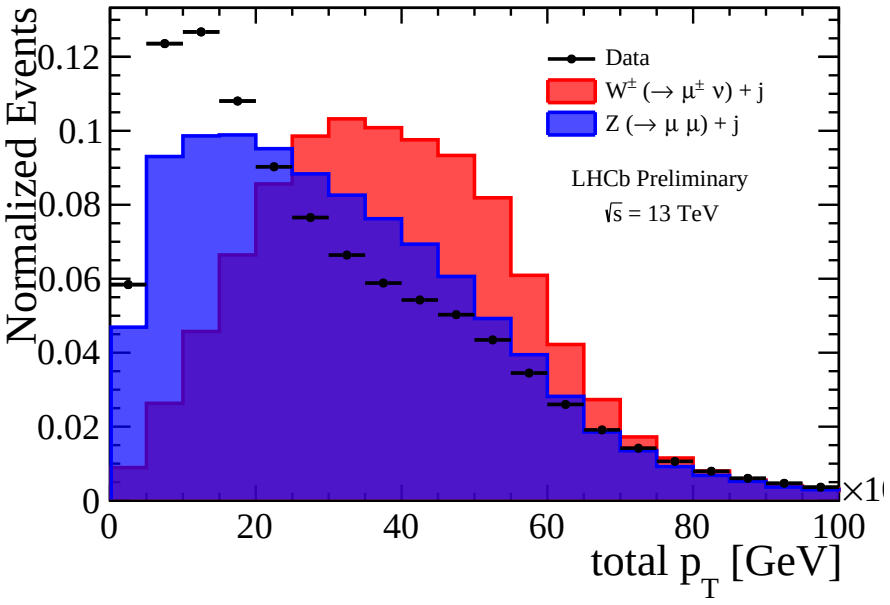


Figure 4.7: $p_{\mathrm{T}}^{\mathrm{total}}$ distributions in data, W+ 1 jet (red) and Z+ 1 jet (blue) events. Distributions are normalized to one.

Events with high $p_{\mathrm{T}}^{\mathrm{total}}$ (typically EW events where a neutrino is present, such as W+ jets) therefore tend to be more *unbalanced* in the $p_{\mathrm{T}}^{\mathrm{total}}$, while QCD background events have a more *balanced* $p_{\mathrm{T}}^{\mathrm{total}}$. This is why $p_{\mathrm{T}}^{\mathrm{total}}$ is a good variable to create SR and CR. The SR (*i.e.* the region with *unbalanced* events) is defined for $p_{\mathrm{T}}^{\mathrm{total}} > 20$ GeV, while the CR (*i.e.* the region with *balanced* events) is defined for $p_{\mathrm{T}}^{\mathrm{total}} < 20$ GeV. While the $p_{\mathrm{T}}^{\mathrm{total}} < 20$ GeV cut mainly selects QCD events, there might be differences between CR and the actual QCD distributions in SR: this is expected given that the cut on $p_{\mathrm{T}}^{\mathrm{total}}$ selects events that might have different kinematic features. The effect of the cut on $p_{\mathrm{T}}^{\mathrm{total}}$ is evident in Figure 4.8, where the distributions of $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ normalized to one for SR and CR for $I_\mu < 0.8$ are shown: despite selecting mostly QCD events because the $I_\mu < 0.8$ cut rejects almost all EW processes, differences are evident.

To take into account these differences, a reweighting procedure is put in place: the CR I_μ templates are reweighted to match SR I_μ templates. Weights are computed by requiring that the $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$

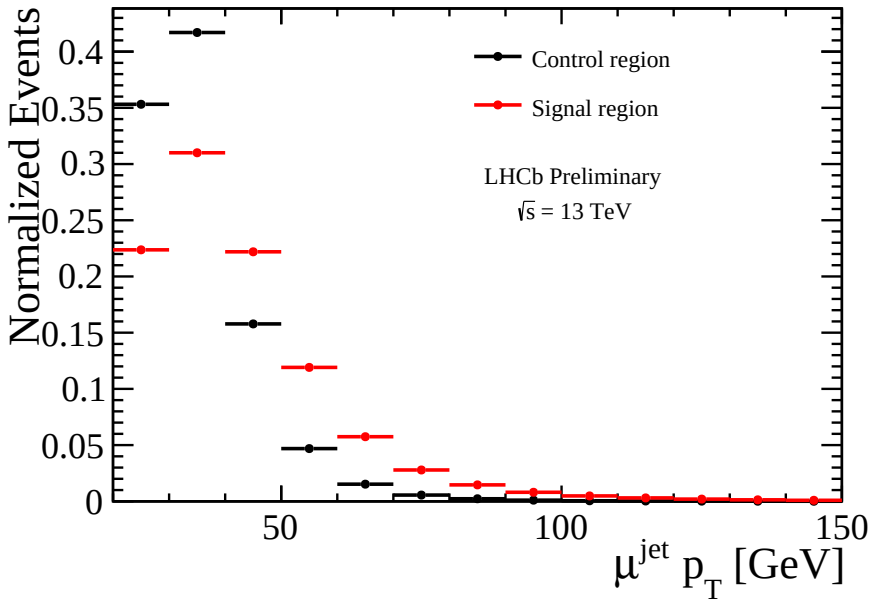


Figure 4.8: $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ distributions in SR (red dots) and CR (black dots), for $I_{\mu} < 0.8$. Distributions are normalized to one and show the evident differences between the kinematics of the two regions.

distribution in the SR matches the $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ in CR. This is required in the $I_{\mu} < 0.8$ region, where only QCD events are selected, to avoid any contamination from EW processes. Figure 4.9 shows the weights obtained as a function of $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$.

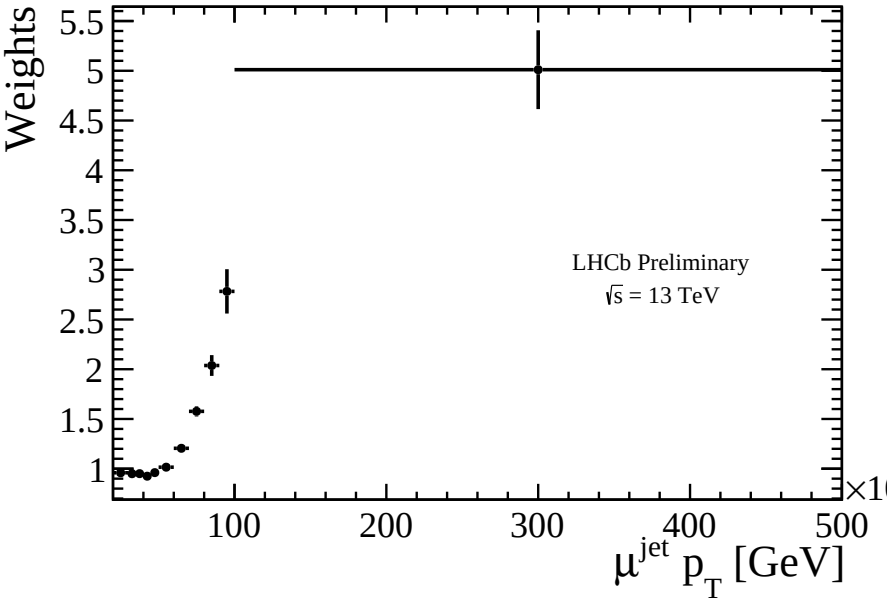


Figure 4.9: Weights obtained by comparing $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ distributions in SR and CR. Weights are shown as a function of $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$. A variable binning is chosen to ensure that all bins have enough statistics to compute the weights.

Figure 4.10 shows the effect of the reweighting procedure on the $p_{\mathrm{T}}(\mu^{\mathrm{jet}})$ distributions, where indeed the reweighted distribution now matches the SR one, particularly for the low p_{T} range where most

of the events are present. Once weights are computed, they are applied to the I_μ templates obtained from the CR and used as I_μ templates to describe QCD background in the SR. This is shown in Figure 4.11, where I_μ templates are normalized to unity in the range $[0.4, 0.8]$ to visualize the effect of the reweighting: indeed, the reweighted I_μ distribution now matches the SR I_μ distribution in the region where mostly QCD events are present. This reweighting procedure, therefore, allows modeling the QCD I_μ distribution even for high values of I_μ.

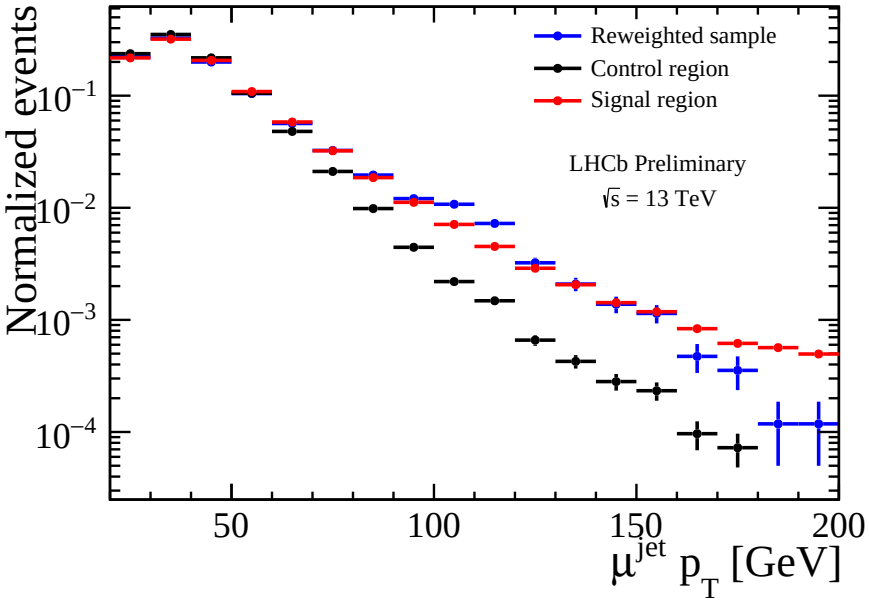


Figure 4.10: $p_T(\mu^{jet})$ distributions for SR (red dots), CR (black dots), and the reweighted distribution (blue dots). Distributions are normalized to one and show the evident matching between SR and reweighted distributions.

Table 4.13 list the final analysis cut for the W+ jets measurements: they pretty much resemble the fiducial region cuts listed in Table 4.1, apart from the last request on p_T^{total}, where μ is substituted by μ^{jet}. This difference will be taken into account in Section 4.7.

Object	Fiducial Cut
Muons	$p_T > 20$ GeV $2.0 < \eta < 4.5$
Jets	$p_T > 20$ GeV $2.2 < \eta < 4.2$
Event	$p_T^{total} > 20$ GeV

Table 4.13: Analysis cuts applied to the final state objects in the W+ jets analysis. To be compared with Table 4.1, where $p_T(\mu + \text{jets})$ is used instead of p_T^{total}.

4.5 Fit procedure

The goal of the analysis is to perform a cross-section measurement of the W+ jets process, both in the full fiducial region and differentially with respect to several kinematic observables, namely the muon pseudorapidity η^μ, the leading jet transverse momentum p_T^{jet} and the leading jet pseudorapidity η^{jet}. The binning schemes for the kinematic observables involved in the differential cross-section

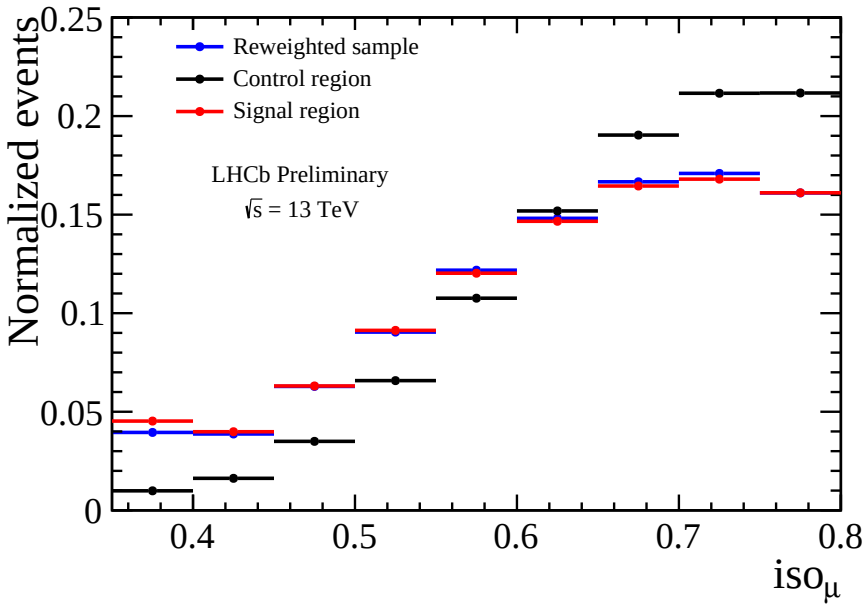


Figure 4.11: I_μ distributions for SR (red dots), CR (black dots) and the reweighted distribution (blue dots) for $I_\mu < 0.8$. Distributions are normalized to one and show the evident matching between SR and reweighted distributions.

measurements are listed in Table 4.14.

Variable	$W+$ 1 jet	$W+$ 2 jets
$p_{\mathrm{T}}^{\mathrm{jet}}$	$[20, 25, 30, 40, 50, 60, 100]$	$[20, 30, 40, 60, 100]$
η^{jet}	$[2.2, 2.7, 3.2, 3.7, 4.2]$	$[2.2, 2.7, 3.2, 3.7, 4.2]$
η^{μ}	$[2.0, 2.5, 3.0, 3.5, 4.5]$	$[2.0, 2.5, 3.0, 3.5, 4.5]$

Table 4.14: Binning schemes for the relevant observables considered in the differential cross-section measurements.

To measure the cross-section, it is necessary to measure the yield of $W+$ jets events. As already described in Section 1.7, both ATLAS and CMS have performed this kind of measurement, by fitting the distribution of the W boson transverse mass m_T^W, defined as

$$m_T^W = \sqrt{2 p_l p_\nu (1 - \cos\phi_{l\nu})} \tag{4.7}$$

where p_l (p_ν) is the transverse momentum of the lepton (neutrino) coming from the W boson decay, and $\phi_{l\nu}$ is the azimuthal angle between the lepton and the neutrino.

In the LHCb experiment, this kind of analysis procedure is not possible, because LHCb is not a hermetic detector, and therefore no information on the missing energy is available. As proposed in the previous Run 1 analysis, a fit to the muon isolation I_μ defined in Equation 4.6 is considered. I_μ templates for signal and EW backgrounds are obtained from MC simulations, while the QCD I_μ template is obtained with the reweighting procedure described in Section 4.4.2. Before performing the fit, another aspect has to be considered: in the evaluation of the QCD background template in the SR, $W+$ jets events and other EW processes might enter in the QCD template; this is shown in Figure 4.7, where indeed the cut at $p_{\mathrm{T}}^{\mathrm{total}} > 20$ GeV leaves a part of the $W+$ jets and $Z+$ jets events in the CR, where the starting template for the QCD evaluation are obtained. Given that EW backgrounds have

been simply estimated from simulations, or from a data-driven technique for $Z+$ jets events, they can be simply removed from the QCD template by properly computing the expected number of events in the CR. A factor $f_{\mathrm{CR/SR}}$ is computed for each process from simulations, both in the fiducial region and for each bin of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ}. To properly correct for the presence of $W+$ jets events in the CR, two possible alternatives are available:

- simply remove the expected $W+$ jets events in the CR using expectations from simulations;
- a recursive fit, which gets the $W+$ jets events in the SR, compute the expected $W+$ jets event in the CR using a $f_{\mathrm{CR/SR}}$, remove those events from the QCD template and re-do the fit till reaching convergence.

While the first approach is definitely easier, it relies on the knowledge of the $W+$ jets cross-section, which is indeed the purpose of this analysis. Therefore, the second approach has been chosen. A maximum-likelihood binned fit to I_{μ} in the SR is performed, where the number of $W+$jets events $N_{\mathrm{fit}}^{W+\mathrm{jets}}(x_i)$ in the i-bin for a given kinematic variable x is obtained as

$$N_{\mathrm{fit}}^{W+\mathrm{jets}}(x_i) = N^{\mathrm{QCD}}(x_i)T^{\mathrm{QCD}}(x_i) + \sum_{p} N^{p}(x_i)T^{p}(x_i) \tag{4.8}$$

where p stands for all the EW processes considered in the fit, namely $W+$ jets, $Z+$ jets, $t\bar{t}$, single top, WW and WZ, and $x \in (p_{\mathrm{T}}^{\mathrm{jet}}, \eta^{\mathrm{jet}}, \eta^{\mu})$. After each fit, the QCD template T^{QCD} is updated by removing the number of events in the CR for a given process p.

The resulting fit to I_{μ} for $W+1$ jet and separately for W^+ and W^- are shown in Figures 4.12 and 4.13. The fit in different bins of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ} for W, W^+, W^- and for both the $W+$ 1 jet and $W+$ 2 jets selections are presented in App. B.

4.6 Reconstruction and selection efficiencies

The results from the fitting procedure obtained in the previous section have to be corrected for the reconstruction and selection efficiencies $\varepsilon_{\mathrm{reco}}$ and $\varepsilon_{\mathrm{sel}}$ defined in Equation 4.5. The evaluation of $\varepsilon_{\mathrm{reco}}$ and $\varepsilon_{\mathrm{sel}}$ is performed in the fiducial region, defined in Table 4.1. Combined together, these efficiencies account for the total efficiency in reconstructing and selecting an event.

4.6.1 Muon reconstruction efficiencies

The evaluation of the muon reconstruction at the LHCb experiment has been already described in Section3.2, where the total muon reconstruction efficiency $\varepsilon_{\mathrm{reco},\mu}$ is evaluated using $Z(\rightarrow \mu\mu)$ events reconstructed in data. These efficiencies have been measured directly from data using tag-and-probe methods: the decay of Z bosons provides a highly pure sample of events containing high-p_{T} muon pairs which can be used to determine the muon reconstruction efficiencies. Studies have also shown that these efficiencies are not affected by the jet kinematics, therefore they can be applied in this analysis.

$\varepsilon_{\mathrm{reco},\mu}$ is evaluated using MC simulations of $W+$ jets events. If only one muon is present in the final state, as in the $W+$ jets samples, the total efficiency simply corresponds to the efficiency of reconstruction, identifying, and triggering on the muon candidate. $\varepsilon_{\mathrm{reco},\mu}$ is calculated on an event-by-event basis: in this way, overall muon reconstruction efficiencies for the given bins of the considered phase-space region are obtained. For each bin i of a given kinematic variable, the efficiency is computed as

$$\varepsilon_{\mathrm{reco},i} = \frac{\sum_{i=1}^{n_{\mathrm{events}}} 1}{\sum_{i=1}^{n_{\mathrm{events}}} 1/\varepsilon_{\mathrm{reco},\mu}}, \tag{4.9}$$

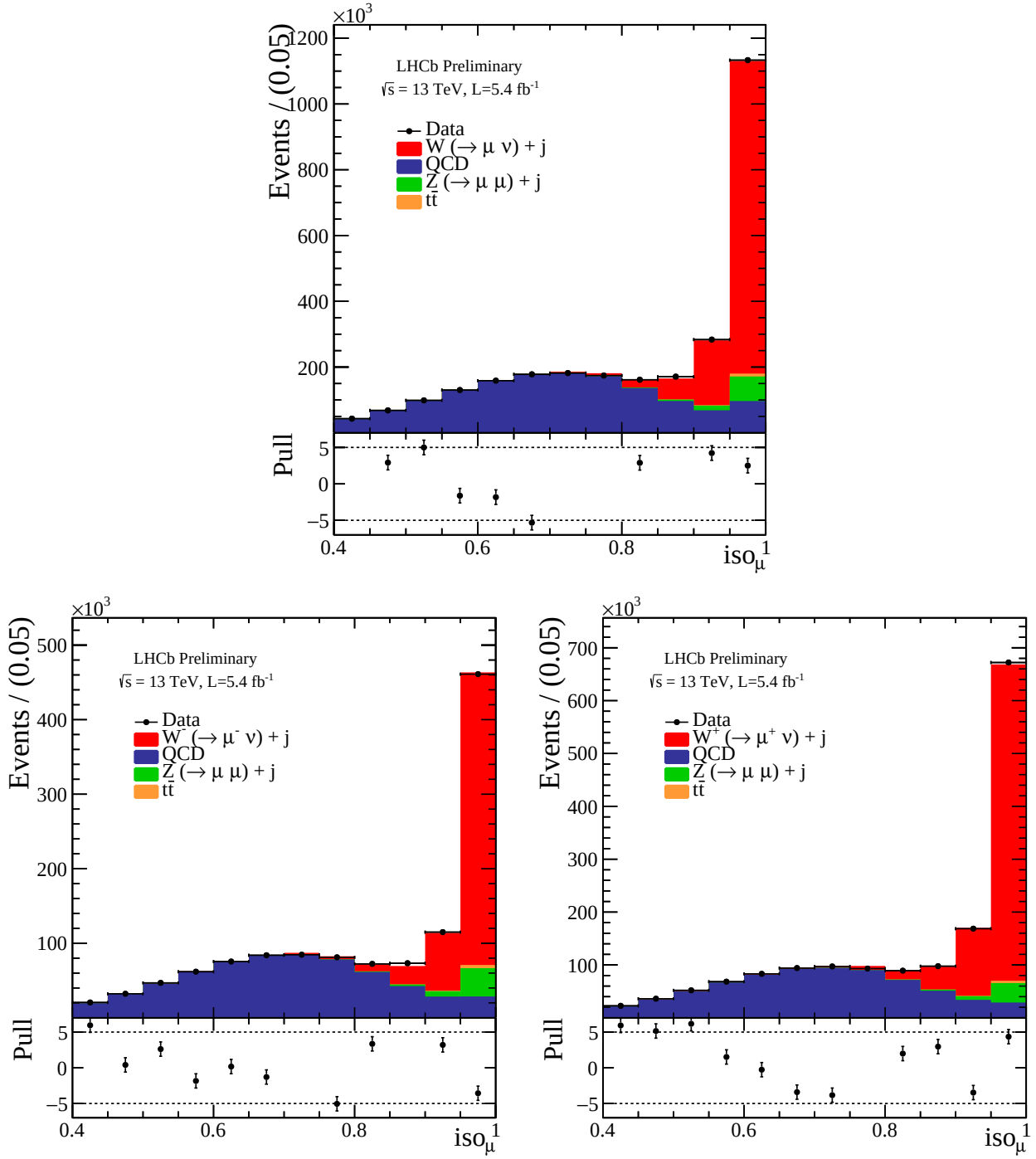


Figure 4.12: Fit to muon isolation I_μ for $W+$ 1 jet selection. Results are shown for $W+$ 1 jet (upper plot), W^-+ 1 jet (lower left plot), and W^++ 1 jet (lower right plot).

which is the number of events in a given bin divided by their efficiency-corrected sum, obtained by weighting each event by the inverse of the efficiency to reconstruct that event. Efficiencies are directly computed on $W+$ jets simulated samples, to avoid the pollution from the background in the data sample. The muon reconstruction efficiency evaluated for $W+1$ jet and separately for W^+ and W^- is shown in Figure 4.14 as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ}. As expected, no evident dependence is present as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ and η^{jet}, while the decreasing trend as a function of η^{μ} is due to the trigger

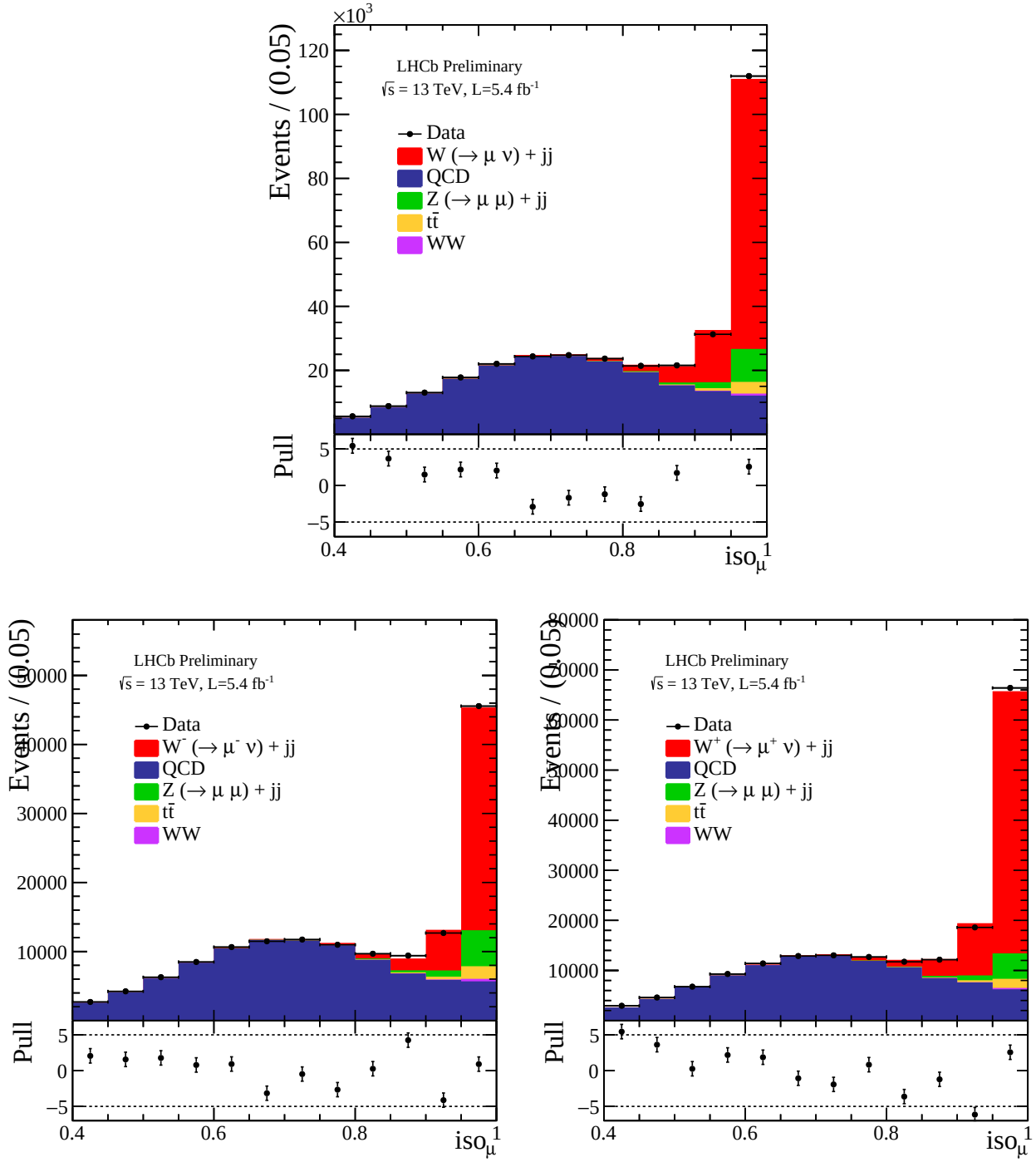


Figure 4.13: Fit to muon isolation I_μ for W+ 2 jets selection. Results are shown for W+ 2 jets (upper plot), W^-+ 2 jets (lower left plot), and W^++ 2 jets (lower right plot).

efficiency, as shown in Figure 3.3.

4.6.2 Jet reconstruction efficiency

The jet reconstruction efficiency $\varepsilon_{\text{reco,jet}}$ is the probability for a jet to be reconstructed and correctly identified. The overall jet reconstruction efficiency is dominated by the identification contribution: jets are reconstructed 99.7% of the times, while the jet identification takes into account the correct

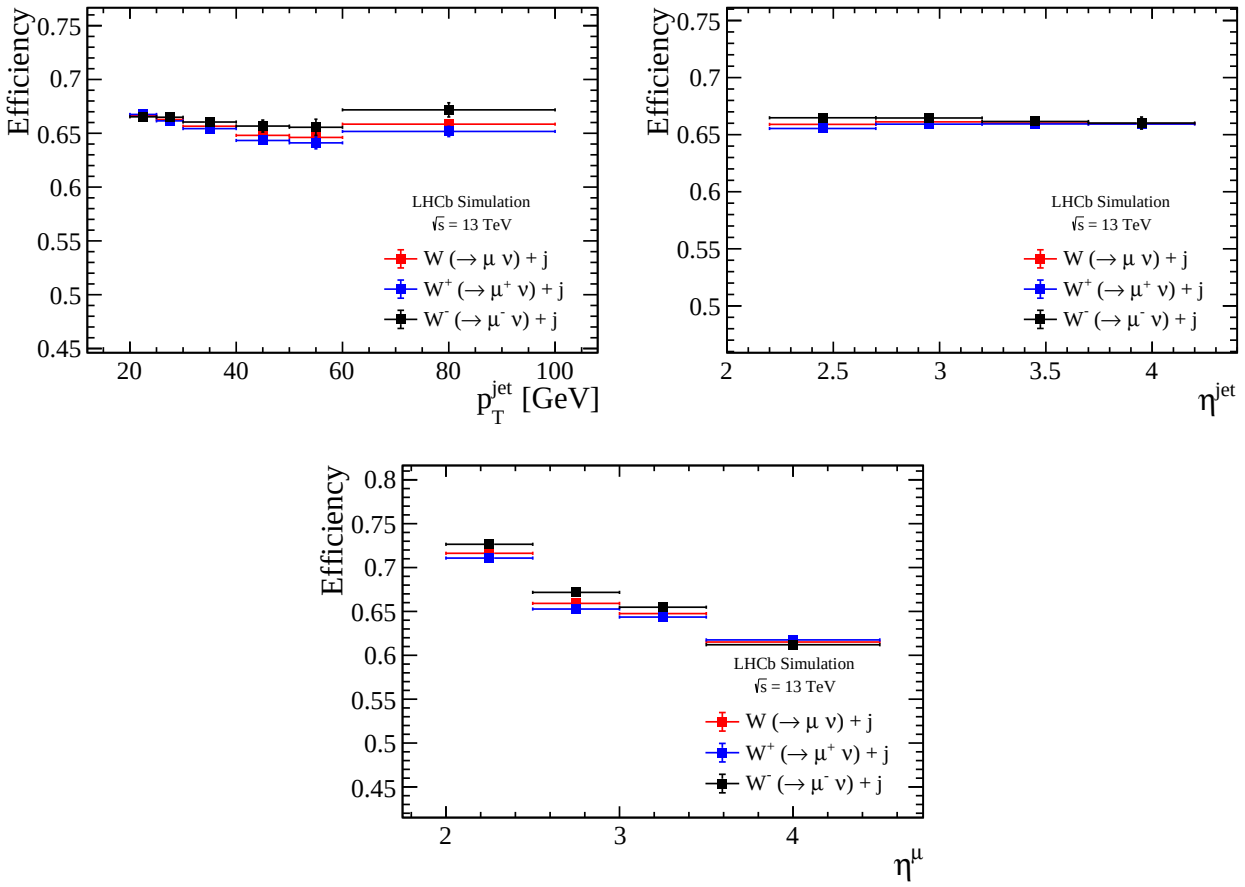


Figure 4.14: Muon reconstruction efficiency for the $W+1$ jet case, as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{jet} (upper right plot), and η^{μ} (lower plot). Results are shown for $W+$ 1 jet (red dots), W^-+ 1 jet (black dots), and W^++ 1 jet (blue dots).

matching of a reconstructed jet to a truth jet: the jet matching procedure consists in selecting those reconstructed jets that have a true jet within a distance $\Delta R < 0.5$ in the (η, ϕ) space. The jet reconstruction efficiency has been measured in $Z(\rightarrow \mu\mu)$+jet simulations, and it has been evaluated as a function of the kinematic variables used in this analysis. In Figure4.15, the jet reconstruction for the $W+1$ jet selection is shown. The dependence from the jet transverse momentum is expected, where indeed the jet identification requirements listed in Section 3.3.5 have a much larger impact at lower jet transverse momenta: for example, the requirement that the jet contains a charged particle with $p_{\mathrm{T}} > 1.4$ GeV has a larger impact at low p_{T}; at the same time, the jet reconstruction is almost flat as a function of η^{jet}; this is expected, given that the η^{jet} region $(2.2 < \eta < 4.2)$ has been chosen to not have any dependence on η^{jet}. Finally, no dependence is present as a function of η^{μ} and as a function of the muon charge.

4.6.3 W selection efficiency

Besides the probability of properly reconstructing a high-energetic muon and jets, additional requirements are placed on the $W+$ jets sample in order to reduce the background contributions. These have been already listed in Tables 4.8 and 4.9. These additional requirements are grouped together and the efficiency with which signal events pass these requirements is referred to as the W selection efficiency. This efficiency includes the requirements on the muon impact parameter IP, the energy

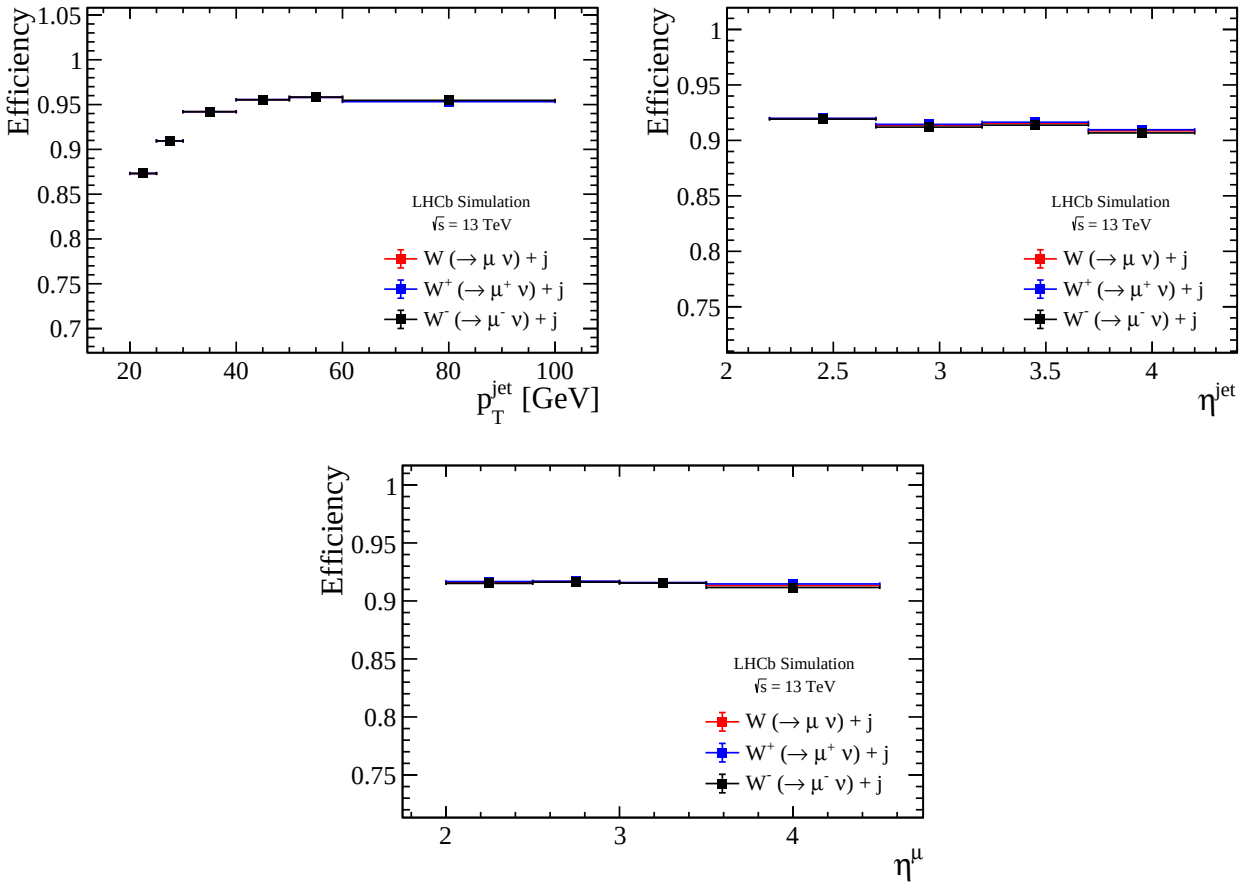


Figure 4.15: Jet reconstruction efficiency for the W+ jet case as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{jet} (upper right plot) and η^{μ} (lower plot). Results are shown for W+ 1 jet (red dots), W^-+ 1 jet (black dots), and W^++ 1 jet (blue dots).

released by the muon in the calorimeter system normalized by its momentum $(E_{\mathrm{ECAL}} + E_{\mathrm{HCAL}})/p$, the second muon veto to reject events compatible with Z boson candidates, and multiple candidates removal. The overall W selection efficiency as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ} are shown in Figure 4.16. The main contribution comes from the multiple candidates' removal, which indeed creates the trend as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$: when considering lower values of $p_{\mathrm{T}}^{\mathrm{jet}}$, more multiple candidates are created, and therefore the efficiency is lower. No evident trend is present for η^{jet} and η^{μ}.

4.6.4 Total reconstruction efficiency

The reconstruction and selection efficiencies obtained in Sections 4.6.1, 4.6.2 and 4.6.3 are combined together to give the total reconstruction efficiency, whose trend as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ} is shown in Figure 4.17.

4.7 Acceptance factor

An acceptance factor $\mathcal{A}$ is introduced to take into account differences in the fiducial region and the actual region where the analysis is performed: indeed, while the fiducial cut on $p_{\mathrm{T}}^{\mathrm{total}}$ is $p_{\mathrm{T}}(\mu + \mathrm{jets}) > 20$ GeV, the analysis cut requires $p_{\mathrm{T}}(\mu^{\mathrm{jet}} + \mathrm{jets}) > 20$ GeV, as already described in Section 4.4.2. Consequently, the acceptance factor accounts for differences in these two variables arising from the

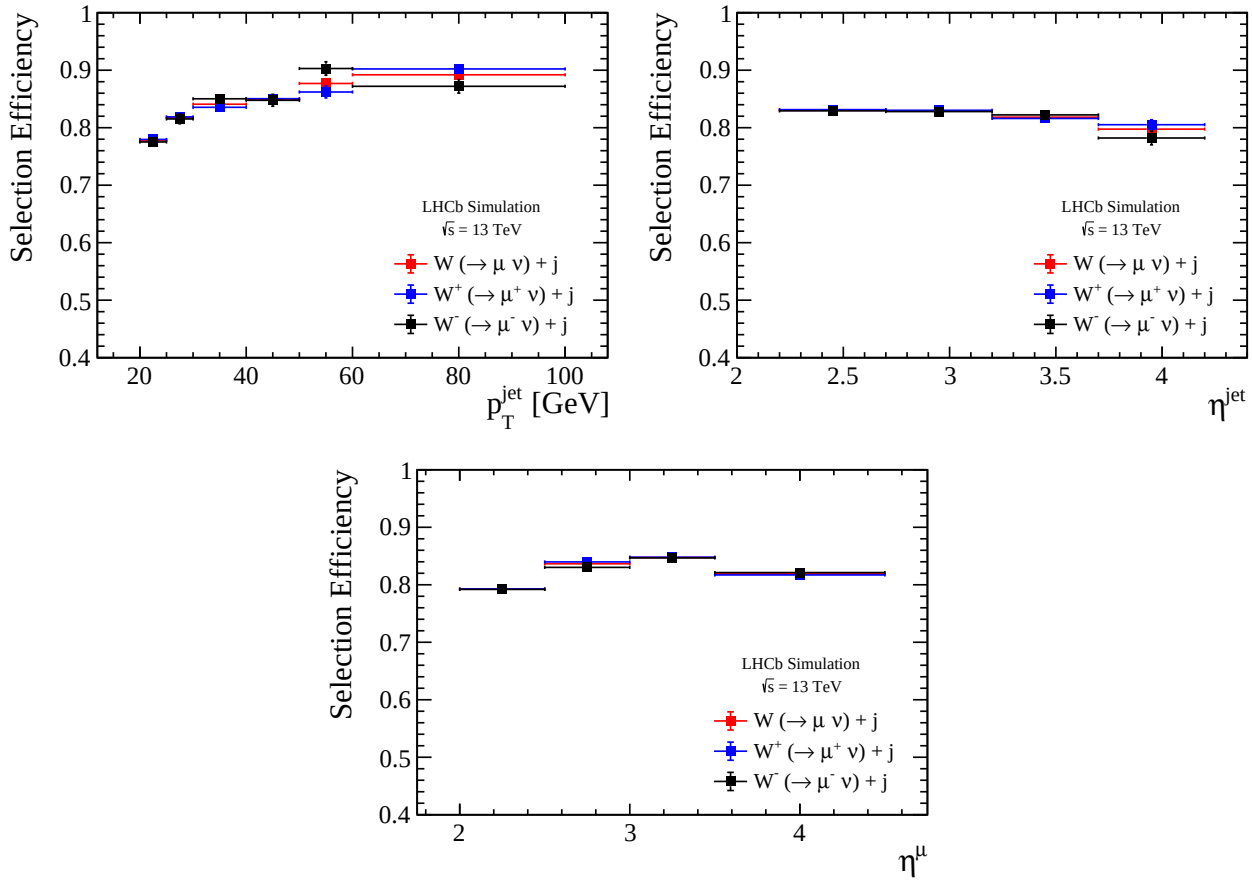


Figure 4.16: W selection efficiency for the $W+$ jet case as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{jet} (upper right plot) and η^{μ} (lower plot). Results are shown for $W+$ 1 jet (red dots), $W^{-}+$ 1 jet (black dots), and $W^{+}+$ 1 jet (blue dots).

extra activity which may be present in the neighborhood of the signal muon. For a given bin i of the considered kinematic variable, the acceptance factor $\mathcal{A}_i$ is computed as

$$\mathcal{A}_i = \frac{N(p_{\mathrm{T}}(\mu + \mathrm{jets}) > 20 \text{ GeV})}{N(p_{\mathrm{T}}(\mu^{\mathrm{jet}} + \mathrm{jets}) > 20 \text{ GeV})} \tag{4.10}$$

where $N(p_{\mathrm{T}}(\mu + \mathrm{jets}) > 20$ GeV) ($N(p_{\mathrm{T}}(\mu^{\mathrm{jet}} + \mathrm{jets}) > 20$ GeV)) is the number of events in the bin i that satisfy the requirement $p_{\mathrm{T}}(\mu + \mathrm{jets}) > 20$ GeV ($p_{\mathrm{T}}(\mu^{\mathrm{jet}} + \mathrm{jets}) > 20$ GeV). The acceptance factors as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet}, and η^{μ} are shown in Figures 4.18 for the $W+$ 1 jet selection.

4.8 Unfolding

The measured distributions as a function of jet variables must be corrected back to the true distributions in the same events: indeed, effects due to the finite resolution of the detector might alter the reconstructed distributions. By selecting events that contain 'real' reconstructed jets, and plotting both the reconstructed and true p_{T} distributions, Figure 4.19 shows that some correction is needed, to properly account for the distribution of $p_{\mathrm{T}}^{\mathrm{jet}}$ and η^{jet}. This technique is called unfolding, and it will be applied only to the distribution of $p_{\mathrm{T}}^{\mathrm{jet}}$ and η^{jet}, given that resolution effects are almost negligible when considering muons.

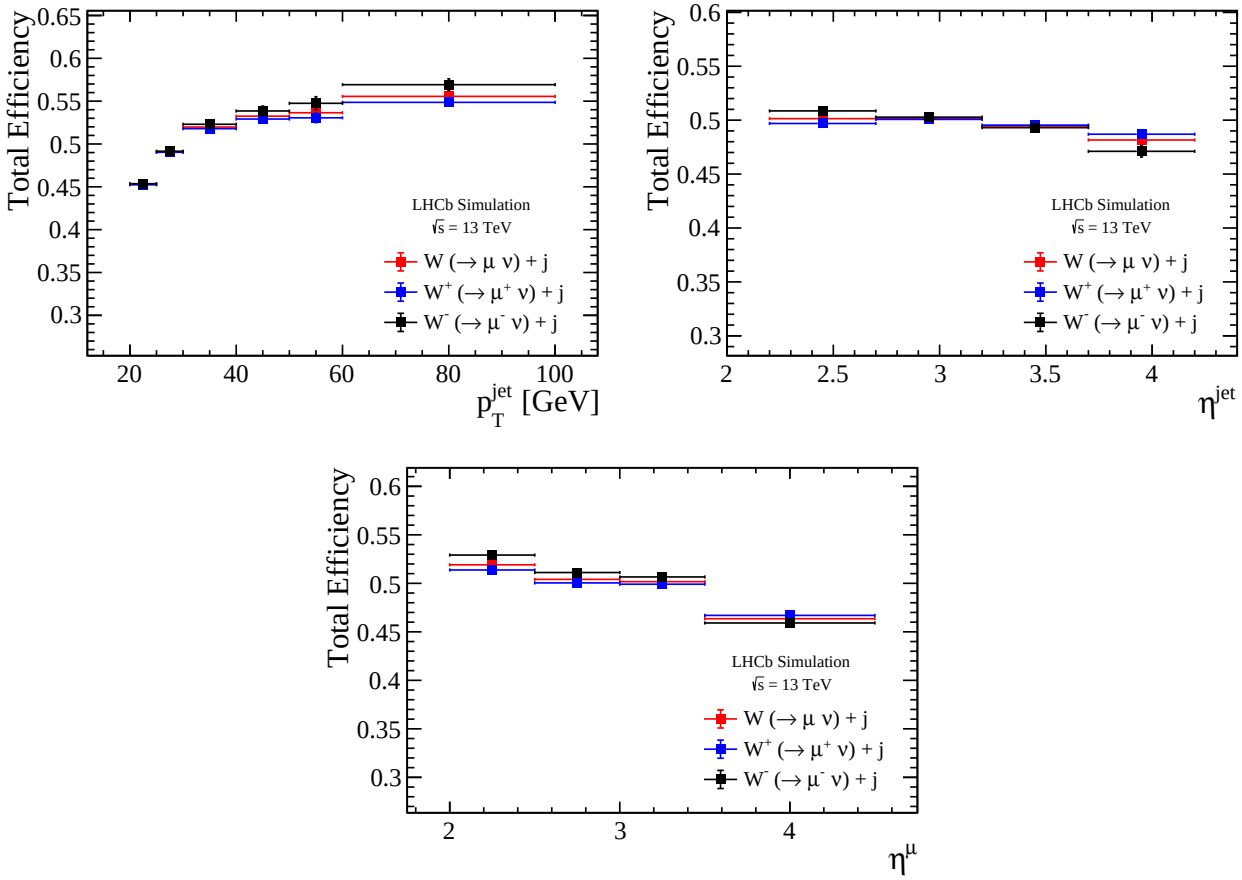


Figure 4.17: Total efficiency for the $W+$ jet case as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{jet} (upper right plot) and η^{μ} (lower plot). Results are shown for $W+$ 1 jet (red dots), $W^{-}+$ 1 jet (black dots), and $W^{+}+$ 1 jet (blue dots).

The first step of unfolding is to compute the response matrix. Response matrices U_{ij} are obtained for $p_{\mathrm{T}}^{\mathrm{jet}}$ and η^{jet} distributions. If m is the number of bins for the reconstructed variable and n is the number of bins for the truth-level variable, then U_{ij} is a matrix with $m \times n$ elements. For the consistency of the unfolding procedure, m must be greater or equal to n. The unfolding matrix U_{ij} is obtained by counting the events in the (reconstructed,truth-level) space. Events that pass the analysis cuts and where jets are matched with truth-level jets are used in the computation of U_{ij}. To compute the response matrices, the same binning scheme used for the analysis presented in Table 4.14 is considered. Figures 4.20 and 4.21 shows the response matrices for $p_{\mathrm{T}}^{\mathrm{jet}}$ and η^{jet} for the $W + 1$ jet and $W + 2$ jets case selection respectively.

There are several techniques available to perform unfolding. For this analysis, the Singular Value Decomposition (SVD) technique is used [89, 90].

4.9 Systematic uncertainties

Systematic uncertainties may affect the analysis at different levels:

- they may affect the fitting procedure (Fit) by changing the templates used in the fits or the normalization of backgrounds;

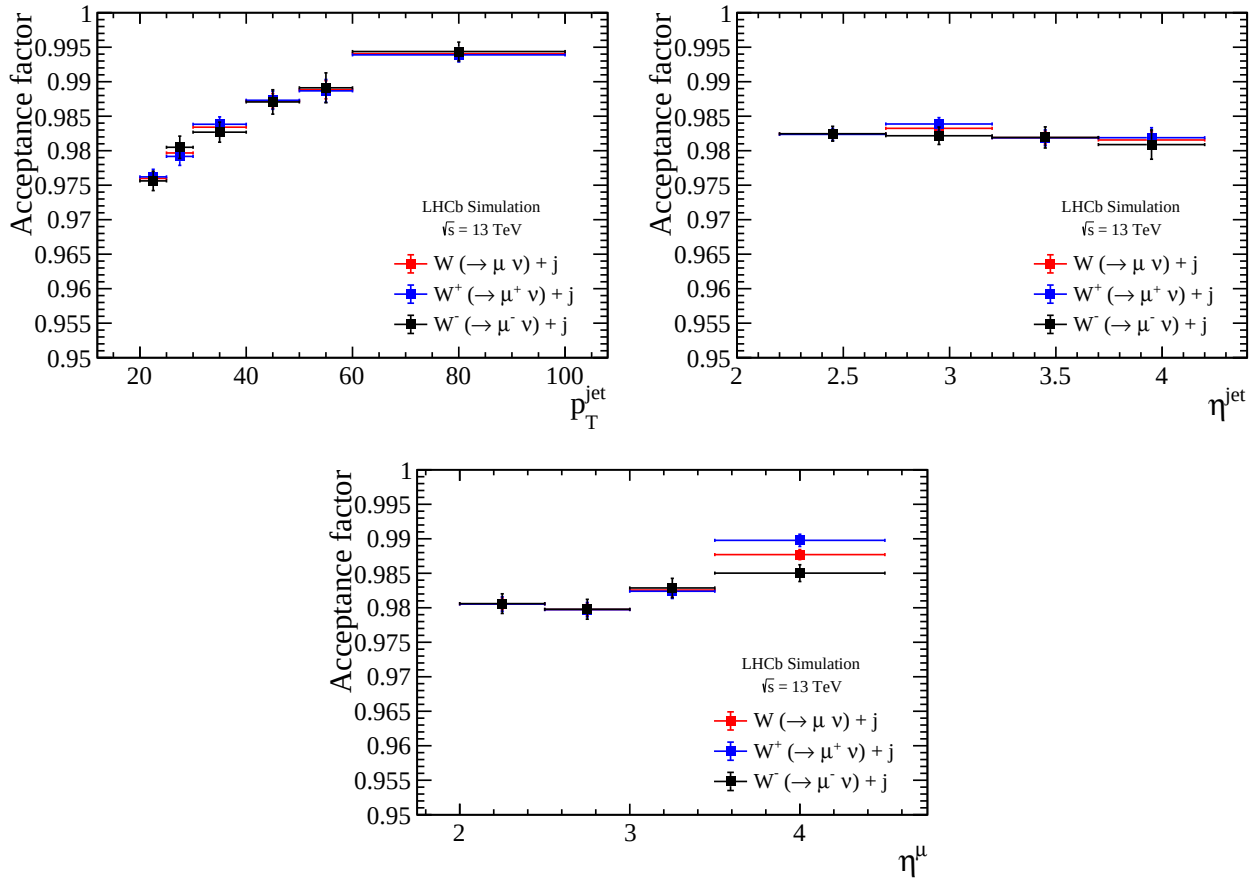


Figure 4.18: The acceptance factor $\mathcal{A}$ as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{jet} (upper right plot), and η^{μ} (lower plot) for the $W+$ 1 jet. Results are shown for $W+$ 1 jet (red dots), $W^{-}+$ 1 jet (black dots), and $W^{+}+$ 1 jet (blue dots).

- they may affect the efficiency evaluation (ε);
- they may affect the unfolding procedure (U_{ij}).

Table 4.15 present a list of the considered systematic uncertainties for this analysis and which aspect they might impact on. The systematic uncertainty coming from luminosity is simply added in quadrature to the other systematic uncertainties. In the following, an overview of the systematic uncertainties considered for this analysis is presented.

Source	Fit	ε	U_{ij}
Muon reconstruction		X	
Jet reconstruction			X
Jet Energy Resolution		X	X
Jet Energy Scale		X	X
Selection		X	
QCD background modeling	X		
Templates	X		
Luminosity			

Table 4.15: Systematic uncertainties considered in this analysis and where they impact.

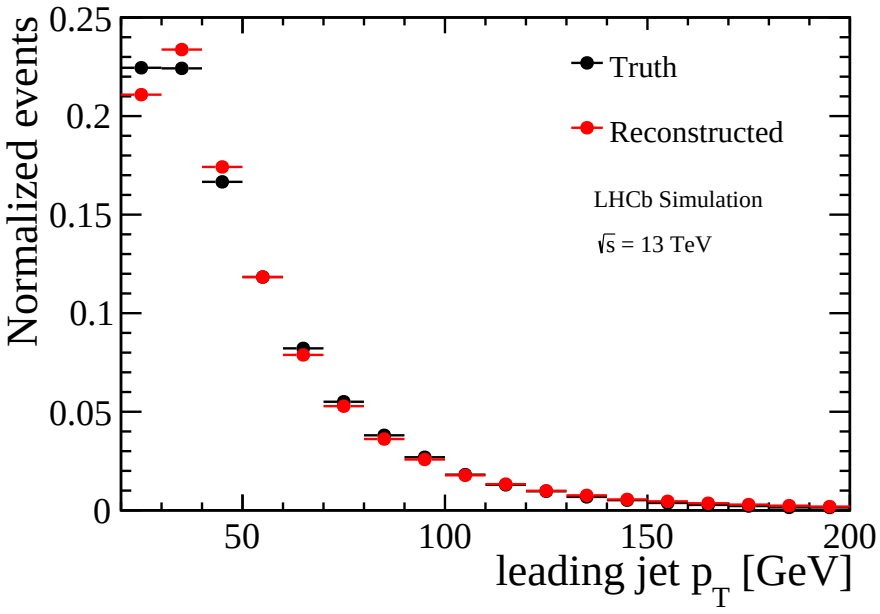


Figure 4.19: Comparison between truth-level (black dots) and reconstructed (red dots) p_{T} of the leading jet for W+ 2 jets events.

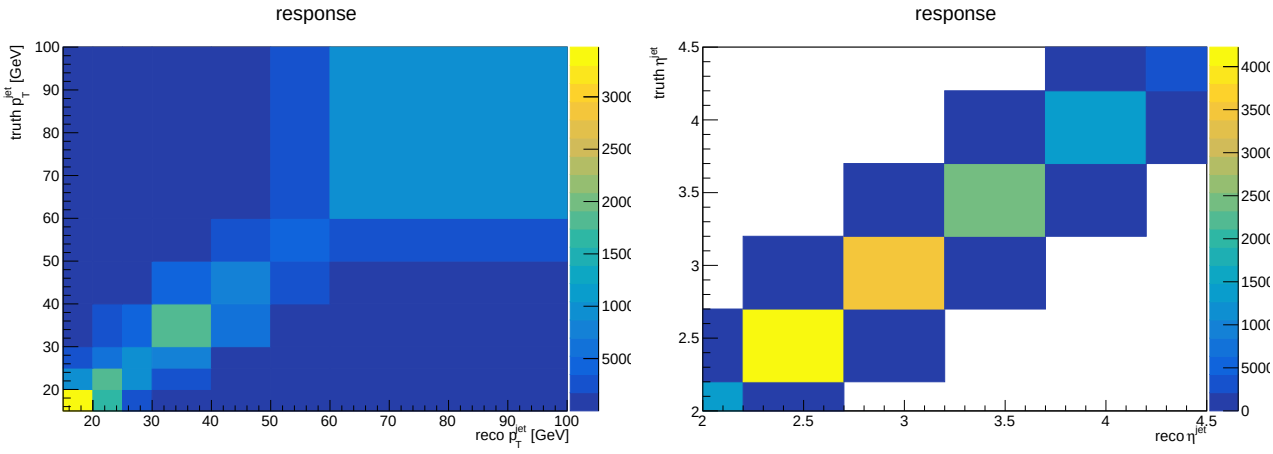


Figure 4.20: (left) Response matrix for $p_{\mathrm{T}}^{\mathrm{jet}}$ for W+ 1 jet selection. (right) Response matrix for η^{jet} for W+ 1 jet selection.

4.9.1 Muon reconstruction

In Section 3.2, the muon reconstruction efficiencies have been computed using values obtained from $Z \to \mu\mu$ events reconstructed in data. Several checks have been performed using also MC simulations, and no discrepancies have been found. Therefore, no systematic uncertainties are placed due to differences between simulation and data. Instead, the systematic uncertainties computed in the efficiencies evaluation for $Z \to \mu\mu$ events are taken into account. These values are listed in Table 4.16 and are propagated in the cross-section formula presented in Equation4.20.

When combined in quadrature, these three contributions will give a systematic uncertainty for muon reconstruction of 0.5%.

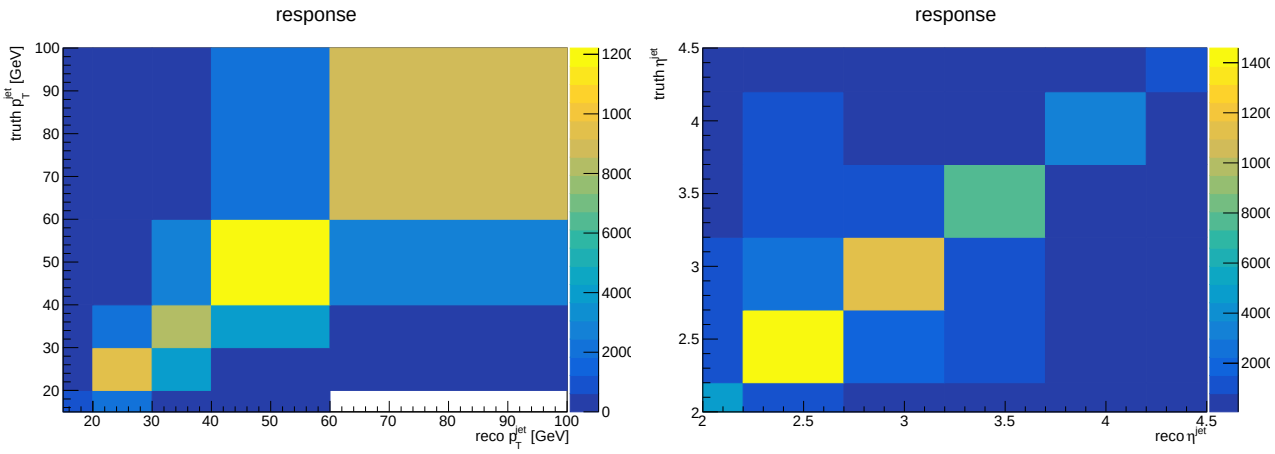


Figure 4.21: (left) Response matrix for $p_{\mathrm{T}}^{\mathrm{jet}}$ for $W+$ 2 jets selection. (right) Response matrix for η^{jet} for $W+$ 2 jets selection.

muon efficiency	systematic uncertainty
Tracking	0.47%
Trigger	$< 0.1\%$
ID	$< 0.1\%$

Table 4.16: Summary of contributions to muon reconstruction systematic uncertainty.

4.9.2 Jet reconstruction and identification

While the muon reconstruction systematic uncertainties affect very little the analysis, the uncertainty associated with differences between data and simulations in the jet reconstruction and identification heavily affects the jet reconstruction efficiency applied to the results of the I_μ fits. To properly detect differences between data and simulated events, $Z(\rightarrow \mu\mu)$+jet events are reconstructed in data and MC simulations: both muons from the Z boson are reconstructed, and the fiducial cuts listed in Table 4.10 are applied. The dominant contribution to the jet reconstruction and identification efficiency $\varepsilon_{\mathrm{reco}}^{\mathrm{jet}}$ comes from the jet identification requirements. As explained in Section 3.3, jet identification is based on cuts applied to several variables: the number of tracks inside the jet pointing to the PV (N_{trk}), the maximum fraction of transverse momentum carried by a single particle inside the jet (mtf) and the maximum transverse momentum carried by a track inside the jet (cpf). Figure 4.22 shows the distributions of data and simulated $Z+$ jet events for N_{trk}, mtf and cpf.

A systematic uncertainty may arise due to the mis-modeling of these observables in MC simulations. Specifically, the idea is to consider how well the regions removed by the jet identification requirements are modeled in simulations. Since these regions are indeed removed by the jet identification requirements, the comparison is done by looking at differences between data and simulations close to the cut value. To quantify the quality of the modeling close to the identification requirements, regions close to the jet identification cuts are considered by tightening the jet identification cuts themselves:

- the cut on N_{trk} is tightened to $N_{\mathrm{trk}} \geq 3$ (default: $N_{\mathrm{trk}} \geq 2$);
- the cut on mtf is tightened to mtf< 0.7 (default: mtf< 0.8);
- the cut on cpf is tightend to cpf> 0.2 (default: cpf> 0.1).

Once the regions to perform comparisons between data and simulations have been defined by the tightening of the jet identification cuts, the level of agreement l between data and simulation is

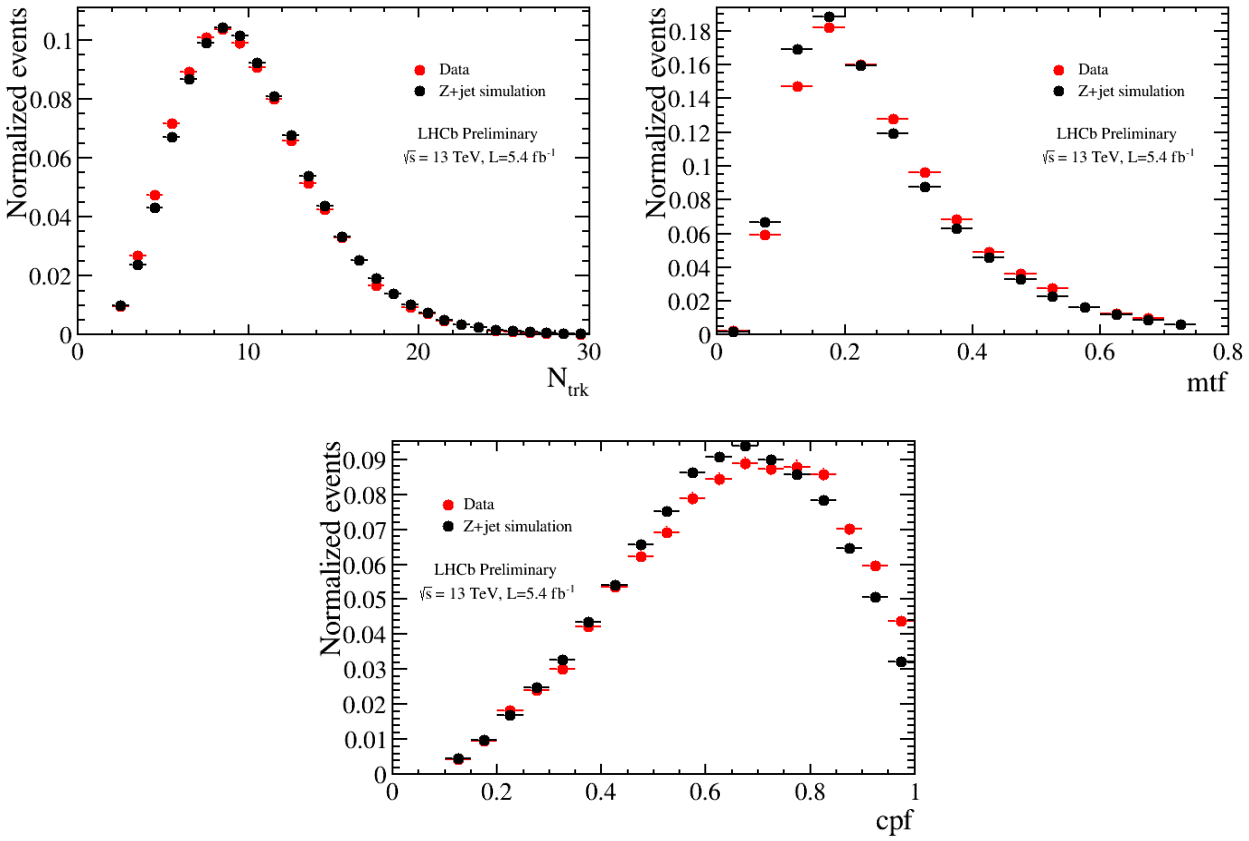


Figure 4.22: Distribution of N_{trk} (upper left plot), mtf (upper right plot) and cpf (lower plot) for $Z \to \mu\mu$+jet in data and simulations.

evaluated as

$$l = \frac{|f_{\text{data}} - f_{\text{sim}}|}{f_{\text{sim}}} \tag{4.11}$$

where f_{data} (f_{sim}) is the fraction of events rejected by tightening the jet identification requirements in data (simulations). The level of agreement l as a function of $p_{\text{T}}^{\text{jet}}$ and η^{jet} is shown in Tables 4.17 and 4.18, both for the full fiducial region and for the $W+$ 1 jet selection binning scheme.

$p_{\text{T}}^{\text{jet}}$ (GeV)	N_{trk}	mtf	cpf
> 20.0	7.6 ± 0.2	3.9 ± 0.2	0.8 ± 0.1
20.0-25.0	10.4 ± 0.7	6.2 ± 0.5	7.9 ± 0.5
25.0-30.0	11.7 ± 0.7	11.1 ± 0.7	6.8 ± 0.5
30.0-40.0	10.3 ± 0.5	2.8 ± 0.3	0.4 ± 0.3
40.0-50.0	0.2 ± 0.5	10.3 ± 0.7	2.2 ± 0.4
60.0-100.0	4.1 ± 0.6	5.3 ± 0.7	2.1 ± 0.6

Table 4.17: Level of agreement l for N_{trk}, mtf, and cpf in the full fiducial region (first line) and as a function of $p_{\text{T}}^{\text{jet}}$ for the $W+$ 1 jet selection.

As a conservative approach, the level of agreement $\tilde{l}$ in the fiducial region and in each $p_{\text{T}}^{\text{jet}}$ and η^{jet} bins is set by the variable which has the worse agreement. This choice is performed independently for each bin. The systematic uncertainty on the jet reconstruction efficiency is then computed as

$$\delta_{\text{ID}} = (1 - \varepsilon_{\text{jet}}) \cdot \tilde{l} \tag{4.12}$$

η^{jet}	N_{trk}	mtf	cpf
2.2-2.7	12.2 ± 0.5	0.3 ± 0.3	21.8 ± 1.2
2.7-3.2	30.8 ± 1.1	20.9 ± 1.1	4.8 ± 0.4
3.2-3.7	22.6 ± 0.8	9.6 ± 0.5	3.4 ± 0.3
3.7-4.2	38.2 ± 1.8	8.3 ± 0.6	2.8 ± 0.5

Table 4.18: Level of agreement l for N_{trk}, mtf, and cpf and as a function of η^{jet} for the $W+$ 1 jet selection.

where ε_{jet} is the jet reconstruction and identification efficiency in the fiducial region or in that specific bin. Basically, the systematic uncertainty is computed by applying the level of agreement $\tilde{l}$ (*i.e.* an estimate of the goodness of the modeling of data by simulations) to the number of rejected events after the application of the jet identification requirements. This computation is performed as a function of $p_{\text{T}}^{\text{jet}}$, η^{jet} and in the full fiducial region. Values between 0.2% and 1.4% are found, which set the systematic uncertainties on the $W+$ jets cross-section coming from the jet identification.

4.9.3 Jet Energy Resolution

The uncertainty associated with differences between data and simulations in the jet energy resolution[1] affects both the unfolding procedure and the efficiency evaluation. To compute the systematic uncertainty related to the jet energy resolution, $Z(\rightarrow \mu\mu)$+jet events reconstructed in data and MC simulations are considered, selected using the same fiducial requirements listed in Table 4.10. $Z(\rightarrow \mu\mu)+$ jet events are rather useful to check for differences between data and simulation for the jet energy resolution: the jet and the Z boson are said to be produced "back-to-back", which means that the direction of the jet is almost opposite to the direction of the Z boson. Therefore, given that p_{T}^{Z} is known with very high precision since the Z boson is reconstructed into two muons, the variable r defined as

$$r = \frac{p_{\text{T}}^{\text{jet}}}{p_{\text{T}}^{Z}} \tag{4.14}$$

is sensitive to differences between data and simulations in the $p_{\text{T}}^{\text{jet}}$ distribution. The jet energy resolution uncertainty is found by smearing the $p_{\text{T}}^{\text{jet}}$ distribution in Z+jet simulated samples using a Gaussian distribution with mean $\mu = 1$ and with maximum width determined by obtaining an agreement between simulation and data for r less than 1σ. This is done for different bins of $p_{\text{T}}^{\text{jet}}$. The smearing factors as a function of $p_{\text{T}}^{\text{jet}}$ and η^{jet} are listed respectively in Tables 4.19 and 4.20. As expected, the smearing factor is rather constant as a function of η^{jet} and is equal to the smearing factor computed in the whole fiducial region, since this effect should not have any dependence on η^{jet}.

Figure 4.23 (left) shows the jet energy resolution in data and simulations in the whole fiducial region as a function of the smearing factor, while Figure 4.23 (right) shows the distance between data and simulations as a function of the smearing factor r. It is evident that for the whole fiducial region, a minimum is found at 0.2.

The smearing factor is applied to $p_{\text{T}}^{\text{jet}}$ and both efficiencies and unfolded distributions are recomputed. For each bin of $p_{\text{T}}^{\text{jet}}$, the relative impact of the systematic uncertainty on the jet energy resolution δ_{jer} is evaluated as

$$\delta_{\text{jer}} = \frac{|N_{\text{jer}} - N|}{N} \tag{4.15}$$

[1]The jet energy resolution is defined as

$$\text{JER} = \frac{E_{\text{true}}^{\text{jet}} - E_{\text{reco}}^{\text{jet}}}{E_{\text{true}}^{\text{jet}}}, \tag{4.13}$$

where $E_{\text{true}}^{\text{jet}}$ ($E_{\text{reco}}^{\text{jet}}$) is the energy of the true (reconstructed) jet.

jet p_{T} (GeV)	Smearing factor
> 20.0	0.19
20.0-25.0	0.14
25.0-30.0	0.19
30.0-40.0	0.18
40.0-50.0	0.23
50.0-60.0	0.20
60.0-100.0	0.19

Table 4.19: Smearing factor for the Gaussian function applied to $p_{\mathrm{T}}^{\mathrm{jet}}$. The smearing factor is shown for the whole fiducial region (first line) and as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$.

jet η	Smearing factor
> 2.2	0.19
2.2-2.7	0.19
2.7-3.2	0.20
3.2-3.7	0.20
3.7-4.2	0.19

Table 4.20: Smearing factor for the Gaussian function applied to $p_{\mathrm{T}}^{\mathrm{jet}}$. The smearing factor is shown for the whole fiducial region (first line) and as a function of η^{jet}.

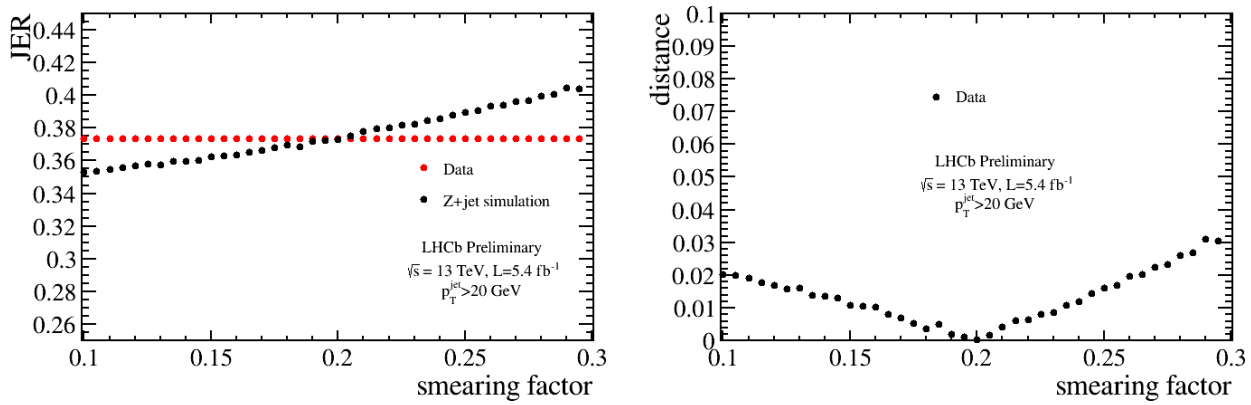


Figure 4.23: Jet energy resolution in data and simulations in the whole fiducial region as a function of the smearing factor (left). Distance between jet energy resolution in data and simulations as a function of the smearing factor (right).

where N_{jer} (N) is the number of unfolded events after having applied efficiencies with (without) the application of the smearing factor. The systematic uncertainty δ_{jer} on the jet energy resolution is found to have an impact between 1% to 4% on the cross-section measurement.

4.9.4 Jet Energy Scale

The uncertainty associated with differences between data and simulations in the jet energy scale[2] affects both the unfolding procedure and the efficiency measurement. As for the jet energy resolution, $Z(\rightarrow \mu\mu)$+jet events in data and from simulations are considered, and they are reconstructed and selected using the fiducial cuts listed in Table 4.10. The uncertainty associated with the jet energy scale is computed as the factor one needs to apply to $p_{\mathrm{T}}^{\mathrm{jet}}$ in simulations to have an agreement with data less than 1σ. Differently from the jet energy resolution, here the focus is on the peak value of the r distribution defined in Equation 4.14. The jet energy scale factors as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ and η^{jet} are listed respectively in Table 4.21 and 4.22. No evident trend is found for η^{jet}, where the scaling factor is similar to the one computed in the whole fiducial region

$p_{\mathrm{T}}^{\mathrm{jet}}$ (GeV)	Scaling factor
> 20.0	1.02
20.0-25.0	0.98
25.0-30.0	1.01
30.0-40.0	1.02
40.0-50.0	1.04
50.0-60.0	1.03
60.0-100.0	1.05

Table 4.21: Scaling factor for the whole fiducial region (first line) and for different bins of $p_{\mathrm{T}}^{\mathrm{jet}}$.

η^{jet}	Scaling factor
2.2-2.7	1.01
2.7-3.2	1.02
3.2-3.7	1.03
3.7-4.2	1.01

Table 4.22: Scaling factor for different bins of η^{jet}.

Figure 4.24 (left) shows the jet energy scale in data and simulations in the whole fiducial region as a function of the scaling factor, while Figure 4.24 (right) shows the distance between data and simulations as a function of the scaling factor. It is evident that for the whole fiducial region, a minimum is found at 1.02.

The scaling factor is applied to $p_{\mathrm{T}}^{\mathrm{jet}}$ and both efficiencies and unfolded distributions are recomputed. For each bin of $p_{\mathrm{T}}^{\mathrm{jet}}$, the relative impact of the systematic uncertainty on the jet energy scale δ_{jer} is evaluated as

$$\delta_{\mathrm{jes}} = \frac{|N_{\mathrm{jes}} - N|}{N} \tag{4.17}$$

where N_{jes} (N) is the number of unfolded events after having applied efficiencies with (without) the application of the scaling factor. The systematic uncertainty on the jet energy scale is found to have an impact between 2% to 5% on the final cross-section measurements. Finally, Figure 4.25 shows the comparison between data and simulations, where the jet energy resolution smearing factor and the jet energy scale factor are applied. The better agreement between data and simulations after the application of smearing and scaling factors is evident.

[2]The jet energy scale is defined as

$$\mathrm{JES} = \frac{E_{\mathrm{MC}}^{\mathrm{jet}}}{E_{\mathrm{data}}^{\mathrm{jet}}}, \tag{4.16}$$

where $E_{\mathrm{MC}}^{\mathrm{jet}}$ ($E_{\mathrm{data}}^{\mathrm{jet}}$) is the energy of a jet coming from simulated samples (data).

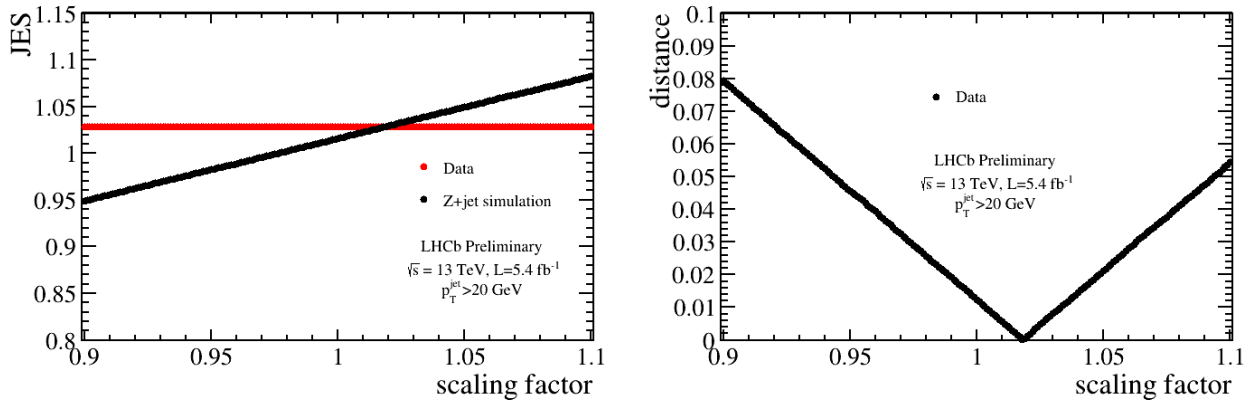


Figure 4.24: Jet energy scale in data (red dots) and $Z(\to \mu\mu)$+ jets simulations in the whole fiducial region as a function of the scaling factor (left). Distance between jet energy scale in data and $Z(\to \mu\mu)$+ jets simulations as a function of the scaling factor (right).

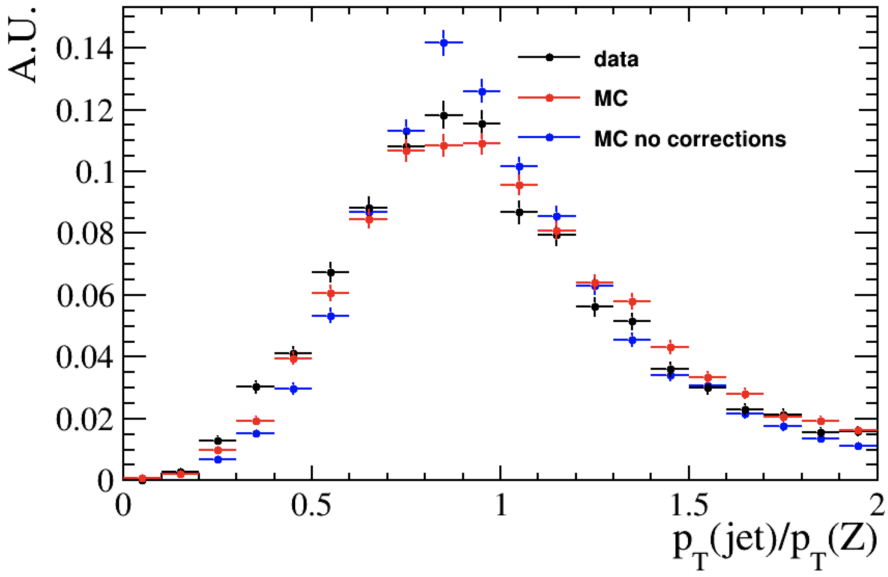


Figure 4.25: $\frac{p_T(\text{jet})}{p_T(Z)}$ in data (black dots) and $Z(\to \mu\mu)$+ jets simulations before (blue dots) and after (red dots) application of scaling and smearing factors.

4.9.5 Selection

In Section 4.6.3, the additional selection requirements applied on top of the fiducial region requirements have been evaluated directly on W+ jets simulated samples. A systematic uncertainty is estimated to take into account possible differences when applying the same requirements in data instead of simulations. To do so, $Z(\to \mu\mu)$+ jets events have been reconstructed and selected in data and simulations, following the standard fiducial requirements listed in Table 4.1. To closely describe the same final state as for W+ jets events, only one muon is reconstructed, randomly chosen between the two muons. The other one is not reconstructed and therefore is treated as missing energy due to the neutrino. The systematic uncertainty in the selection efficiency is evaluated in all bins of the relevant kinematic quantities and in the fiducial region, and it is defined as

$$\delta_{\text{sel}} = \frac{\varepsilon_{data} - \varepsilon_{MC}}{\varepsilon_{MC}} \tag{4.18}$$

where ε_{data} (ε_{MC}) is the selection efficiency evaluated in data (simulations). This systematic uncertainty is found to vary between 1% and 2%. With a similar approach, a systematic uncertainty is associated with differences between simulation and data in the evaluation of the acceptance factor introduced in Section 4.7. This systematic uncertainty is found to vary between 0.1% and 0.5%.

4.9.6 Templates

As described in Section 4.5, the muon isolation I_μ templates used in the fit procedure are taken from simulations. A systematic uncertainty may be associated with differences in the I_μ templates in data and simulations. As for the systematic uncertainty evaluation for the selection efficiency described in Section 4.9.5, $Z(\rightarrow \mu\mu)+$ jets events have been simulated and selected in data using the standard fiducial requirements listed in Table 4.1. Only one muon coming from the Z boson decay is reconstructed, faking the presence of a neutrino by not reconstructing the other muon. The I_μ templates in data and simulations are compared, and weights are computed as

$$w_{\text{template}} = \frac{I_{\mu,data}}{I_{\mu,MC}} \tag{4.19}$$

where $I_{\mu,data}$ ($I_{\mu,MC}$) represents the I_μ template in data (simulations). The weights w_{template} are computed on a bin-by-bin basis. Finally, w_{template} are applied to the I_μ templates used in the fit. Figure 4.26 shows the weights to be applied to I_μ in simulations to match the distribution in data. Particular good agreement is found in the $I_\mu > 0.8$ region, where indeed the EW processes are present.

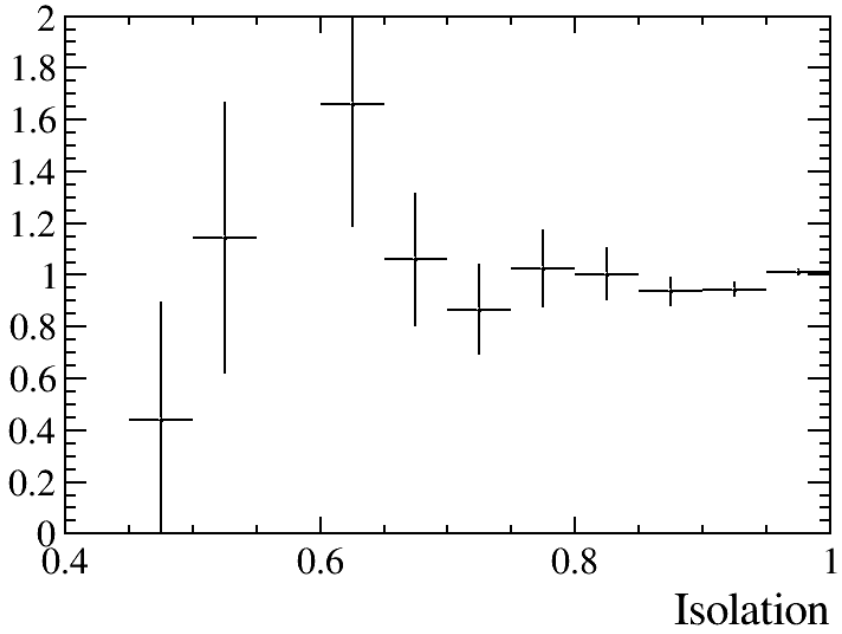


Figure 4.26: Weights resulting from comparison between I_μ in data and $Z(\rightarrow \mu\mu)+$ jets simulations. The region $I_\mu > 0.8$ is found to be well modeled, particularly the last bin.

Once the I_μ templates have been reweighted, the fit to the muon isolation is performed again, and the difference between the new fitted value of $W+$ jets events and the nominal value obtained in Section 4.5 is taken as a systematic uncertainty. This systematic uncertainty is found to be around 1%, since the weights acting on the last I_μ bin are very close to one.

4.9.7 QCD background modeling

The quality and effectiveness of the fit procedure described in Section4.5 relies on the capability of properly modeling the QCD background, which is by far the most important background in this analysis, particularly in the high isolation region. The data-driven modeling of the QCD background has been explained in Section 4.4.2, where it is clear that this is dependent on the reweighting of the muon isolation templates, which happens in the $I_\mu < 0.8$ region. To account for the choice of the cut value for I_μ, this value is changed and the analysis is performed again using the new weights obtained with the new value for the cut on I_μ. The systematic uncertainty is taken as the difference between the new fitted value of $W+$ jets events and the nominal value obtained in Section **??**. The new values for the cut on I_μ are chosen to be 0.75 and 0.85, mainly for two reasons: the values are symmetric to the nominal value of 0.8, and still, a low number of EW events are selected with these cuts, thus not biasing the evaluation of the weights. The systematic uncertainty is found to be $< 0.1\%$, showing the robustness of the reweighting procedure.

4.9.8 Other sources of systematic uncertainties

Other sources of systematic uncertainties are found to be negligible:

- jet fake rate: in the evaluation of the jet reconstruction efficiency described in Section 4.6.2, there might be reconstructed jets that are not associated to any true jets. This contribution is usually referred as "fake jets", and it might affect both the efficiency evaluation and the unfolding procedure. This contribution has been evaluated using the same procedure described in Section 4.6.2, and its effects are found to be negligible.

4.9.9 Luminosity

The luminosity determination has a precision of 2.0% [91]. The corresponding uncertainty is quoted separately from the other sources of systematic uncertainty, and it is propagated in the computation of the cross-section.

4.9.10 Summary of systematic uncertainties

A summary of all the systematic uncertainties in the fiducial region measurement and their impact on the cross-section is shown in Table 4.23. Here Wj (Wjj) refers to $W+$ 1 jet ($W+$ 2 jets) selection, and similarly for W^+ and W^-.

Source	Wj	W^+j	W^-j	Wjj	W^+jj	W^-jj
Muon reconstruction	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%
Jet reconstruction	0.9%	0.9%	0.9%	1.3%	1.3%	1.3%
Jet Energy Resolution	2.3%	2.3%	2.3%	3.2%	3.2%	3.2%
Jet Energy Scale	3.3%	3.3%	3.3%	4.6%	4.6%	4.6%
Selection	1.3%	1.1%	1.7%	1.3%	1.1%	1.7%
Templates	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%
QCD background modeling	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$	$< 0.1\%$
Total systematic uncertainty	4.4%	4.4%	4.6%	6.0%	5.9%	6.1%
Luminosity	2%	2%	2%	2%	2%	2%

Table 4.23: Summary of relative systematic uncertainties for $W+$ 1 jet and $W+$ 2 jets selections.

4.10 Results

The differential cross-section for W production in association with one or more jets in a given phase space bin i for a variable x is given by the formula

$$\frac{d\sigma}{dx}(i) = f_i \cdot \frac{\mathcal{A}_i \cdot N_{\text{fit},i}}{\mathcal{L} \cdot \varepsilon_i} \tag{4.20}$$

where f_i is the correction factor accounting for the unfolding, $\mathcal{A}_i$ is the acceptance factor computed in Section 4.7, $N_{\text{fit},i}$ is the number of signal events fitted, $\mathcal{L}$ is the total integrated luminosity, $\varepsilon_i = \varepsilon_{\text{reco},i} \cdot \varepsilon_{\text{sel},i}$, where $\varepsilon_{\text{reco},i}$ and $\varepsilon_{\text{sel},i}$ are respectively the reconstruction and selection efficiencies, and $x \in (p_{\text{T}}^{\text{jet}}, \eta^{\text{jet}}, \eta^{\mu})$. The total cross-section in the fiducial region is evaluated in a similar way as

$$\sigma = f \cdot \frac{\mathcal{A} \cdot N_{\text{fit}}}{\mathcal{L} \cdot \varepsilon} \tag{4.21}$$

where now all the contributions have been evaluated in the whole fiducial region. The fiducial-region cross-sections for $W+$ 1 jet, $W+$ 2 jets and separately for $W^+ +$ 1 jet, $W^- +$ 1 jet, $W^+ +$ 1 jets and $W^- +$ 2 jets are

$$\begin{aligned}
\sigma(W + 1 \text{ jet}) &= 462.29 \pm 0.48 \pm 20.34 \pm 9.24 \text{ pb} \\
\sigma(W^+ + 1 \text{ jet}) &= 287.73 \pm 0.37 \pm 12.66 \pm 5.75 \text{ pb} \\
\sigma(W^- + 1 \text{ jet}) &= 85.87 \pm 0.30 \pm 8.17 \pm 3.71 \text{ pb} \\
\sigma(W + 2 \text{ jets}) &= 47.69 \pm 0.25 \pm 2.95 \pm 0.95 \text{ pb} \\
\sigma(W^+ + 2 \text{ jets}) &= 29.87 \pm 0.21 \pm 1.85 \pm 0.59 \text{ pb} \\
\sigma(W^- + 2 \text{ jets}) &= 17.88 \pm 0.16 \pm 1.14 \pm 0.35 \text{ pb}
\end{aligned} \tag{4.22}$$

where the first uncertainty is statistical, the second is systematic, and the third is due to the luminosity determination.

The differential cross-section results for the different kinematic variables considered in this analysis, separated for the muon charge, are presented in Tables 4.24, 4.25 and 4.26 for the $W+1$ jet selection, and in Tables 4.27, 4.28 and 4.29 for the $W+2$ jets selection, where the first uncertainty is statistical, the second is systematic, and the third is due to the luminosity determination. The W^-+ jets cross-sections are lower than the W^++ jets ones, as expected since these processes are produced in proton-proton collisions.

$p_{\text{T}}^{\text{jet}}$ (GeV)	σ_{Wj} (pb)	σ_{W^+j} (pb)	σ_{W^-j} (pb)
20-25	$130.69 \pm 0.27 \pm 7.38 \pm 2.61$	$78.84 \pm 0.21 \pm 4.44 \pm 1.57$	$51.89 \pm 0.17 \pm 2.92 \pm 1.03$
25-30	$83.75 \pm 0.20 \pm 3.72 \pm 1.67$	$50.55 \pm 0.157 \pm 2.19 \pm 1.01$	$33.19 \pm 0.13 \pm 1.55 \pm 0.66$
30-40	$49.58 \pm 0.11 \pm 1.47 \pm 0.99$	$29.69 \pm 0.08 \pm 0.87 \pm 0.59$	$19.76 \pm 0.067 \pm 0.61 \pm 0.39$
40-50	$27.62 \pm 0.07 \pm 0.94 \pm 0.55$	$16.82 \pm 0.06 \pm 0.72 \pm 0.33$	$11.34 \pm 0.05 \pm 0.44 \pm 0.22$
50-60	$17.23 \pm 0.06 \pm 0.95 \pm 0.34$	$10.57 \pm 0.04 \pm 0.53 \pm 0.21$	$6.36 \pm 0.03 \pm 0.39 \pm 0.12$
60-100	$6.19 \pm 0.02 \pm 0.37 \pm 0.12$	$3.91 \pm 0.01 \pm 0.23 \pm 0.07$	$2.33 \pm 0.01 \pm 0.14 \pm 0.05$

Table 4.24: Measured cross-sections for $W+$ 1 jet as a function of $p_{\text{T}}^{\text{jet}}$.

4.10.1 Comparison with theoretical predictions

The results obtained are compared with theoretical predictions. In particular, two theoretical predictions are computed using MadGraph:

η^{jet}	σ_{Wj} (pb)	σ_{W^+j} (pb)	σ_{W^-j} (pb)
2.2-2.7	$173.47 \pm 0.28 \pm 7.97 \pm 3.46$	$102.54 \pm 0.22 \pm 4.81 \pm 2.05$	$72.63 \pm 0.18 \pm 3.26 \pm 1.45$
2.7-3.2	$141.54 \pm 0.25 \pm 7.21 \pm 2.83$	$85.76 \pm 0.20 \pm 4.28 \pm 1.71$	$56.06 \pm 0.16 \pm 2.97 \pm 1.12$
3.2-3.7	$103.73 \pm 0.21 \pm 4.66 \pm 2.07$	$65.43 \pm 0.16 \pm 2.94 \pm 1.30$	$38.17 \pm 0.13 \pm 1.71 \pm 0.76$
3.7-4.2	$59.29 \pm 0.17 \pm 3.21 \pm 1.18$	$37.31 \pm 0.13 \pm 1.90 \pm 0.74$	$21.66 \pm 0.10 \pm 1.32 \pm 0.43$

Table 4.25: Measured cross-sections for $W+$ 1 jet as a function of η^{jet}.

η^{μ}	σ_{Wj} (pb)	σ_{W^+j} (pb)	σ_{W^-j} (pb)
2.0-2.5	$137.08 \pm 0.24 \pm 6.30 \pm 2.74$	$85.21 \pm 0.19 \pm 4.01 \pm 1.70$	$51.99 \pm 0.15 \pm 2.28 \pm 1.03$
2.5-3.0	$134.23 \pm 0.24 \pm 5.90 \pm 2.68$	$85.70 \pm 0.19 \pm 3.68 \pm 1.71$	$48.48 \pm 0.15 \pm 2.23 \pm 0.96$
3.0-3.5	$107.46 \pm 0.22 \pm 4.62 \pm 2.14$	$65.36 \pm 0.17 \pm 2.81 \pm 1.31$	$42.69 \pm 0.13 \pm 1.83 \pm 0.85$
3.5-4.5	$52.85 \pm 0.11 \pm 2.37 \pm 1.05$	$28.79 \pm 0.08 \pm 1.27 \pm 0.57$	$23.88 \pm 0.07 \pm 1.05 \pm 0.47$

Table 4.26: Measured cross-sections for $W+$ 1 jet as a function of η^{μ}.

$p_{\text{T}}^{\text{jet}}$ (GeV)	σ_{Wjj} (pb)	σ_{W^+jj} (pb)	σ_{W^-jj} (pb)
20-30	$11.33 \pm 0.12 \pm 0.75 \pm 0.26$	$6.54 \pm 0.09 \pm 0.43 \pm 0.13$	$4.75 \pm 0.08 \pm 0.32 \pm 0.09$
30-40	$10.90 \pm 0.12 \pm 0.45 \pm 0.21$	$6.81 \pm 0.09 \pm 0.28 \pm 0.13$	$4.07 \pm 0.08 \pm 0.17 \pm 0.08$
40-60	$6.59 \pm 0.06 \pm 0.43 \pm 0.13$	$4.33 \pm 0.04 \pm 0.26 \pm 0.08$	$2.29 \pm 0.04 \pm 0.16 \pm 0.04$
60-100	$2.23 \pm 0.02 \pm 0.18 \pm 0.04$	$1.45 \pm 0.02 \pm 0.12 \pm 0.02$	$0.78 \pm 0.01 \pm 0.06 \pm 0.01$

Table 4.27: Measured cross-sections for $W+$ 2 jets as a function of $p_{\text{T}}^{\text{jet}}$.

η^{jet}	σ_{Wjj} (pb)	σ_{W^+jj} (pb)	σ_{W^-jj} (pb)
2.2-2.7	$15.69 \pm 0.15 \pm 0.99 \pm 0.31$	$9.61 \pm 0.11 \pm 0.61 \pm 0.19$	$6.10 \pm 0.09 \pm 0.38 \pm 0.12$
2.7-3.2	$13.33 \pm 0.13 \pm 0.92 \pm 0.26$	$8.56 \pm 0.10 \pm 0.59 \pm 0.17$	$4.83 \pm 0.08 \pm 0.33 \pm 0.09$
3.2-3.7	$10.61 \pm 0.11 \pm 0.66 \pm 0.21$	$6.91 \pm 0.08 \pm 0.43 \pm 0.13$	$3.81 \pm 0.07 \pm 0.24 \pm 0.07$
3.7-4.2	$7.72 \pm 0.09 \pm 0.56 \pm 0.15$	$4.77 \pm 0.070 \pm 0.33 \pm 0.09$	$2.52 \pm 0.05 \pm 0.19 \pm 0.05$

Table 4.28: Measured cross-sections for $W+$ 2 jets as a function of η^{jet}.

η^{μ}	σ_{Wjj} (pb)	σ_{W^+jj} (pb)	σ_{W^-jj} (pb)
2.0-2.5	$13.46 \pm 0.13 \pm 0.82 \pm 0.26$	$8.74 \pm 0.10 \pm 0.55 \pm 0.17$	$4.69 \pm 0.08 \pm 0.27 \pm 0.09$
2.5-3.0	$12.39 \pm 0.13 \pm 0.74 \pm 0.24$	$8.26 \pm 0.10 \pm 0.48 \pm 0.16$	$4.08 \pm 0.08 \pm 0.25 \pm 0.08$
3.0-3.5	$10.43 \pm 0.11 \pm 0.62 \pm 0.20$	$6.54 \pm 0.08 \pm 0.38 \pm 0.13$	$3.92 \pm 0.07 \pm 0.23 \pm 0.07$
3.5-4.5	$5.58 \pm 0.06 \pm 0.33 \pm 0.11$	$3.07 \pm 0.04 \pm 0.18 \pm 0.06$	$2.54 \pm 0.03 \pm 0.15 \pm 0.05$

Table 4.29: Measured cross-sections for $W+$ 2 jets as a function of η^{μ}.

- a LO prediction which includes the processes $W + N$ jets, with $N = 0, 1, 2, 3$. Pythia8 is used for parton showering and hadronization, and the matrix element computation is matched to the parton showering using the k_{T}-MLM scheme with the merging scale set at 20 GeV;
- a NLO prediction hich includes the processes $W + N$ jets, with $N = 0, 1, 2$ interfaced with Pythia8. The matrix element computation is matched to the parton showering using the FxFx merging scheme, with a merging scale parameter set to 20 GeV.

For both theoretical predictions, the NNPDF2.3 PDF [51] set has been used. For these theoretical predictions, three uncertainties are considered:

- the uncertainty δ_{scale} due to missing higher-order contributions is estimated by varying the

renormalization scale μ_R and the factorization scale μ_F independently by a factor of two around their nominal values, keeping the restriction that $0.5 < \mu_R/\mu_F < 2$. A total of 6 points are evaluated, and the scale uncertainty is given as the maximum deviation above and below the central value observed when using the different scale choices;

- the uncertainty δ_{PDF} due to the description of the chosen PDF set is obtained by re-computing the prediction for each observable with the 100 replicas provided and calculating the uncertainty according to the usual standard deviation formula;
- the uncertainty δ_{α_s} due to the choice of the value of the strong coupling constant $\alpha_s(M_Z) = 0.118$, is determined by re-evaluating the observables at different values of α_s, namely $\alpha_s(M_Z) = 0.117$,0.119.

Finally, the total error for the theoretical prediction is obtained following the recommendation of the Higgs cross-section working group [92] as

$$\delta_{\text{th}} = \sqrt{\delta^2_{\text{PDF}} + \delta^2_{\alpha_s}} + \delta_{\text{scale}} \tag{4.23}$$

Results are compared graphically with theoretical predictions from MadGraph. The comparison between the measured cross-sections in the fiducial region and the corresponding theoretical expectations are shown in Figure4.27: good agreement is found between the measured and the theoretical cross-section. The measurements are definitely limited by systematic uncertainties.

Figures 4.28 and 4.30 show the differential cross-section as a function of $p_{\text{T}}^{\text{jet}}$, η^{jet} and η^{μ} respectively for $W+1$ jet and $W+2$ jets, while Figures 4.29 and 4.31 show the differential cross-section separately for W^+ and W^- as a function of $p_{\text{T}}^{\text{jet}}$, η^{jet} and η^{μ} respectively for $W+1$ jet and $W+2$ jets. Here, the measured cross-sections are compatible with the theoretical expectations within the errors. A small deviation is found for low and high values of the $W+$ 2 jets distributions; the same deviation is found also separately for $W^+ +$ 2 jets and $W^- +$ 2 jets events.

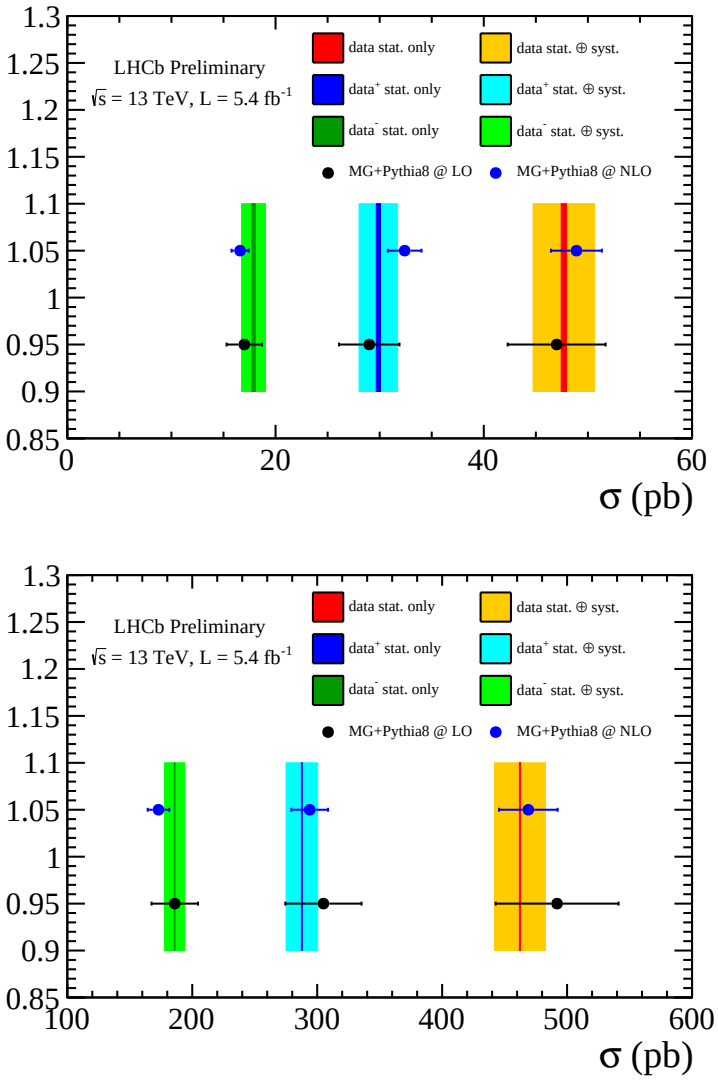


Figure 4.27: Comparisons between measured cross-sections for $W+$ 1 jet (upper plot) and $W+$ 2 jets (lower plot) and theoretical predictions.

4.11 Final considerations

The analysis presented in this Chapter describes the first measurement of $W+$ jets events in the forward region at a center-of-mass energy $\sqrt{s} = 13$ TeV. The measurement has been performed in the fiducial region defined in Table 4.1, as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ}, and depending on the charge of the muon. Events are required to have at least one jet (two jets) in the LHCb acceptance for the $W+$ 1 jet ($W+$ 2 jets) selection. Results are in agreement with SM predictions obtained with MadGraph, and the only slight discrepancy is found in the η^{jet} distribution for the $W+$ 2 jets selection. Few more considerations:

- this analysis extends the results already obtained by a previous LHCb measurement of $W+$ jets cross-section at $\sqrt{s} = 8$ TeV [49];
- to properly check for differences in the matrix element generation, results should be checked against other MC generators, e.g. Powheg [28] and Alpgen [93];

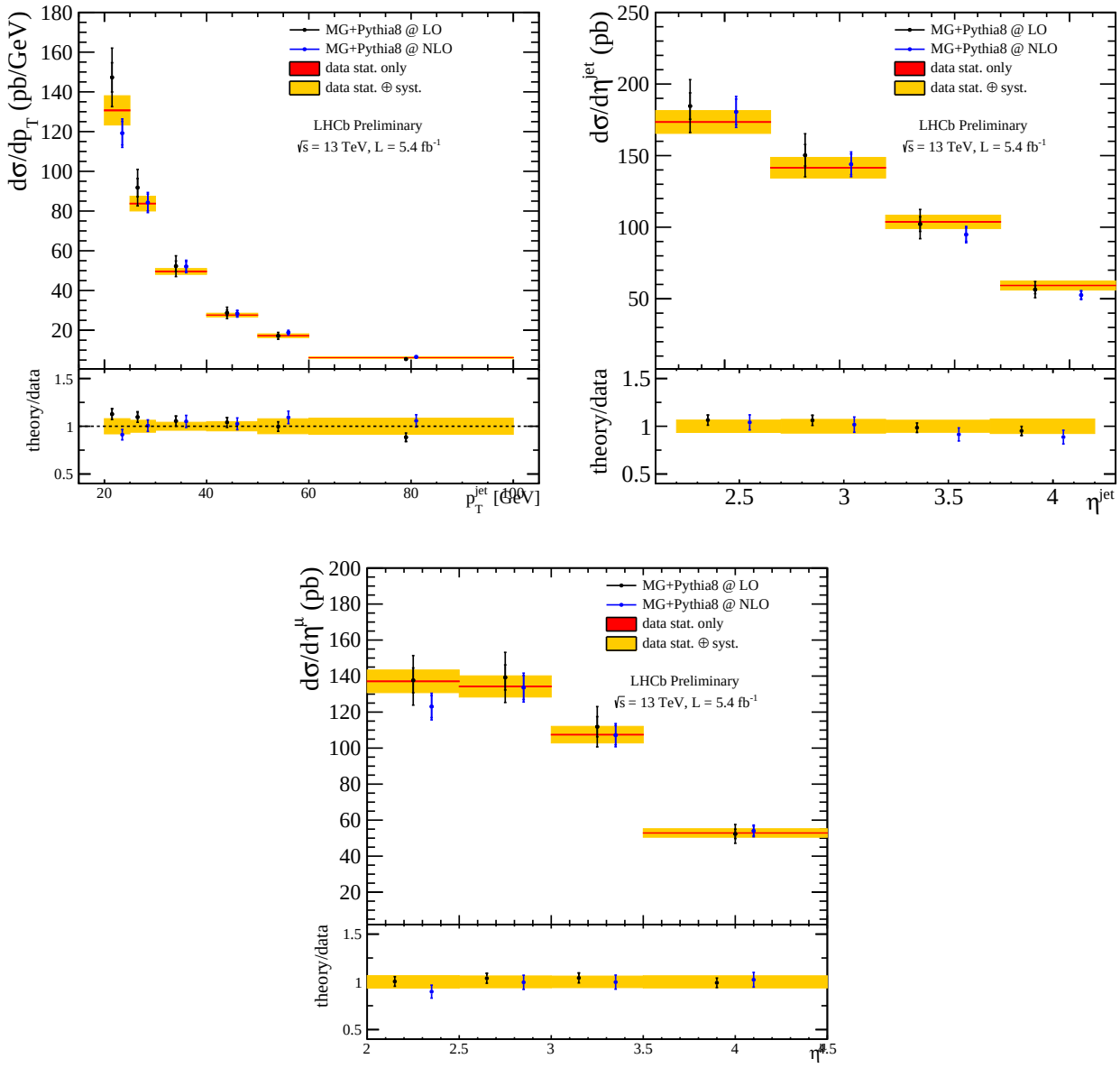


Figure 4.28: Differential cross-section measurement for $W+$ 1 jets as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{T} (upper right plot), and η^{μ} (lower plot) and comparison with theoretical predictions.

- this measurement can be used to update PDFs fits, providing a better understanding of the forward region of proton-proton collisions. Comparisons with other PDFs sets can be obtained by extending the theoretical predictions computed in this analysis;
- this analysis paves the way to several measurements: $W + b$-jet, $W + c$-jet, $W + b\bar{b}$ and $W + c\bar{c}$ cross-section measurements [44, 47] can be performed in the same fiducial region and as a function of the same kinematic observables, following an analysis procedure similar to the one presented in this Chapter. The only additional step would be the identification of the flavor of the quark generating the jet, using the tools described in Section 3.4;

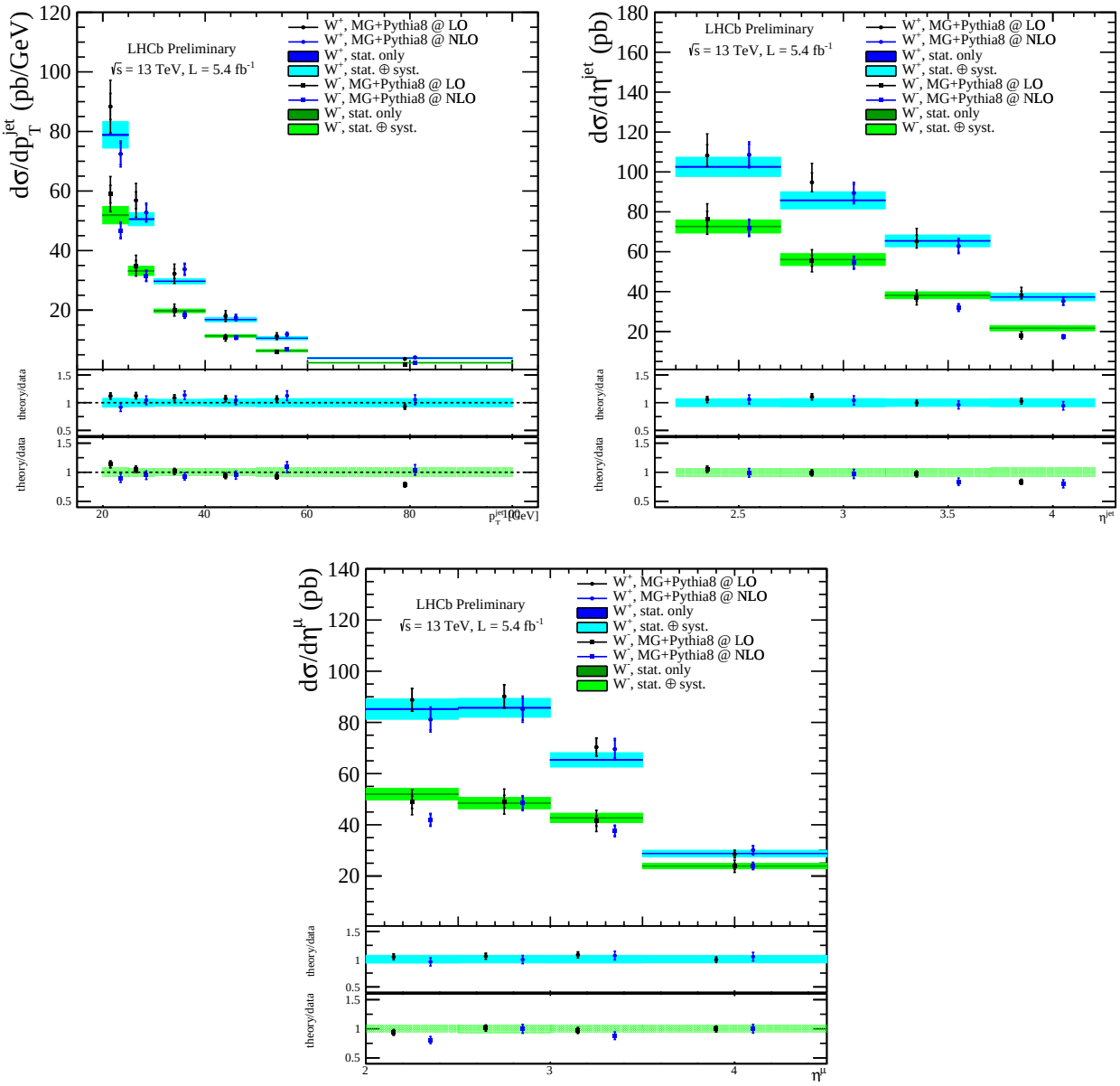


Figure 4.29: Differential cross-section measurement for W^++ 1 jet and W^-+ 1 jet as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{T} (upper right plot), and η^{μ} (lower plot) and comparison with theoretical predictions.

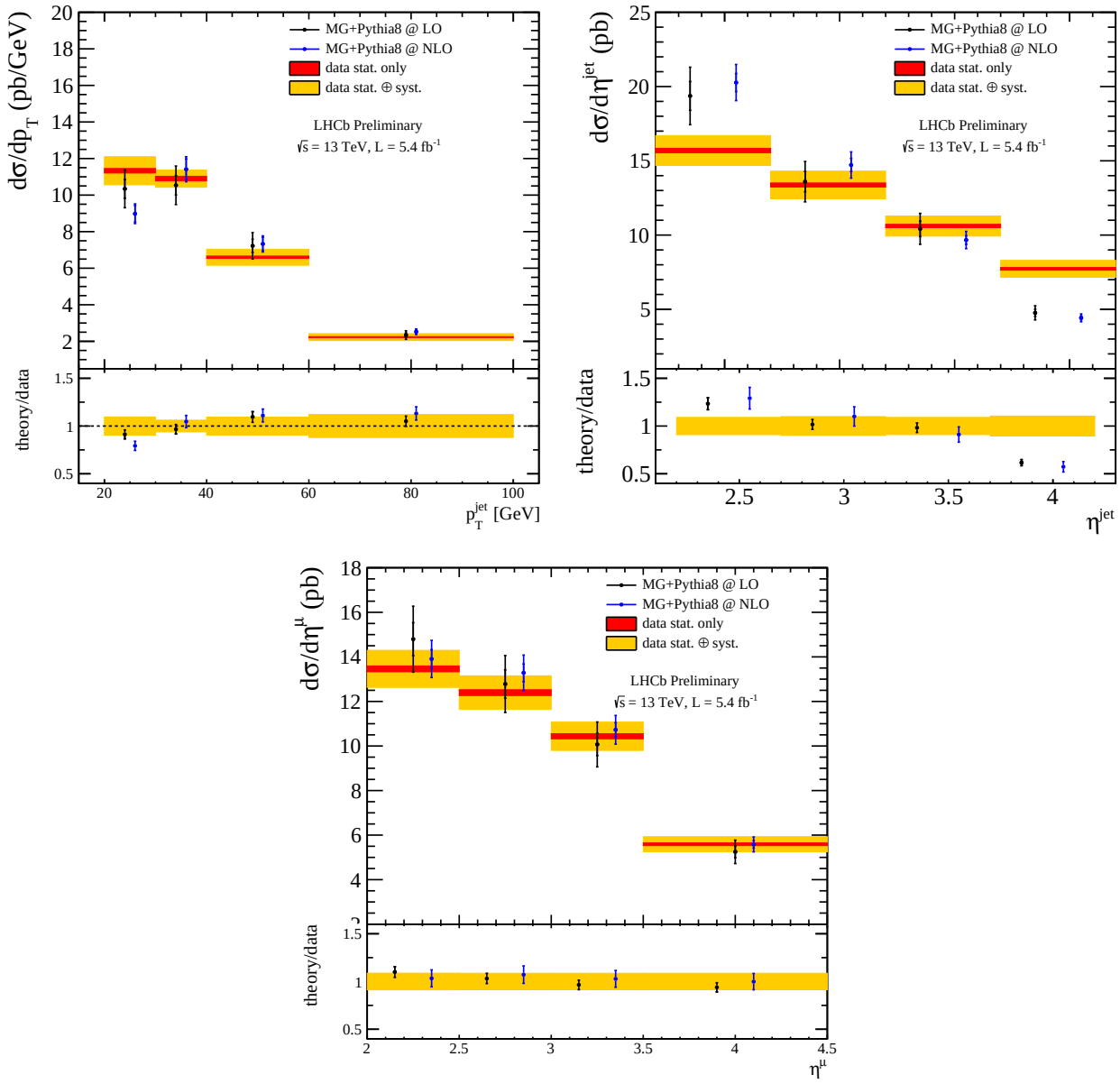


Figure 4.30: Differential cross-section measurement for $W+$ 2 jets as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{T} (upper right plot), and η^{μ} (lower plot) and comparison with theoretical predictions.

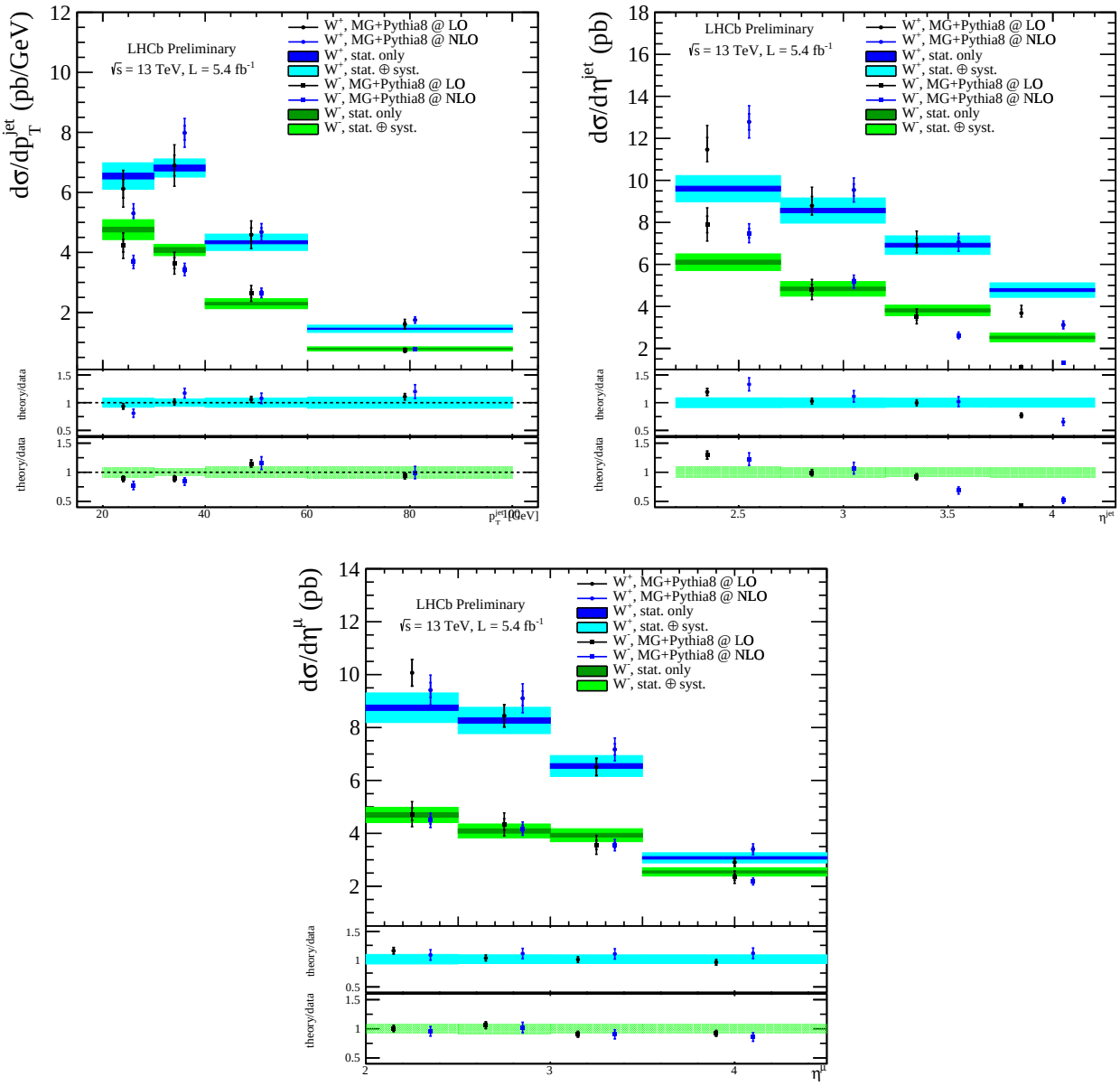


Figure 4.31: Differential cross-section measurement for W^++ 2 jets and W^-+ 2 jets as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ (upper left plot), η^{T} (upper right plot), and η^{μ} (lower plot) and comparison with theoretical predictions.

Chapter 5

Measurement of $W(\rightarrow \mu\nu)W(\rightarrow \text{jets})$ production

At the LHCb experiment, WW production has never been measured: an attempt using the fully leptonic $e\mu$ channel has led to the conclusion that such a measurement will be possible in the next data-taking campaign, using roughly ~ 16 fb^{-1} [50], as explained in Section 1.7.3. While using a final state with a very low background, due to the opposite charge of the two leptons which mainly leaves $t\bar{t}$ events as the main background, no multivariate techniques have been used to further discriminate between signal and backgrounds. At the moment of writing this thesis, a proposal for measuring WW production in the $e\mu$ channel using Machine Learning techniques is in place. In this Chapter, a measurement of WW is presented: here one W boson is required to decay hadronically into two jets, while the other W boson is reconstructed from its semileptonic decay into a muon and a neutrino. An approach based on Machine Learning tools is presented to discriminate WW production from other backgrounds, in particular $W+$. jets The main backgrounds for this analysis are $W+$jets, $Z+$ jets, and events coming from QCD processes, which have already been studied in Chapter 4. The full Run 2 dataset is analyzed, corresponding to a total integrated luminosity $\mathcal{L} = 5.4$ fb^{-1}. This Chapter has the following structure:

- a brief description of the analysis strategy is presented in Section 5.1;
- Section 5.2 describes the signal and background selections for this specific analysis, stemming from the selections already studied in Section 4.2;
- Section 5.3 describes the Deep Neural Network classifier and the input features used to separate WW events from $W+$ jets events, which is the main background;
- Section 5.4 describes the fit procedure used to extract the WW yield, while Sections 5.5 and 5.6 compute the reconstruction efficiencies and the acceptance factor, similarly to what done in Sections 4.6 and 4.7;
- finally, WW cross-section result and final considerations are reported in Section 5.8.

5.1 Analysis strategy

The main purpose of this analysis is to estimate the number of $W(\rightarrow \mu\nu))W(\rightarrow$ jets) events reconstructed in a fiducial region defined by the cuts listed in Table 5.1. In the rest of this chapter, writing "WW" will correspond to the $W(\rightarrow \mu\nu))W(\rightarrow$ jets) channel. The analysis strategy to estimate the number of WW events is divided into the following steps:

- data and simulated samples are selected using the same fiducial cuts applied to the $W+$ jets analysis. On top of these cuts, the requirement that the signal muon is well isolated is considered, in order to reject QCD events;

- an ML classifier, namely a Deep Neural Network, is trained to distinguish between WW and $W+$ jets events, the latter being the highest background amongst the other EW backgrounds in the high isolation range. Variables coming from the kinematics of the event objects are used as input features of the classifier. The performance of the classifier is studied and it is applied to data and simulated samples.

- the QCD contribution is evaluated using a control region, defined by inverting the cut on the muon impact parameter IP_μ and therefore selecting events that do not come from EW processes. The output of the classifier applied to the QCD is reweighted to match the distribution in the signal region;

- a binned maximum likelihood fit is performed, simultaneously fitting the output of the classifier and the isolation I_μ of the muon;

- finally, systematic uncertainties are evaluated and comparisons with theoretical predictions are presented.

A flowchart summarising the analysis steps is shown in Figure 5.1.

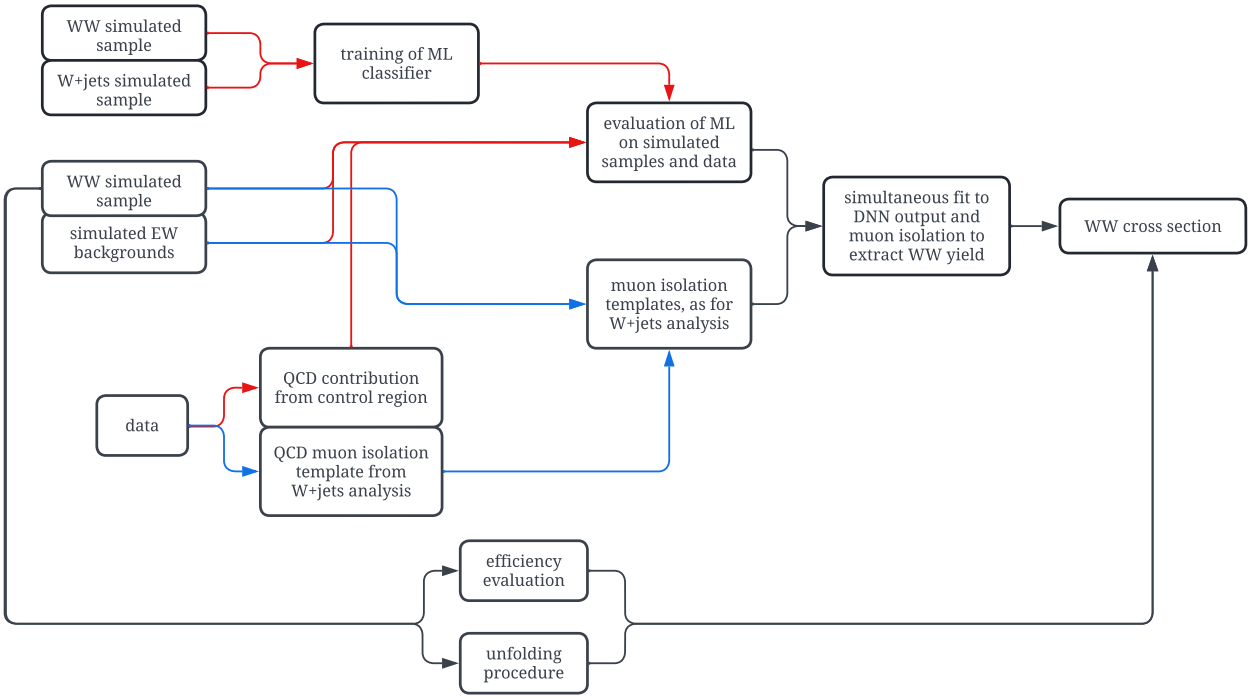


Figure 5.1: Flowchart of the analysis. Red lines describe the necessary step to train the Deep Neural Network to perform classification between WW and $W+$ jets events, while blue lines describe the steps to obtain the muon isolation templates. Finally, black lines show the steps that lead to the simultaneous fit and to the cross-section measurement.

5.2 Signal and backgrounds selection

The signal selection for this specific process strictly follows the selection cuts already described in Section 4.2: the signature is one muon with $p_{\mathrm{T}} > 20$ GeV and at least two jets with $p_{\mathrm{T}} > 20$ GeV are required to be in the LHCb acceptance. The main cuts defining the fiducial region are listed in Table 5.1.

Object	Fiducial Cut
Muons	$p_{\mathrm{T}} > 20$ GeV $2.0 < \eta < 4.5$
Jets	$p_{\mathrm{T}} > 20$ GeV $2.2 < \eta < 4.2$
Event	$p_{\mathrm{T}}(\mu + \text{jets}) > 20$ GeV

Table 5.1: Fiducial cuts applied to the final state objects for the WW analysis.

As already studied in the previous chapter, a muon coming from the decay of a vector boson tends to be more isolated, *i.e.* the activity contained inside a jet of radius $R = 0.4$ surrounding the muon is lower with respect to a muon coming from hadron decay-in-flight or heavy flavor decays. The muon isolation I_μ is defined by Equation 4.6. In the $W+$ jets analysis, the whole I_μ is considered, in order to properly constrain the contribution coming from QCD events, which are mainly present in the $I_\mu < 0.8$ region. In this analysis instead, given that the amount of WW events to be measured is rather low, a cut on I_μ is considered to reduce the QCD contribution. Therefore, on top of the fiducial cuts in Table 5.1, the muon isolation I_μ is required to be greater than 0.8, where the EW processes are dominant. Table 5.2 shows the fraction of events (both for QCD and EW processes) that have $I_\mu > 0.8$.

Process	$I_\mu > 0.8$	$I_\mu < 0.8$
data	0.57	0.43
WW	0.98	0.02
$W(\to \mu\nu)+$ jets	0.98	0.02
$Z(\to \mu\mu)+$ jets	0.98	0.02
$t\bar{t}$	0.95	0.05
WZ	0.98	0.02
single top	0.98	0.98
QCD	0.28	0.72

Table 5.2: Fraction of events with $I_\mu > 0.8$ and $I_\mu < 0.8$ in data and in MC simulations. It is evident that the $I_\mu > 0.8$ requirement mainly removes the QCD background.

5.3 Deep Neural Network for signal discrimination

The number of expected WW events in the μjj channel is rather low compared to the other backgrounds present: similarly to what is done in Section 1.7.3, a simple calculation of the global significance $\mathcal{S}$ indeed gives

$$\mathcal{S} = \frac{N_{WW}}{\sqrt{N_{WW} + N_{backgrounds}}} = 1.8 \tag{5.1}$$

where N_{WW} ($N_{backgrounds}$) indicates the number of WW (background) events. Therefore it is almost impossible to measure WW events without applying more stringent selections that keep WW events

while reducing the background contributions. While at CMS and ATLAS, further discrimination is achieved by posing more stringent cuts on the kinematics of the events [43] (e.g. on the p_{T} of the jets or the p_{T} of the W boson), this is not possible at the LHCb experiment, particularly because the muon and the jets p_{T} ranges are quite limited by the acceptance of the detector. To achieve further discrimination between WW and background contributions, Machine Learning techniques can be used. In this analysis, a Deep Neural Network (DNN) [94] has been trained to separate WW events from $W+$ jets events, which is the most important background in the $I_{\mu} > 0.8$ region, as shown in Section 4.5. The input features for the DNN are the following:

- p_{T}, η and ϕ of the two jets;
- p_{T} and η of the muon;
- $\Delta\eta$, $\Delta\phi$ and ΔR between each jet and the muon;
- the invariant mass $m_{\mathrm{inv}}(jj)$ of the two jets;
- the three components of the momentum P_x, P_y, P_z, P_{T} and η of the jets system $J = j_0 + j_1$;
- $\Delta\eta$, $\Delta\phi$ and ΔR between the jets system J and the muon;
- the invariant mass M_{inv} of the muon and the two jets;
- the total transverse momentum of the muon and the two jets.

The input variables are shown in Figure 5.3 for WW and relevant EW backgrounds: while some variables show great discrimination between WW and other backgrounds, some others do not, and that is why variables should be combined with an ML classifier to enhance the discrimination power of the input features.

The DNN structure consists of a 1-dimensional convolutional layer followed by a dense structure with three layers. The depth of the convolutional layer is set to 32, while the number of nodes for the three layers is set to 64, 64, and 48 respectively. In the end, a "softmax" activation function is used to get the output of the classification. The DNN is trained using the ADAM [95] optimizer, for a maximum of 200 epochs. A "patience" checkpoint is set to 25, therefore the training procedure is stopped if the loss function does not decrease for 25 epochs. A batch size equal to 8192 is used, and the training is performed on an NVIDIA GeForce GTX TITAN X GPU. The DNN is trained on samples of WW and $W+$ jets events. A total amount of 500000 events are used, equally distributed between WW and $W^{\pm}+$ jets events. The total dataset is divided into training, validation and testing datasets, to check for overfitting. The DNN outputs a value in the range $[0, 1]$, and the DNN output is matched to the probability for an event to be classified as a WW or a $W+$ jets event: if the DNN output is close to 0 (1), then the event is more likely to be a WW ($W^{\pm}+$ jets) event. Figure 5.6 shows the loss function and the training accuracy as a function of the number of epochs during the training: no signal of overfitting is found, assuring the quality of the training procedure. Figure 5.7 shows the ROC curve [96] and the confusion matrix: a fair separation is obtained between WW and $W+$ jets events.

The DNN output for the considered EW processes is shown in Figure 5.8: the separation between WW and $W+$ jets events is evident, but it is not enough to perform the analysis by simply cutting selecting events in a given region of the DNN output (e.g. for values lower than 0.5). Moreover, since the DNN is optimized to discriminate between WW and $W+$ jets events, the separation between WW and other EW processes (such as WZ and $t\bar{t}$) is fairly poor. Therefore a fit to the output DNN is performed, as it is explained in Section 5.4.

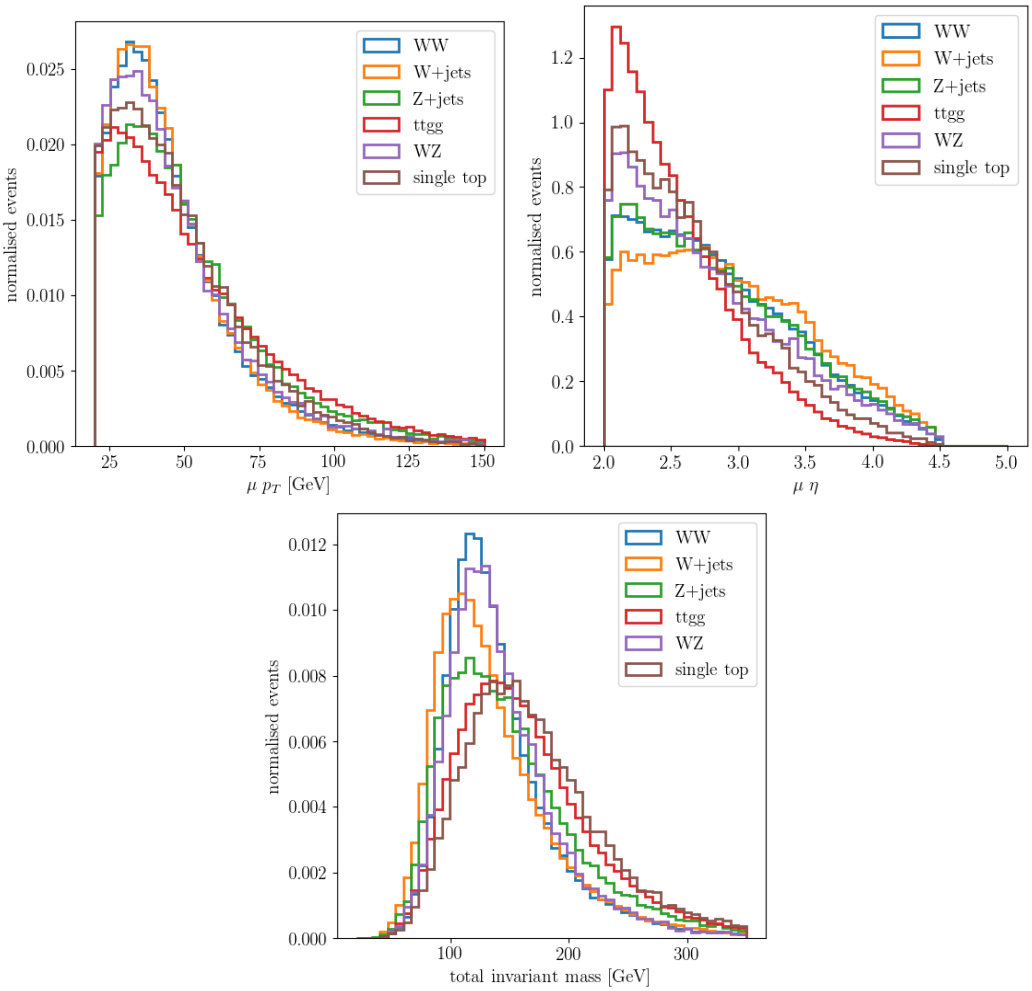


Figure 5.2: Muon p_{T} (upper left plot), muon η (upper right plot), and $(\mu + j_0 + j_1)$ invariant mass M_{inv} (lower plot). Distributions are normalized to one.

5.3.1 Modeling of QCD DNN output

While the DNN output can be easily obtained for all the EW processes, as they are obtained from MC simulations, for QCD events this is not possible: as for the determination of the $W+$ jets yield described in Section 4.5, it is not possible to evaluate the templates of QCD events in the $I_\mu > 0.8$ region, as this would need an awful amount of simulated events. In that specific situation, a data-driven technique was used by defining signal and control regions. Here a similar approach is used. Signal (SR) and control (CR) regions are defined by exploiting the cuts on I_μ and on the muon impact parameter IP_μ, The following SR and CR are defined:

- SR is defined for $I_\mu > 0.8$ and $\mathrm{IP}_\mu < 0.04$;
- CR is defined for $I_\mu > 0.8$ and $\mathrm{IP}_\mu > 0.04$.

The choice of using IP_μ to define the SR and CR is the following: for $\mathrm{IP}_\mu > 0.04$ no EW events are present; in this way, the evaluation of the DNN in the CR can be used as a good proxy for the description of QCD events in SR. To check for the quality of the modeling, two other CRs are defined in the following way:

- CR1 is defined for $I_\mu < 0.8$ and $\mathrm{IP}_\mu < 0.04$;

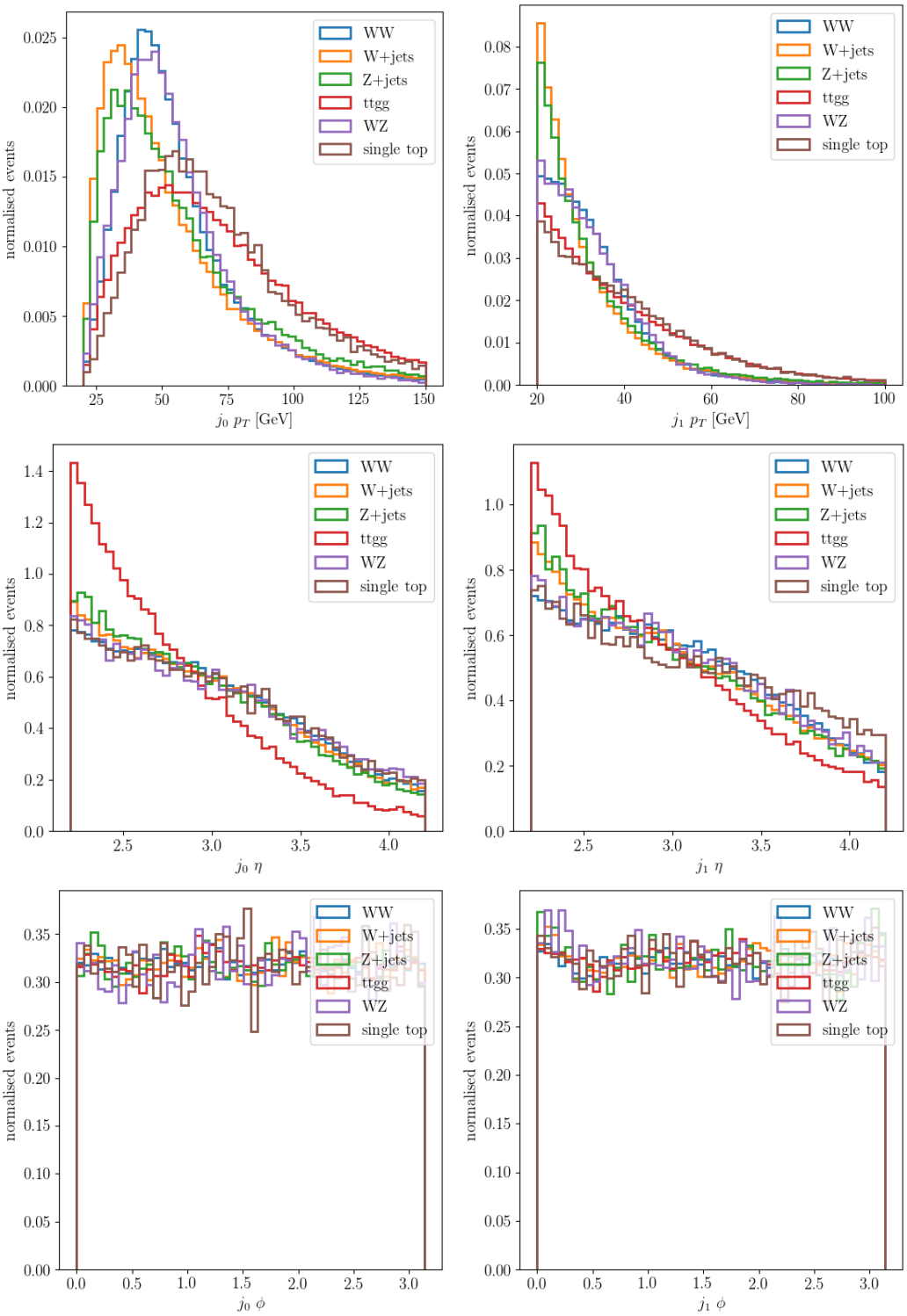


Figure 5.3: p_T, η and ϕ distributions for leading jet j_0 (left column) and sub-leading jet j_1 (right column). Distributions are normalized to one.

- CR2 is defined for $I_\mu < 0.8$ and $\text{IP}_\mu > 0.04$;

The motivation behind this choice is the following: the idea is to model the DNN output of QCD events in SR by evaluating the DNN in CR. Since differences may arise between SR and CR due to the

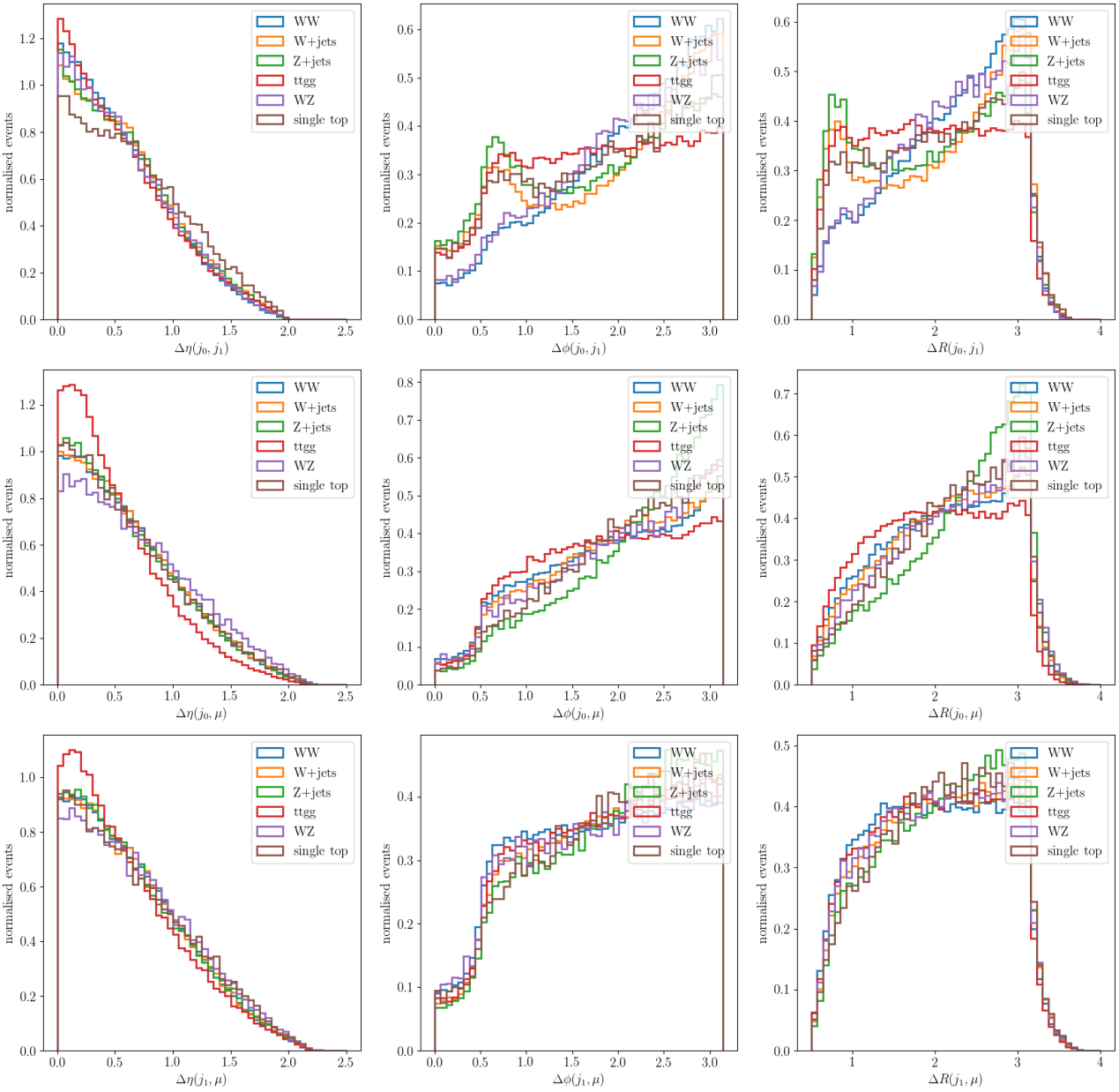


Figure 5.4: $\Delta\eta$, $\Delta\phi$ and ΔR distributions for jets (first row), leading jet and muon (second row), and sub-leading jet and muon (third row). Distributions are normalized to one.

IP_μ requirement, these differences are quantified by checking how well CR2 is describing CR1. Since I_μ and other variables related to it are not used in the classification of WW events, it is reasonable to assume that the relation between SR and CR is well described by the relation between CR1 and CR2. The definition of SR, CR, CR1, and CR2 in the (I_μ, IP_μ) space is represented in Figure 5.9. Both CR1 and CR2 should not contain any EW events, given that EW processes tend to have higher values of muon isolation, as shown in Figure 4.5. Therefore, the quality of modeling SR by using CR can be assessed by fitting CR1 using CR2, and correcting for any differences. The result of a maximum-likelihood bin fit is shown in Figure 5.10. A few considerations can be drawn from this fit:

- no EW backgrounds are fitted, which is expected given that the EW contamination in CR1 and CR2 is expected to be negligible;

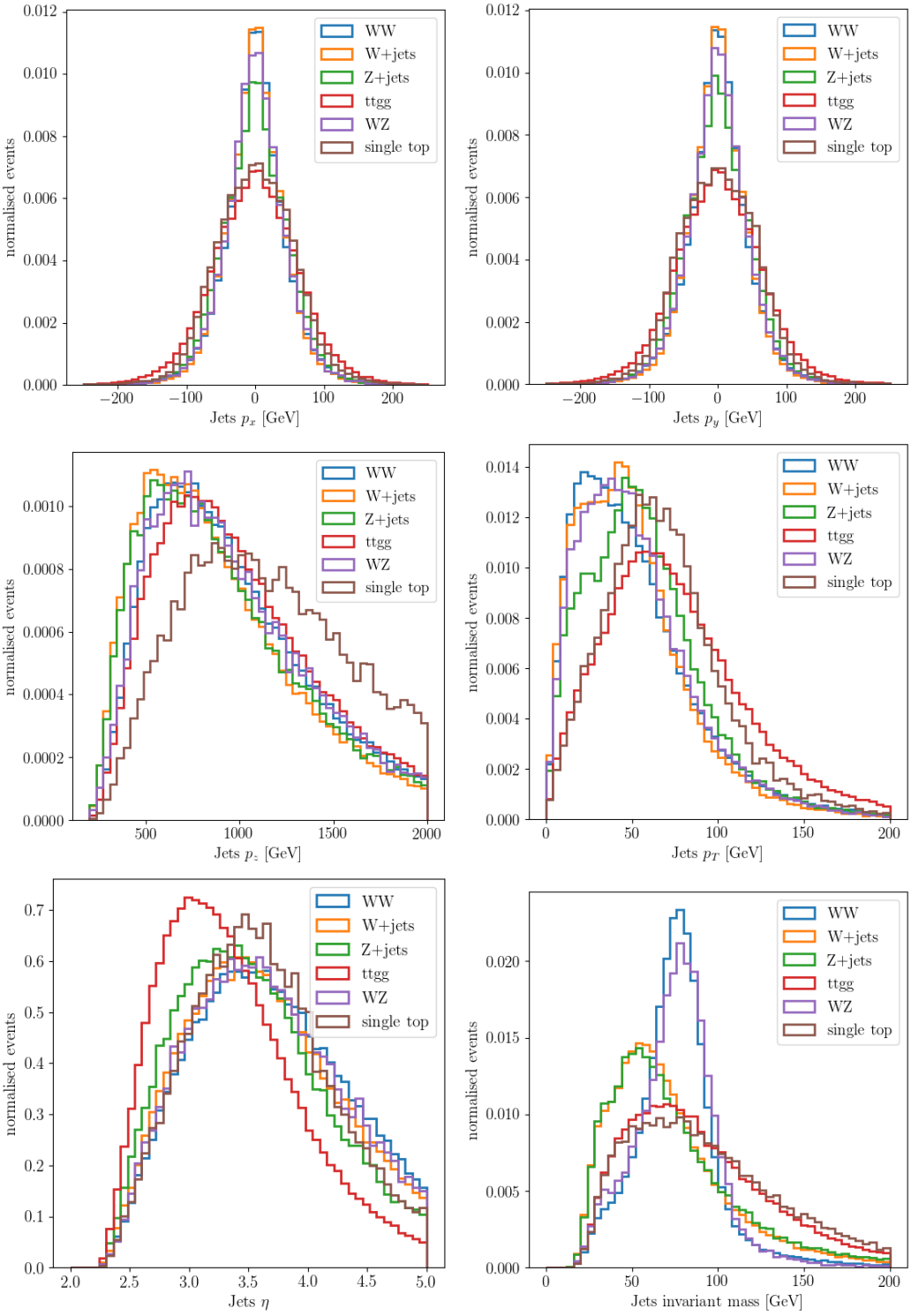


Figure 5.5: P_x, P_y, P_z, P_T, η, and invariant mass of the two jets system. Distributions are normalized to one.

- an overall quite good description is obtained by using CR2 to describe CR1, which means that the cut on IP_μ does not change the overall kinematics of the QCD events;

- despite the good matching between CR1 and CR2 templates, discrepancies are found, particularly at the edges of the DNN output. To correct these differences, weights have been computed

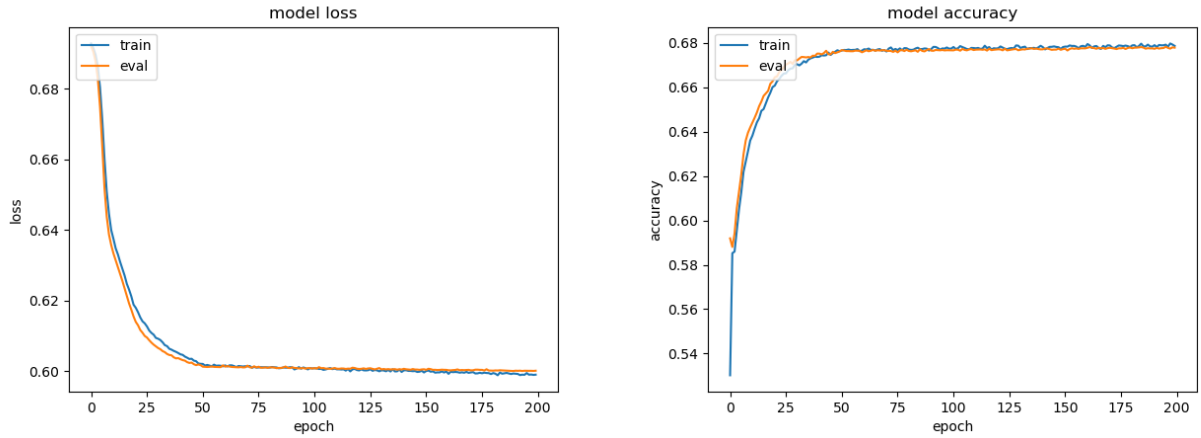


Figure 5.6: Loss function (left) and training accuracy (right) as a function of the number of epochs.

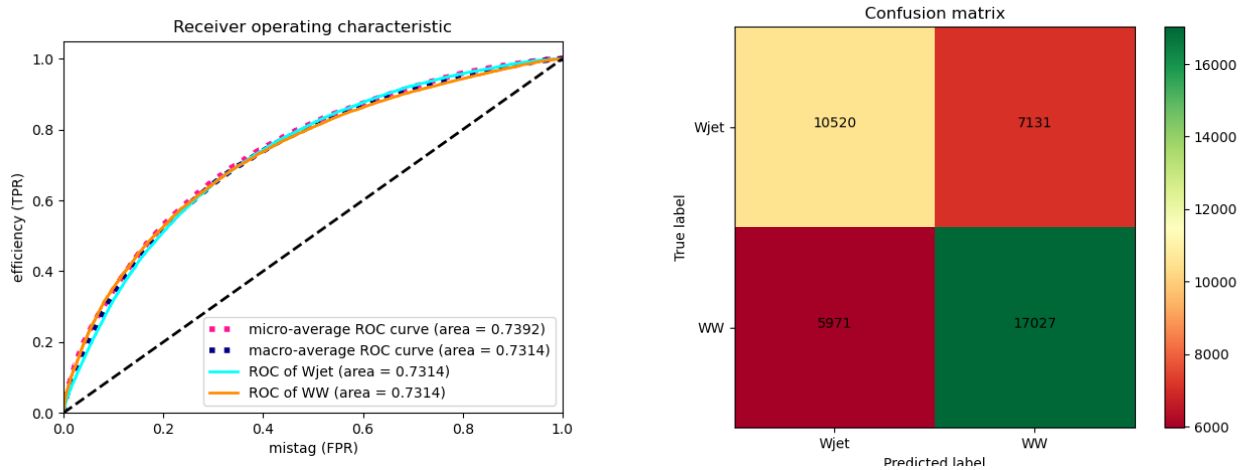


Figure 5.7: ROC curve (left) and confusion matrix (right) for the classification between WW and $W+$ jets events.

to match the DNN output between CR1 and CR2. The weights w are defined as

$$w = \frac{T_{\text{CR1}}^{\text{QCD}}}{T_{\text{CR2}}^{\text{QCD}}}, \tag{5.2}$$

where the $\frac{T_{\text{CR1}}^{\text{QCD}}}{T_{\text{CR2}}^{\text{QCD}}}$ ratio is performed on a bin-by-bin basis. The weights w are shown in Figure 5.11 as a function of the DNN output:

Finally, the DNN output template for QCD events in SR is obtained as

$$T_{\text{SR}}^{\text{QCD}} = T_{\text{CR}}^{\text{QCD}} \cdot w \tag{5.3}$$

and $T_{\text{SR}}^{\text{QCD}}$ is shown in Figure 5.12 together with DNN output for WW and $W+$ jets events, where it is evident that the DNN outputs for $W+$ jets and QCD events share similar shapes.

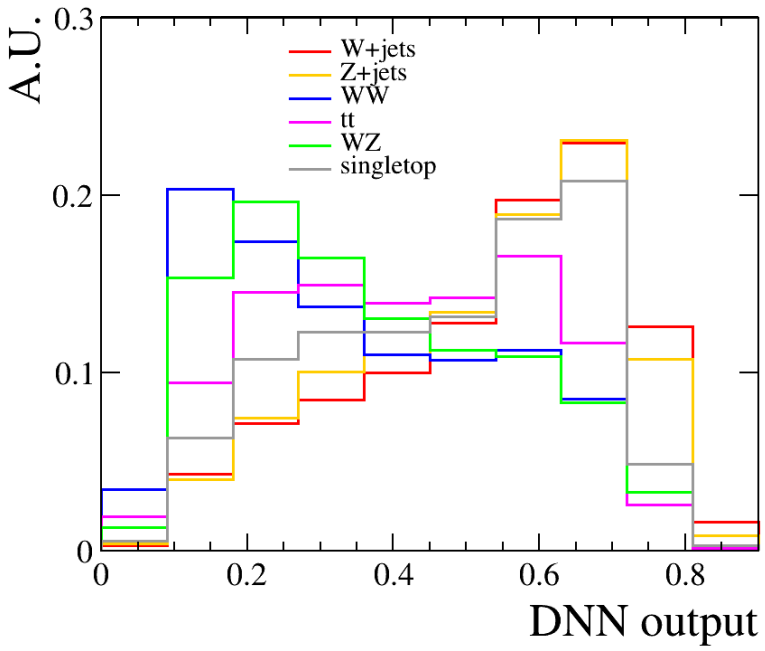


Figure 5.8: DNN output for the EW processes considered in this analysis. Distributions are normalized to one.

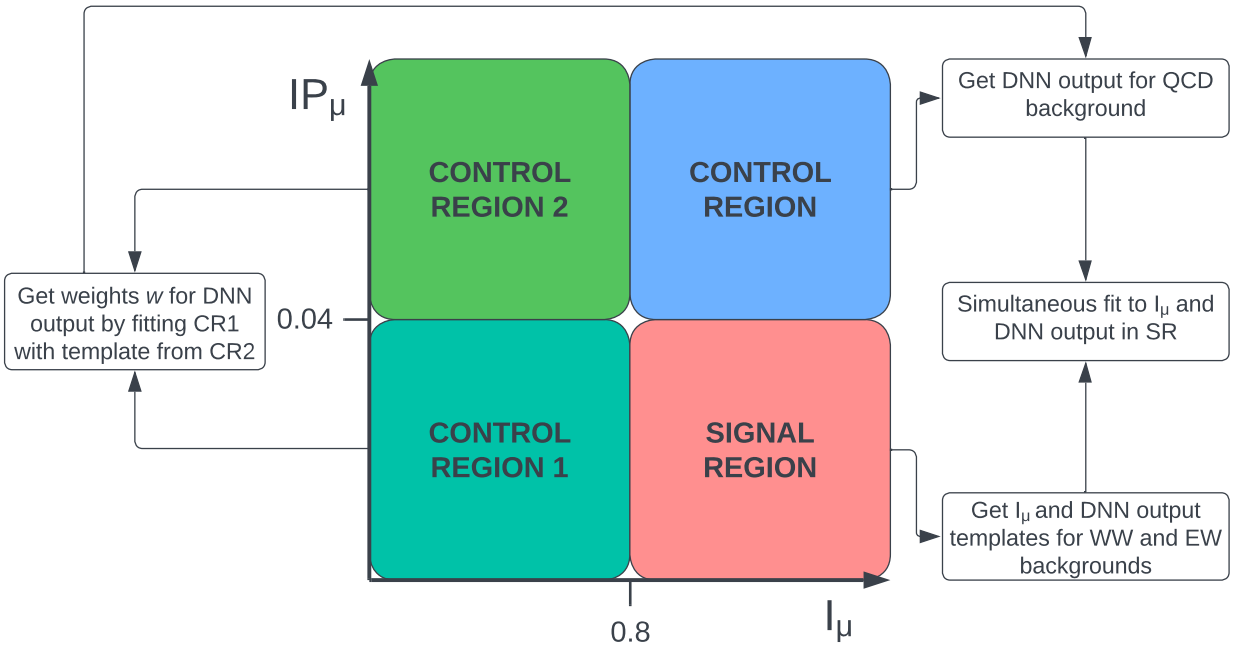


Figure 5.9: Schematic workflow of the fit procedure: the QCD DNN output in SR is obtained by evaluating the DNN in CR, and reweighting the DNN output in CR by the difference in DNN outputs between CR1 and CR2.

5.4 Fit procedure

The number of WW events is extracted through a maximum-likelihood bin fit, similar to what has been done in Section 4.5 for the extraction of $W+$ jets events. As shown in Sections 5.3 and 5.3.1, $W+$ jets events share similar DNN output as $Z+$ jets and QCD events. Therefore, a simple fit to the DNN output is not sufficient to properly constrain the yields of QCD and $W+$ jets events. To overcome

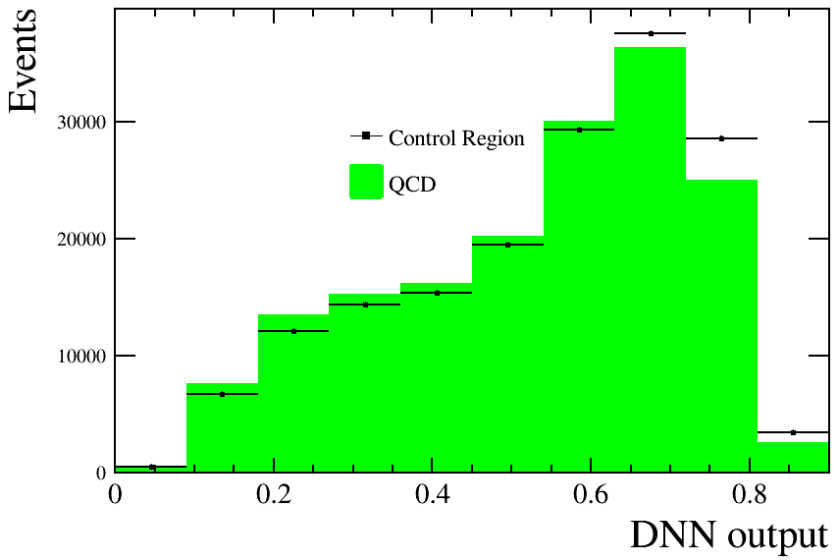


Figure 5.10: Fit of DNN output in CR1 (black dots) using DNN output evaluated in CR2 (green distribution).

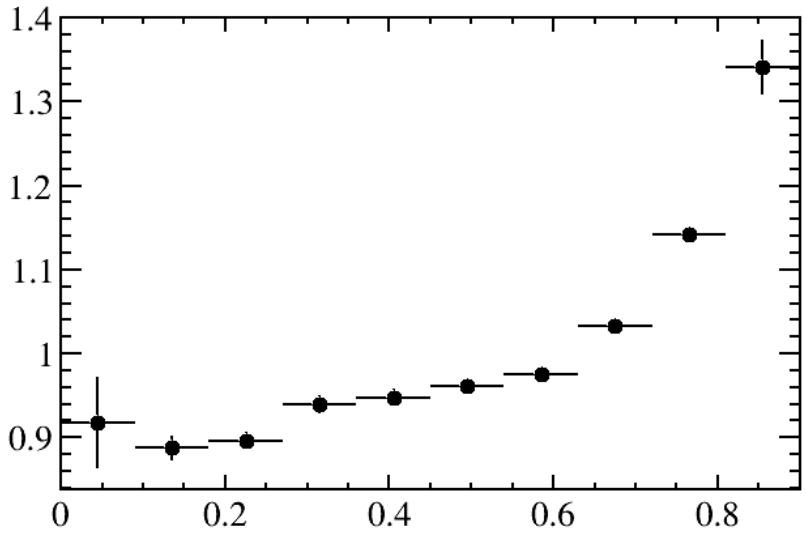


Figure 5.11: Weights w resulting from the comparison of the DNN output between CR1 and CR2. Weights w are shown as a function of the DNN output.

this issue, a simultaneous template fit to the DNN output and the muon isolation I_μ is performed. As already done for the extraction of $W+$ jets events in Section 4.5, $Z+$ jets yield is extracted with the data-driven technique explained in Section 4.4.1, while the $t\bar{t}$, single top and WZ yields are obtained from simulations, as described in Section 4.4.1. The numbers of events from WW, $W+$ jets, and QCD processes are free parameters of the fit. The result of the fit is shown in Figure 5.13, where it is evident that the dominant contribution comes from $W+$ jets events. The numbers of fitted events for WW, $W+$ jets, and QCD events are

$$\begin{aligned} N_{WW} &= 1426 \pm 663 \\ N_{W+\text{jets}} &= 114413 \pm 835 \\ N_{QCD} &= 54105 \pm 403 \end{aligned} \tag{5.4}$$

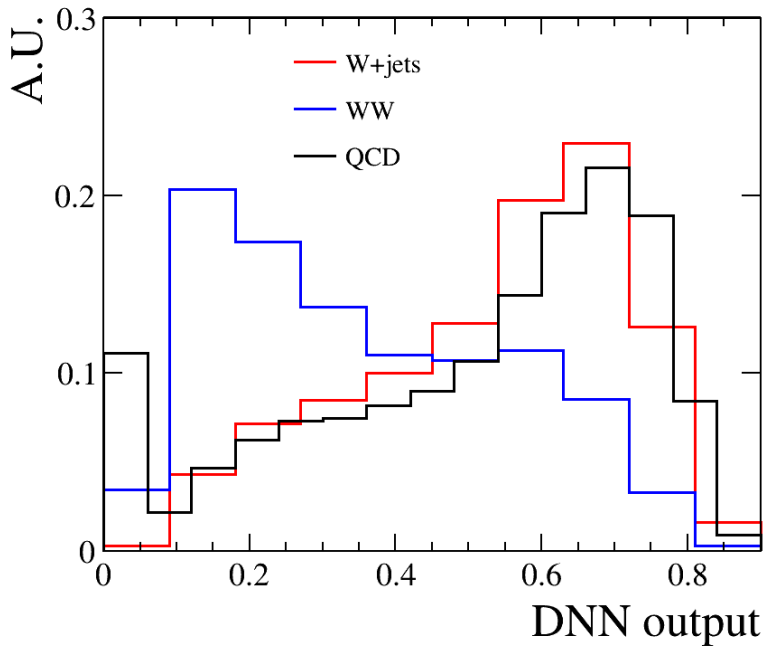


Figure 5.12: DNN output for WW, $W+$ jets, and QCD events. Distributions are normalized to one.

and both N_{Wjj} and N_{QCD} are compatible with the previous results obtained in Chapter 4. A look at the pull distribution of the fit shows that they are not centered around 0, but in this fit no systematic uncertainty contribution has been taken into account yet.

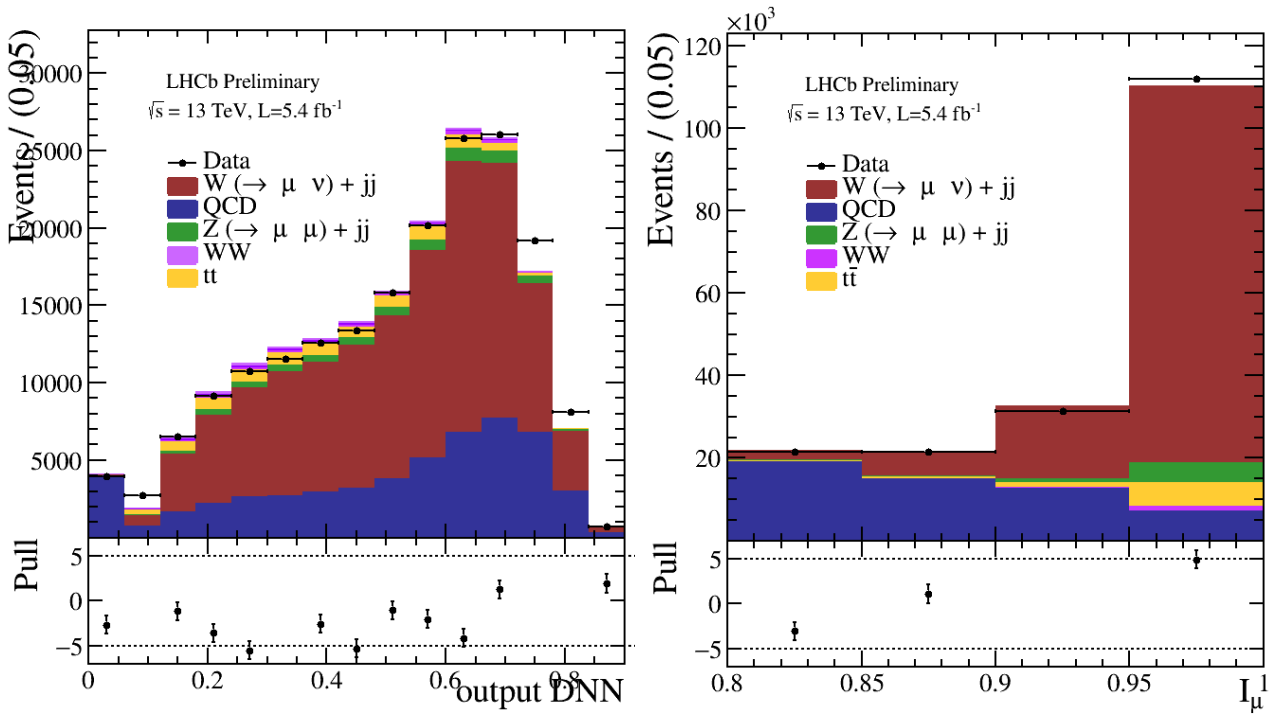


Figure 5.13: Results for the simultaneous fit of the DNN output (left) and the muon isolation I_μ (right).

5.5 Reconstruction and selection efficiencies

Reconstruction efficiencies have been already evaluated in Section 5.5 for the $W+$ jets selection, particularly for muon and jets reconstruction. The total reconstruction efficiency for reconstructing WW events is found to be

$$\varepsilon_{reco}^{WW} = (57.91 \pm 0.08)\% \tag{5.5}$$

which takes into account both muon and jets reconstruction efficiencies. The selection efficiency due to selection requirements put on top of reconstructed events (as described in Section 4.6.3) is found to be

$$\varepsilon_{sel}^{WW} = (68.16 \pm 0.09)\% \tag{5.6}$$

5.6 Acceptance factor and unfolding

The acceptance factor $\mathcal{A}$ defined in Section 4.7 takes into account the difference between the fiducial region cuts and the analysis cuts: while the fiducial region asks for $p_{\text{T}}(\mu+\text{jets}) > 20$ GeV, the analysis cut requires $p_{\text{T}}(\mu^{\text{jet}}+\text{jets}) > 20$ GeV, with μ^{jet} being already defined in Section 4.4.2. The acceptance factor $\mathcal{A}$ has been computed using WW simulated samples, and it is found to be

$$\mathcal{A} = (98.89 \pm 0.03)\% \tag{5.7}$$

The unfolding of the reconstructed events has been done following the same procedure described in Section 4.8. Here no differential measurement is performed, therefore the unfolding factor f simply accounts for detector resolution effects that migrate events inside or outside the fiducial region defined in Table 5.1. The unfolding factor is $f = 0.97 \pm 0.01$.

5.7 Systematic Uncertainties

Several systematic uncertainties affect this measurement. The vast majority of them have been already studied in Section 4.9, and they are easily applied to this measurement. Two other sources of systematic uncertainties are present in this analysis:

- a systematic uncertainty related to the reweighting procedure to obtain the QCD template in the SR;
- a systematic uncertainty related to the definition of the control regions CR, CR1 and CR2.

5.7.1 Reweighting procedure

As already described in Section 5.3.1, the DNN output for the QCD template in SR is obtained starting from CR, and reweighting the template with weights w obtained by comparing the DNN output in CR1 and CR2, as shown by Equation 5.3. The weights obtained by the comparison between CR1 and CR2 come with a statistical error related to the amount of data considered in the $I_\mu < 0.8$ region. Therefore, the statistical error of the weights is propagated to the template fit by letting $T_{\text{SR}}^{\text{QCD}}$ to vary within its statistical error. The difference in the number of fitted WW events with respect to the nominal fit is taken as the systematic uncertainty on the reweighting procedure. This systematic uncertainty is found to have an impact on the WW cross-section measurement of 1%.

5.7.2 Control regions definition

To obtain the DNN output for the QCD events, the definition of signal and control regions has been described in Section 5.7.2. This approach is based on two assumptions:

- there is no dependence on IP_μ, given that the DNN output template obtained in CR is used as a QCD template in SR;

- there is no dependence on I_μ, since the reweighting procedure is based on templates obtained from CR1 and CR2.

These assumptions have been checked by changing the IP_μ and I_μ values defining the various regions. For every new configuration of IP_μ and I_μ values, the simultaneous fit to I_μ and the DNN output is done again, and the difference between the new number of WW fitted events and the nominal value presented in Equation 5.4. The relative impact of this systematic uncertainty on the WW cross-section measurement is 3%.

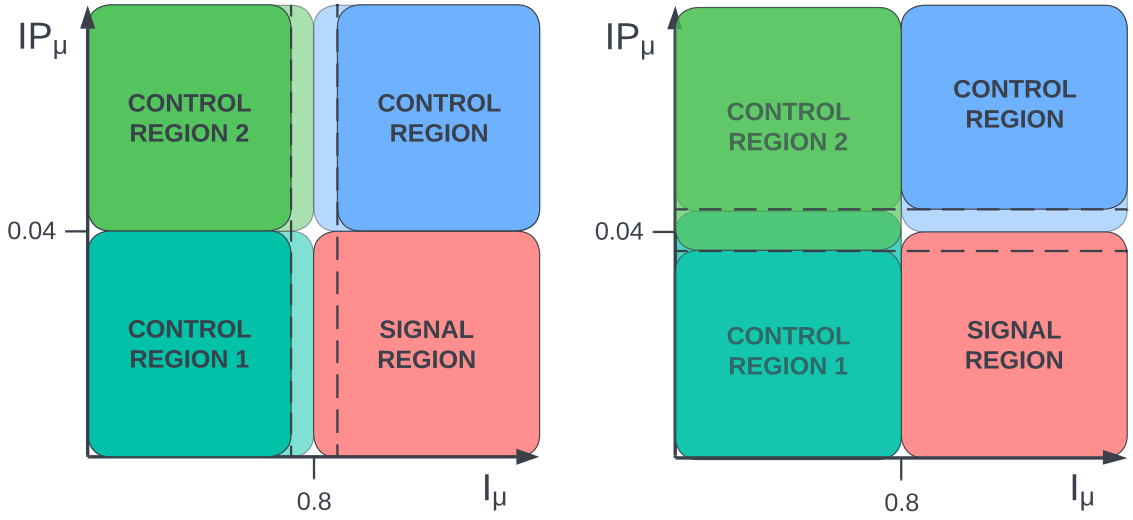


Figure 5.14: Schematic representation of the evaluation of the systematic uncertainty on control regions definition: control regions definitions are varied by varying the cut on I_μ and on IP_μ.

5.7.3 Summary of systematic uncertainties

A summary of the relative systematic uncertainties affecting the WW cross-section measurement is shown in Table 5.3.

Source	systematic error
Muon reconstruction	0.5%
Jet reconstruction	1.2%
Jet Energy Resolution	3.2%
Jet Energy Scale	4.9%
Selection	1.3%
Reweighting procedure	1.0%
Control regions definition	3.0%
Total systematic uncertainty	6.9%
Luminosity	2%

Table 5.3: Summary of relative systematic uncertainties affecting the WW cross section measurement.

5.8 Results and final considerations

The $W(\to \mu\nu)W(\to \text{jets})$ cross section is computed as

$$\sigma\left(W(\to \mu\nu)W(\to \text{jets})\right) = f \cdot \frac{N_{fit} \cdot \mathcal{A}}{\varepsilon_{reco}^{WW} \cdot \varepsilon_{sel}^{WW} \cdot \mathcal{L}} \tag{5.8}$$

where f is the unfolding factor, N_{fit} is the number of fitted WW events obtained in Section 5.4, $\mathcal{A}$ is the acceptance factor obtained in Section 5.6, ε_{reco}^{WW} and ε_{sel}^{WW} are respectively the efficiency to reconstruct and to select WW events obtained in Section 5.5, while $\mathcal{L}$ is the total integrated luminosity. The result obtained for the $W(\to \mu\nu)W(\to \text{jets})$ cross-section measurement is

$$\sigma\left(W(\to \mu\nu)W(\to \text{jets})\right) = 0.64 \pm 0.30 \pm 0.05 \pm 0.01 \text{ pb} \tag{5.9}$$

where the first uncertainty is statistical, the second one is systematic uncertainty, and the third one is related to luminosity determination. The measured cross section is compared with theoretical predictions obtained with MadGraph: similarly to what has been described in Section 4.10.1, MadGraph is interfaced with Pythia8 for parton shower and hadronization; the theoretical prediction is computed at NLO accuracy using the NNPDF2.3 PDF set [51]. Results are shown in Figure 5.15.

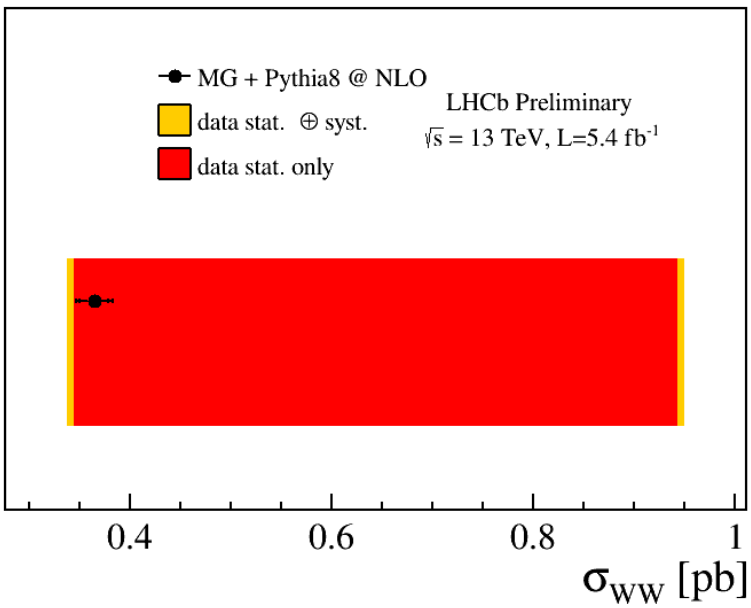


Figure 5.15: Comparison between the WW cross section measurement and the WW theoretical predictions. The red band indicates the statistical uncertainty, while the yellow band indicates the total uncertainty.

A few considerations can be done following the results just obtained:

- the measurement is evidently limited by the statistics available. In the following data-taking campaigns, it will be possible to improve the quality of this measurement. For example, at the end of the Run 3 data-taking campaign, the LHCb experiment will collect an integrated luminosity of almost 15 fb^{-1} [97]. The projected sensitivity to the luminosity collected at the end of Run 3 $\mathcal{L}_{\text{Run 3}}$ can be computed by simply rescaling the actual sensitivity by a factor

$$k = \sqrt{\frac{\mathcal{L}_{\text{Run 2}}}{\mathcal{L}_{\text{Run 3}}}} = 0.6 \tag{5.10}$$

  which will lead to a sensitivity of this measurement of the order of $\sim 30\%$;

- to perform this measurement, the semileptonically W boson is required to decay into a muon and a neutrino, to ensure good reconstruction given the excellent performance in reconstructing muon. Moreover, as already mentioned in Section 2.2.5, the LHCb ECAL is not optimized to reconstruct high p_{T} electrons. In view of the future upgrade of the LHCb experiment, a further optimization of ECAL might be necessary, in order to optimize the electron reconstruction and therefore consider also the $W \to e\nu$ channel. This study is presented in the next Chapter;
- as already mentioned in Section 1.7.3 and at the beginning of this Chapter, the feasibility study to measure WW in the $e\mu$ channel has not considered the usage of ML techniques to further separate WW events from backgrounds. This might help in improving the significance obtained in Section 1.7.3;
- the DNN classifier uses as inputs only information coming from the kinematics of the event. Given the low pile-up environment, at the LHCb experiment, it is possible to reconstruct and identify the particles inside jets; usually, these features are referred to as "jet substructure" observables. An attempt of including also these jet substructure features has been tried in this analysis, but no evident gain in the classification task has been found. In recent years, more complicated ML classifiers have been developed [98–100]: Graph Neural Networks (GNN) [101] have recently been proven to overcome the performance of state-of-the-art DNN. Therefore, a possible improvement in this analysis might come from a different choice of the ML classifier, which might better exploit the different internal structures of jets coming from a W boson decay (as in WW events) or from the whole events (as in $W+$ jets events);
- in this analysis, the proposed DNN has been trained to separate WW from $W+$ jets events since the latter is the higher background. As shown in Figure 5.8, the DNN is not able to separate between WW, $t\bar{t}$ and WZ events. A possible solution is to train a second DNN to perform this separation, and possibly gain further discrimination.

Chapter 6

Study of the calorimeter system upgrade in view of future EW physics analyses

The WW measurement presented in Chapter 5 is limited by the statistics available using the Run 2 dataset. Particularly, the choice of considering only the $W(\rightarrow \mu\nu)$ channel limits the statistics by almost a factor of two. Moreover, similarly to the other LHC experiments, the LHCb experiment will undergo a series of upgrades, that will take the experiment to the High Luminosity phase of the LHC (HL-LHC), where the luminosity delivered to experiments is expected to increase by a factor of 100. The high performance on tracks and vertices reconstruction, described in Chapter 2, needs to be maintained, despite the obvious increase in the occupancy of the detector. In view of the future EW measurements at the LHCb experiment, this Chapter presents the work done to assess the performance in the upgrades conditions, particularly for the calorimeter system. The structure of this Chapter is the following:

- Section 6.1 presents few examples of EW measurements that would benefit from the upgrades of the LHCb experiment;
- Section 6.2 gives an overview of the LHCb experiment's future upgrades, highlighting the main changes and improvements from the detector side;
- Section 6.3 presents a detailed description of the upgrade of the calorimeter system, describing the necessary changes and possible improvements to tackle the reconstruction of events with a higher pile-up environment;
- Section 6.4 focuses on the studies of the ECAL upgrade. Particularly, the reconstruction of high p_{T} electrons is presented;
- finally, Section 6.5 tackles the proposal of removing HCAL. The HCAL removal impact is studied by analyzing the reconstruction performance on jets. Given that HCAL has a role in the PID algorithms used at the LHCb experiment, this Section presents also a new way of using upgraded ECAL to perform PID discrimination between particles.

6.1 EW measurements at the LHCb Upgrades

EW physics analyses involve the reconstruction and selection of high p_{T} objects, such as electrons and muons from vector bosons decay and jets. The quality of the analyses is primarily based on the quality of the reconstruction of these objects: this last consideration is of uttermost importance for objects that are reconstructed by several sub-detectors, jets being the clearest example. Since the focus of this Chapter is on the calorimeter system, here few EW analyses that might benefit from a calorimeter upgrade are briefly described:

- W mass measurement: at the LHCb experiment, the measurement of the W boson mass uses only the $W \to \mu\nu$ channel, since the ECAL ADC saturation is not optimized for reconstructing high p_{T} electrons. Using the 2016 dataset, the W mass if found to be [86]

$$m_W = 80354 \pm 23_{\mathrm{stat}} \pm 10_{\mathrm{exp}} \pm 17_{\mathrm{theory}} \pm 9_{\mathrm{PDF}} \text{ MeV} \tag{6.1}$$

  and the projection to the whole Run 2 dataset is expected to lower the statistical error to ~ 10 MeV. It is clear that the inclusion of the $W \to e\nu$ channel, possible only if high p_{T} are well reconstructed, will help in lowering the statistical error;

- top physics: at the LHCb experiment one of the main channels to study top physics is the μeb channel, where a muon, an electron, and a b-jet are reconstructed in the LHCb acceptance to measure the $t\bar{t}$ cross-section [87]. It is evident that a good reconstruction of both the electron and jets is fundamental to keeping a good precision for this measurement;

- Higgs physics: so far at the LHCb experiment no Higgs boson has been measured. The most promising channels to be studied in the future LHCb upgrade are $h \to b\bar{b}$ and $h \to c\bar{c}$. Particularly, at the end of the HL-LHC phase, the LHCb experiment might be able to obtain new limits on the $h \to c\bar{c}$ cross-section in the same order of ATLAS and CMS [97]. This of course will be possible if and only if a good jet reconstruction performance is maintained throughout the upgrade phases.

6.2 The LHCb experiment upgrade

At the time of writing of this thesis, the LHCb experiment is undergoing the Run 3 data-taking period. The running period of an experiment is divided in runs, interleaved by Long Shutdown (LS) periods to allow for upgrades both in the accelerator machine (LHC) and in all the experiments: this happened between Run 1 and Run 2 in the LS1 when the collision center-of-mass energy has changed from $7-8$ TeV to the actual 13 TeV of Run 2. In order to investigate the evidence of new physics beyond the Standard Model, the number of data gathered from interactions should be increased. Therefore, not only changes in the center-of-mass energy but also in the luminosity of an experiment are considered during LS periods. To account for these changes, the LHC machine will go through a series of upgrades that will take the LHC experiments to the HL-LHC, where an increase in luminosity of factor 100 with respect to Run 2 conditions is expected, reaching a delivered luminosity of $\mathcal{L} = 2 \cdot 10^{34}$ cm^{-2}s^{-1}. As shown in Figure 6.1, parallel to the upgrades of the LHC, the LHCb experiment will undergo several LS periods that will lead the experiment through several upgrade stages.

The Run 3 period coincides with the "Upgrade Ia" phase [102], the first stage of the upgrade where the instantaneous luminosity will increase by a factor of 5, while the "Upgrade Ib" and the "Upgrade II" phases will prepare the LHCb experiment for the HL-LHC phase. As shown in Figure 6.2, the major upgrade in luminosity will happen after LS4, where an increase of luminosity by a factor 10 will allow the LHCb experiment to get a total integrated luminosity of $\mathcal{L}_{\mathrm{int}} = 300$ fb^{-1}: data collected by the end of the HL-LHC period will be more than a factor thirteen higher than data collected in the

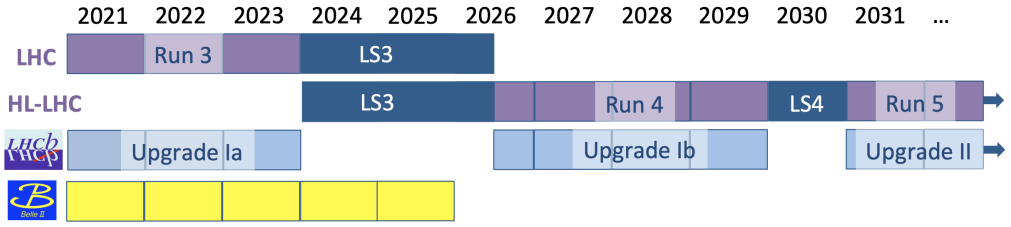


Figure 6.1: Agenda of the future upgrades of the LHC and the LHCb upgrades. Notably, the LHCb experiment will start its HL-LHC phase in Run 4 [97].

pre-HL-LHC period, and a factor six higher than data collected at the end of Run 4, leading to major improvements in all those measurements that are so far statically limited. In this section, an overview of the main aspects of the LHCb upgrades is given.

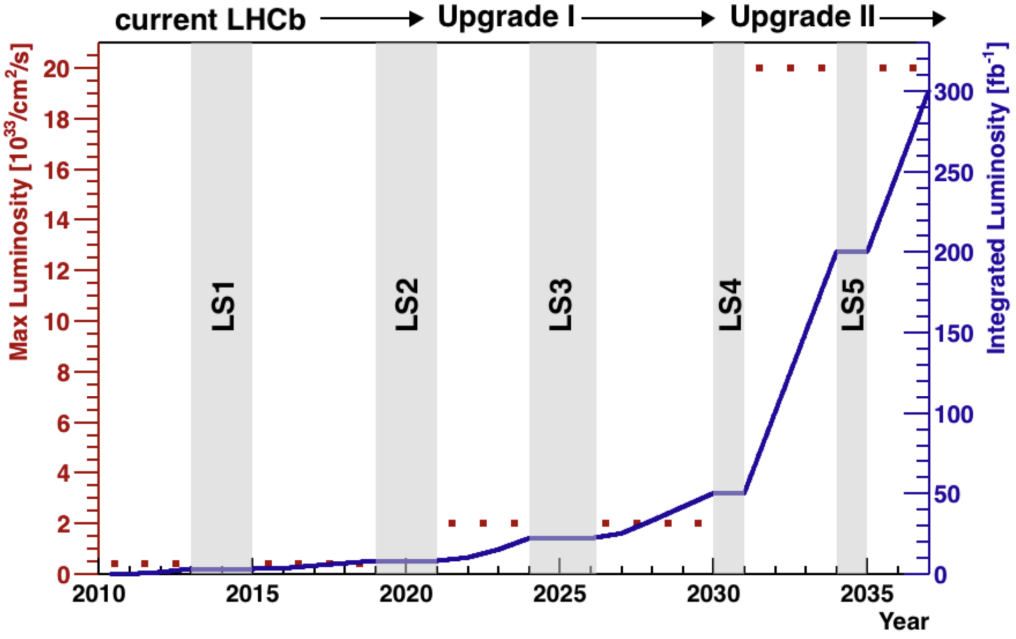


Figure 6.2: Maximum peak luminosity (red) and total integrated luminosity (blue) as a function of the year of data taking. The various LHCb upgrades, as well as the LSs, are highlighted [103].

6.2.1 Upgrade Ia and Ib

The original LHCb detector was designed to collect 8 fb^{-1} of data, at an instantaneous luminosity up to $\mathcal{L} = 2 \times 10^{32}$ cm^{-2}s^{-1}. Given that this target has been well exceeded at the end of Run 2 (reaching $\mathcal{L} = 4 \times 10^{32}$ cm^{-2}s^{-1}), in order to further proceed with the LHCb physics program an upgrade is necessary. The center-of-mass energy will reach the nominal conditions at $\sqrt{s} = 14$ TeV[1], and the instantaneous luminosity will be increased to $\mathcal{L} = 2 \times 10^{33}$ cm^{-2}s^{-1}, a factor five with respect to Run 2 conditions, that will allow collecting almost 50 fb^{-1} of data at the end of Run 4: this will translate in a higher pile-up environment, almost doubling the number of visible interactions with respect to Run 2. To sustain these new experimental conditions, and keep excellent performance, the upgrade of the LHCb experiment will concern all sub-detectors as well as the readout system. The most notable change is the usage of a full-software trigger and readout system, working at the 30 MHz LHC event

[1] At the time of writing this thesis, the LHC has reached center-of-mass energy $\sqrt{s} = 13.6$ TeV.

rate. The Upgrade I [102] phase is divided into two steps: Upgrade Ia, which coincides with Run 3 and whose operations have been carried on during LS2 (therefore it is already in place), and Upgrade Ib, which will happen during LS3 and will operate during Run 4. Figure 6.3 shows the new LHCb Upgrade I detector.

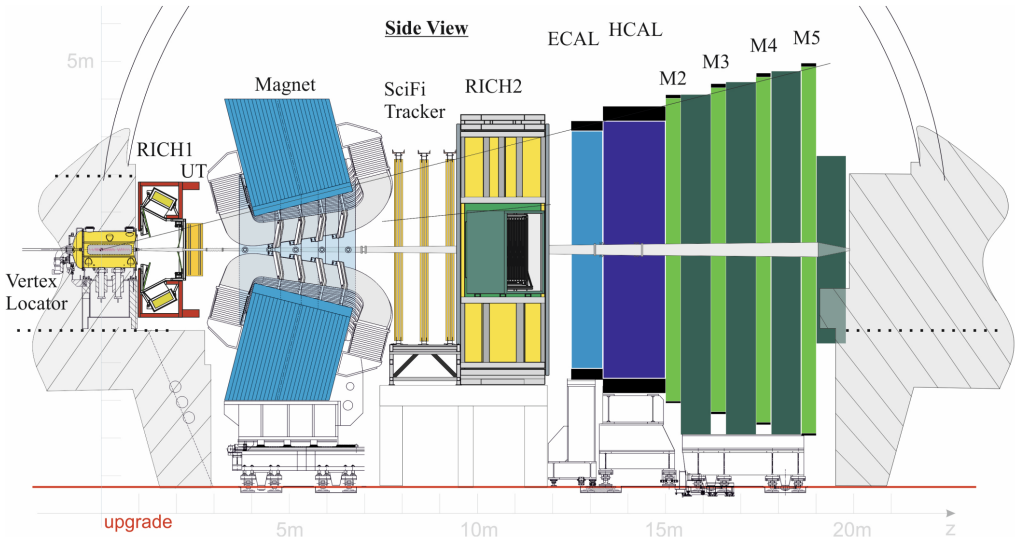


Figure 6.3: Geometry of the LHCb experiment during Upgrade Ia and Ib. The main notable differences are the UT (SciFi) in place of the TT (T1-3 stations), the removal of the SPD/PS system, and the M1 muon chamber [102].

6.2.2 Upgrade II

As shown in Figures 6.1 and 6.2, the most challenging upgrade is Upgrade II, where the luminosity delivered at the LHCb experiment will increase of factor 10, reaching a maximum luminosity of $\mathcal{L} = 1.5 \times 10^{34}$ cm^{-2}s^{-1} that will allow the LHCb experiment to collect 300 fb^{-1} of data. The upgrade operations will happen during LS4 and the upgraded experiment will take data during Run 5, currently scheduled for 2032. Differently with respect to Upgrade I, an essential element that will allow precision measurements in flavor physics at such high luminosity is precision time measurements: a resolution of a few tens of picoseconds will allow a correct matching between tracks, clusters, and the relative interaction points. The performance expected from Upgrade II is similar to Upgrade I, but this time in a pile-up environment 7.5 higher than Upgrade I, therefore a challenge. A schematic view of the LHCb Upgrade II is shown in Figure 6.4

6.3 The calorimeter system upgrade

In preparation for Run 4 and following, the calorimeter system will face several changes. The main motivations for these changes are the following:

- the ECAL will face high radiation levels, particularly in the region close to the beampipe, where a dose of almost 1 MGy is expected at the end of the HL-LHC phase, after an integrated luminosity of 300 fb^{-1}, as shown in Figure 6.5. This dose is well above the operational limits of the current Shashlik modules, which is around 40 kGy. This implies the search for new materials and geometries which guarantee radiation tolerance;

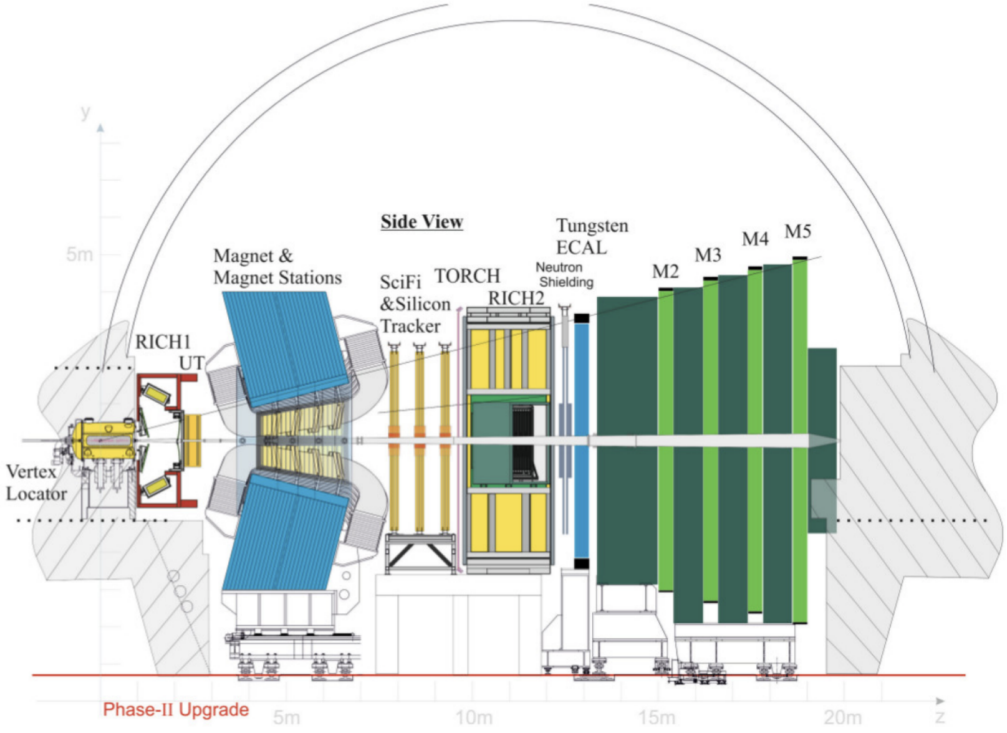


Figure 6.4: Proposed geometry for the LHCb experiment during Upgrade II. The main notable differences are the Magnet stations at the sides of the magnet, the introduction of the Silicon Tracker in the tracking system, and the removal of the HCAL [103].

- the pile-up expected in the Upgrade 2 phase is ~ 50, which is almost a factor 50 bigger than the pile-up measured during Run 2, and 7 times bigger than the pile-up expected in Run 3. Therefore, a finer granularity is needed, together with the possibility of measuring timing information with a resolution of the order of $\sim$ 20 picoseconds, a requirement mandatory even for other sub-detectors (*e.g.* the VELO);
- the upgraded ECAL should have the same performance as the actual calorimeter, particularly regarding the energy resolution, whose performance has been already described in Section 2.2.5.

In order to tackle these challenges in the future, an upgrade of the ECAL is under development. Particularly, the following strategies have been proposed:

- the geometry for arranging the ECAL cells (and therefore the PMTs and the readout) is changed, going from three rectangular regions with increasing cell size (inner, middle and outer with cell size respectively of 4×4, 6×6 and 12×12 cm^2, as shown in Figure 2.13) to "rhomboidal" regions, which follows the radiation map of ECAL shown in Figure 6.5. Particularly, two different proposals are in place for Run 4 and Run 5, shown in Figure 6.6: for Run 4, the inner region will be covered by 2×2 cm^2 cells, with increasing size (from 4×4 to 12×12 cm^2) following the new geometry pattern, while for Run 5 the inner region will be covered by 1.5×1.5 and 3×3 cm^2 cells before increasing the cells size up to 12×12 cm^2; The impact of the new proposed geometries is evident in Figure 6.7, where the actual geometry and the Run 4 geometry are compared by evaluating the particle occupancy for cells with transverse energy $E_{\mathrm{T,cell}} > 50$ MeV in Run 4 conditions: the new geometry keeps low occupancy values, guaranteeing good reconstruction performance [103];
- in the innermost region, the current Shashlik modules will be replaced using a "Spaghetti

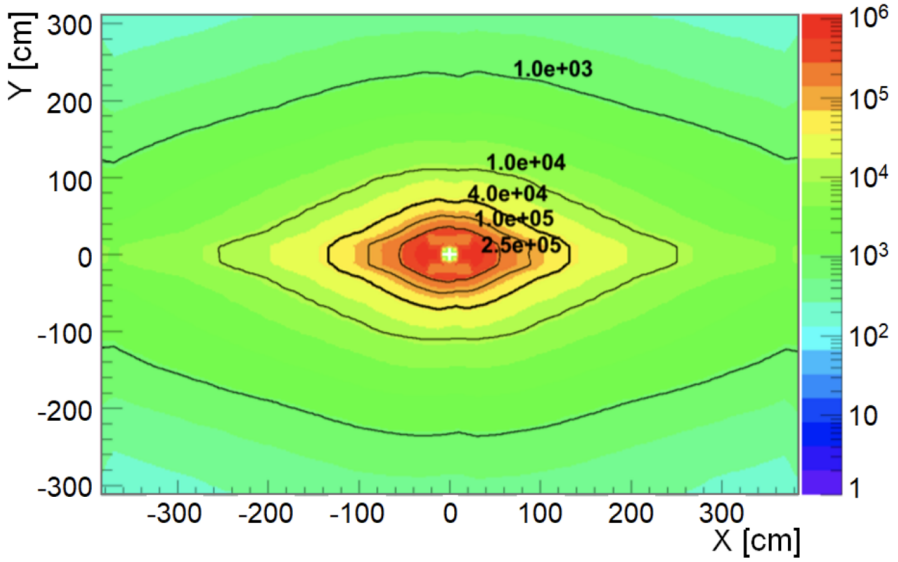


Figure 6.5: Radiation maps (expressed in Gy) of the ECAL after the HL-LHC phase, where the actual ECAL geometry is assumed. The $4.0 \cdot 10^4$ Gy limit corresponds to the operational Shashlik limit, showing that the innermost region would not be operational in these conditions [103].

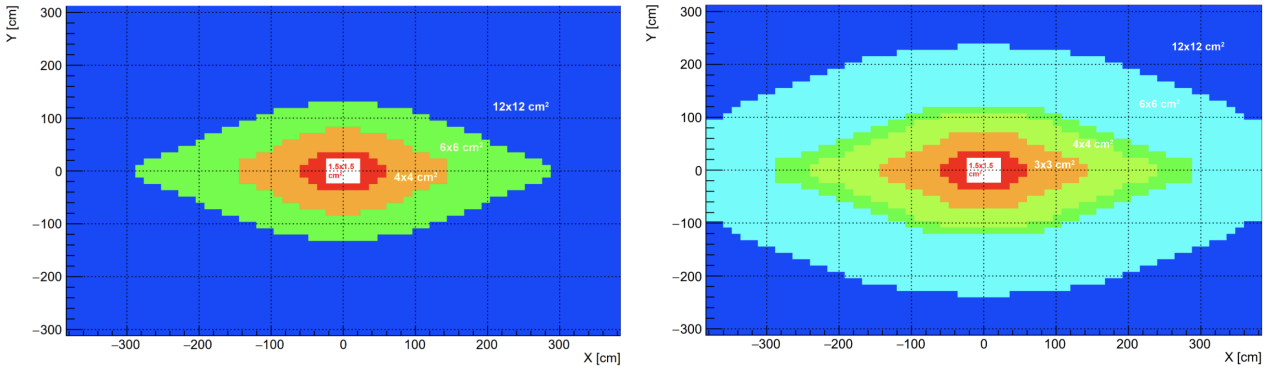


Figure 6.6: Proposed new geometries for the upgraded ECAL: (left plot) Run 4 proposal, (right plot) Run 5 proposal [103].

Calorimeter" (SpaCal) [104] technology, where the scintillating material is placed in long fibers which are immersed into the absorber material. In this way, the scintillating fibers serve both as scintillating material and to drive the scintillating photons to the PMTs. This choice is driven by radiation tolerance while keeping good energy resolution performance. A front view of a SpaCal prototype with different crystal fibers is shown in Figure 6.8. The outermost regions will be instrumented with Shashlik modules since the occupancy in the outermost region is still manageable;

- both the new SpaCal modules and the Shashlik ones can be segmented into a front and a back section. This would require a double readout system, with PMTs both behind and in front of the module. The longitudinal segmentation of SpaCal and Shashlik modules is schematically shown in Figure 6.9, while Figure 6.8 (right) evidence the longitudinal segmentation in a SpaCal prototype. A possible application of this new feature for PID is presented in Section 6.5.2.

Depending on the cell size, the choice of the material of fibers and absorber, and the presence of longitudinal segmentation, several configurations for the SpaCal modules have been proposed, each

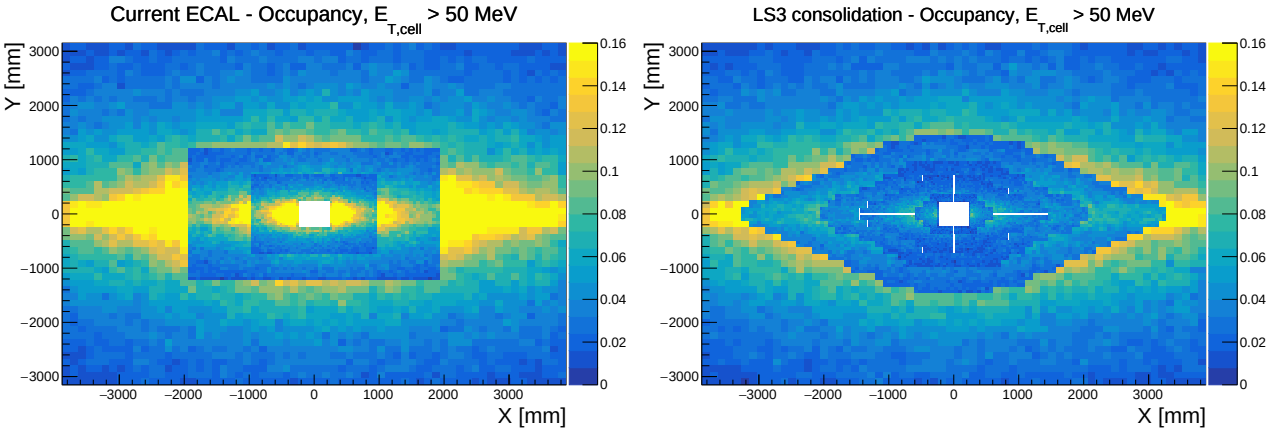


Figure 6.7: Comparison between occupancy maps in Run 4 conditions for actual Run 3 geometry (left plot) and proposed Run 4 geometry (right plot) [103].

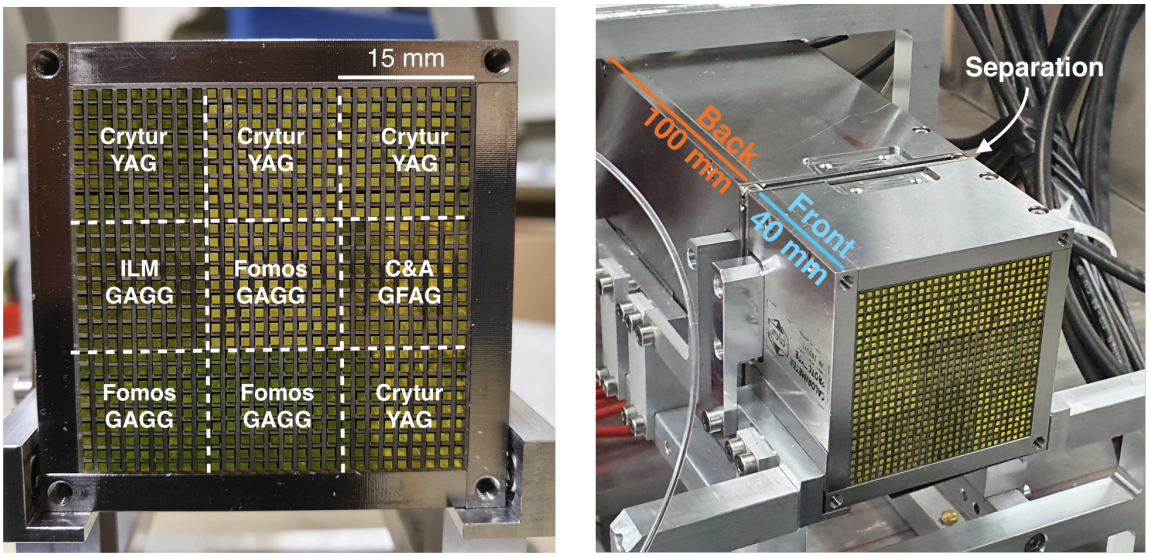


Figure 6.8: Front view (left) and side view (right) of a SpaCal module. The fibers structure of each cell and the longitudinal separation are highlighted [104].

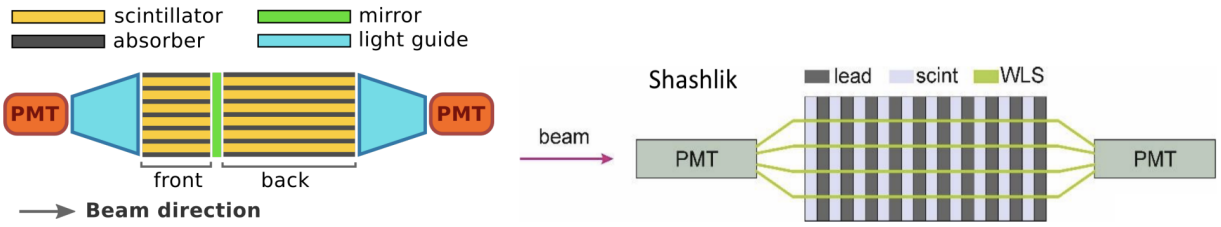


Figure 6.9: Schematic representation of the longitudinal segmentation of a SpaCal module (left) and a Shashlik module (right) [103].

designed to fulfill the requirements for a specific ECAL configuration:

- Pb+Poly configuration: a SpaCal module with lead (Pb) absorber and polystyrene fibers.
- W+Poly configuration: a SpaCal module with tungsten (W) absorber and polystyrene fibers.
- W+GAGG configuration: a SpaCal module with tungsten absorber and crystal fibers. This

configuration will be used for the Upgrade 2 configuration in the innermost region, particularly requiring 32 modules with 64 cells each with 1.5×1.5 cm^2 size.

An overview of the main features of these module configurations are listed in Table 6.1.

Module	W+Poly	W+GAGG	Pb+Poly
Absorber	Tungsten (W)		Lead (Pb)
Fibers	polystyrene	GAGG	polystyrene
Cell size	2×2 cm^2	1.5×1.5 cm^2	3×3 cm^2
Moliére Radius	1.8 cm	1.46 cm	3 cm
Radiation Length	0.72 cm	0.62 cm	1 cm
Longitudinal sections	1 (19 cm)	2 (4.5 cm + 10.5 cm)	1 (29 cm) / 2 (8 cm + 21 cm)

Table 6.1: Summary table of different proposals for SpaCal modules [103].

6.4 Reconstruction of high p_T electrons in the upgraded ECAL

A good reconstruction of high p_T electrons is fundamental from the EW measurements perspective: channels such as $Z \to e^+e^-$, $W^\pm \to e^\pm\nu_e$ and top decays will benefit from good reconstruction performance of ECAL, as described in Section 6.1. Two main issues can be found when approaching high p_T electron reconstruction in the new upgraded conditions:

- higher occupancy expected in the future upgrades will make object reconstruction particularly challenging;
- during Run 1 and Run 2, EW physics measurements have been limited by a rather low ADC saturation of ECAL cells: while this requirement is fundamental to keep high reconstruction performance for b physics, where objects have energies of few GeV, it does not allow meaningful electron reconstruction in the high p_T range. The ADC saturation is calibrated in terms of transverse energy E_T, and during Run 1 and Run 2, the ADC threshold was set at $E_{thr} = 10$ GeV, which produced a broad "peak" for the $Z \to e^+e^-$ process, as shown in Figure 6.10[2].

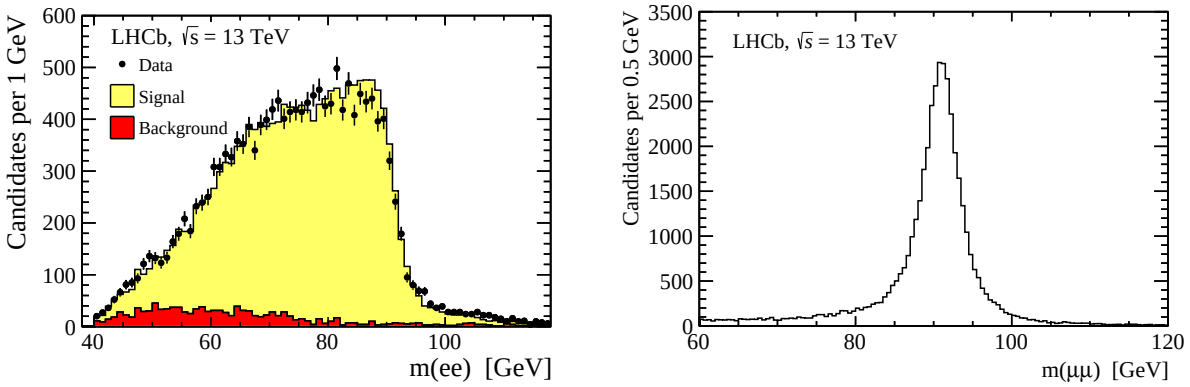


Figure 6.10: Invariant mass of e^+e^- pairs coming from $Z \to ee$ decay, where the effect of low ADC saturation is evident from the left tail. (left plot) The invariant mass of $\mu^+\mu^-$ pairs coming from $Z \to \mu\mu$ decay, highlighting the optimal reconstruction performance for muons [48]. (right plot)

[2]While a tail on the left side of the peak is expected due to Bremsstrahlung photons contribution (that does not happen in the $Z \to \mu^+\mu^-$ case), the broadness of the peak is almost entirely due to ADC saturation.

6.4.1 Simulation tools

At the moment of writing this thesis, no official simulation of the various configurations for the upgraded LHCb detector is available. Therefore, to study the performance of the upgraded ECAL in the several proposed configurations, a simulation framework has been set up. While the simulation of a Shashlik module is rather trivial and follows the actual ECAL simulation, a proper simulation of a SpaCal requires the propagation of optical photons inside the SpaCal module fibers. This task can be achieved by performing full ray-tracing of these photons. While possible, this task is very computationally demanding, with a simulation time of approximately 1-2 hours per GeV of deposited energy. Therefore, the simulation framework for the upgraded ECAL has been built by developing a "hybrid" approach, which allows keeping good realism while lowering the simulation time. In the following, we will refer to this simulation framework as "Hybrid-MC" simulation framework. A flowchart of the hybrid-MC simulation approach is shown in Figure 6.11, where the main steps are highlighted: while the energy deposition and the full ray-tracing of Cherenkov photons are performed using GEANT4 [71, 72], the transport of scintillation photons is done in a parametrized way.

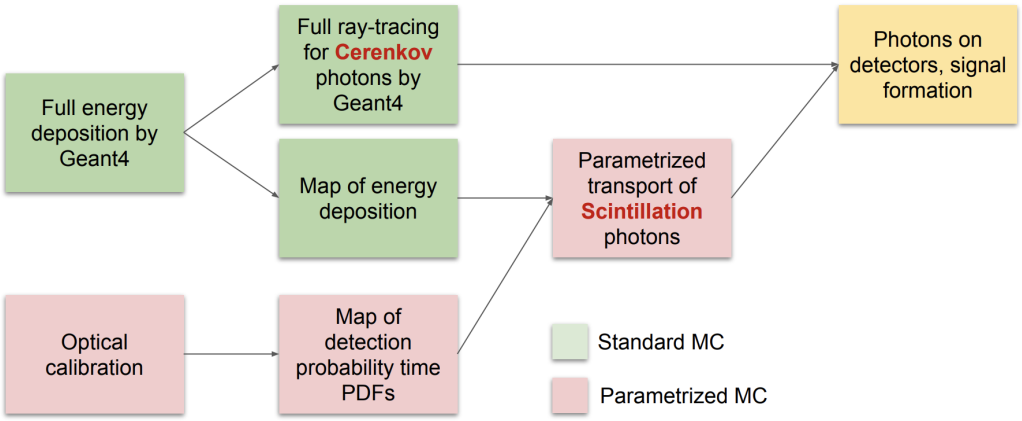


Figure 6.11: Schematic flowchart of the Hybrid-MC simulation framework used in the context of the upgraded ECAL studies. Full-simulation steps are highlighted in green, while parametrized steps are highlighted in red.

6.4.2 High p_T electron reconstruction

The main purpose of this Section is to present reconstruction performance for high p_T electrons in the future LHCb upgrades. To do so, a framework to perform simulation, reconstruction, and analysis of the events has been developed. This framework can be summarised in the following steps:

1. the Hybrid-MC simulation framework described in Section 6.4.1 is interfaced with the full LHCb simulation in the following way:
   (a) the chosen physics process is generated, using standard MC generators as PYTHIA8[29];
   (b) the generated events are passed through the actual LHCb detector simulation and particle-matter interactions are simulated using GEANT4[71, 72] as described in Section 2.2.9. In these studies, the LHCb detector geometry considered for Run 3 is considered. Position and momenta of charged particles are stored to create tracks;
   (c) the detector simulation is performed up to the ECAL surface. A "flux" files collecting all the particles reaching the ECAL surface is created. The flux files is used as input for the Hybrid-MC ECAL simulation, in order to simulate the ECAL response. The output of the Hybrid-MC ECAL simulation is the energy collected by each ECAL cell;

2. from energy deposition in ECAL cells, clusters are reconstructed by combining the cells' energies, using a cellular automaton algorithm similar to the one used in the standard LHCb reconstruction algorithm. Clusters are formed by 3×3 cells [105];

3. clusters' energy and position are corrected by applying energy correction (E correction) and cluster position corrections (L and S corrections), closely following the official LHCb framework [105];

4. clusters are combined with tracks to reconstruct objects and perform analysis.

A flowchart of the pipeline used to perform these studies is shown in Figure 6.12.

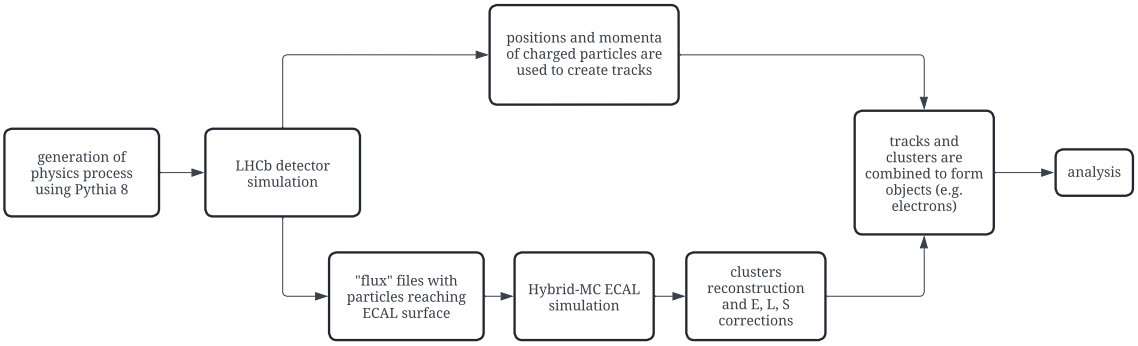


Figure 6.12: Schematic flowchart of the pipeline used to perform $Z \to ee$ studies for the upgraded ECAL. The pipeline can be summarised in four main steps: event generation, Hybrid-MC simulation of ECAL, reconstruction of ECAL clusters, and analysis of reconstructed events.

The simulation of ECAL in upgraded conditions is very time-consuming, particularly due to the increased pile-up with respect to Run 2 conditions. To perform a faster event simulation, signal and minimum bias (MB) processes are simulated separately and in parallel. The merging of signal and MB is performed at pulse level in the Hybrid-MC simulation, to get a realistic ECAL occupancy for the various configurations.

In order to study high p_T electrons, the process $Z \to e^+e^-$ has been considered. 10000 events of $Z \to e^+e^-$ have been simulated using PYTHIA8, requiring both electrons to be in the LHCb acceptance and with $p_T > 10$ GeV. Figure 6.13 shows several important variables at the generator level: the p_T of the electrons, the electron pseudorapidity η, the invariant mass $m(ee)$ and the true energy of the electrons E_e. It can be already seen that the energy range for this kind of process (and generally speaking for EW processes) starts at ~ 100 GeV.

The $Z \to ee$ events are propagated through the LHCb detector up to the ECAL surface. Figure 6.15 shows the vertices where particles are produced in 100 $Z \to ee$ events, in the yz plane of the detector, from the interaction point up to the ECAL surface. Already from this plot, a few considerations can be done:

- electrons and positrons coming from $Z \to ee$ decay tends to be very energetic, and therefore the curvature due to the magnetic field is rather low, obtaining tracks that are quite straight;

- a lot of particles are generated from the interaction of the neutron shielding [106] placed in front of the ECAL.

The last consideration is indeed confirmed by plotting the "event display" of the same 100 events, this time by plotting separately electrons and photons. This is shown in Figure

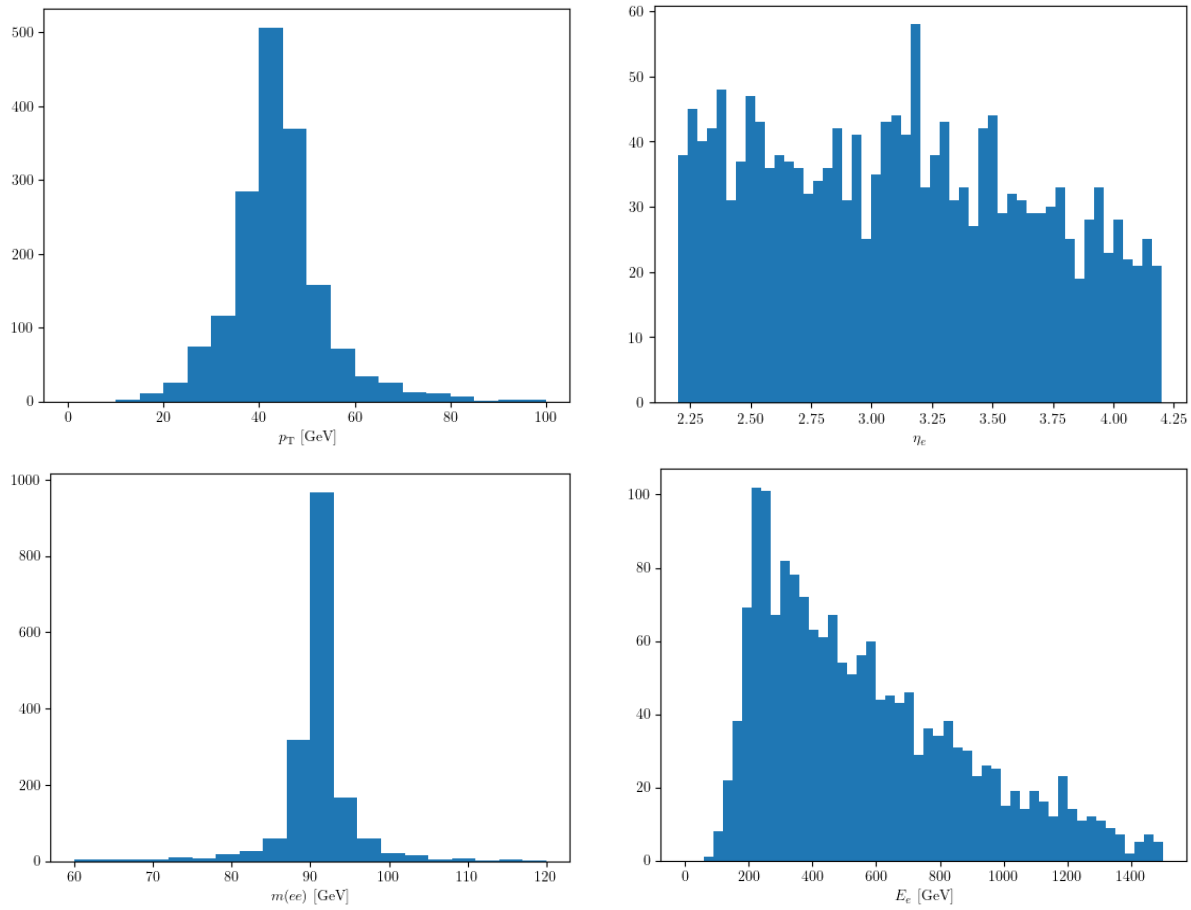


Figure 6.13: p_T, η and E distributions for electrons coming from $Z \to ee$ process, and invariant mass distribution $m(ee)$ for electron-positron pairs.

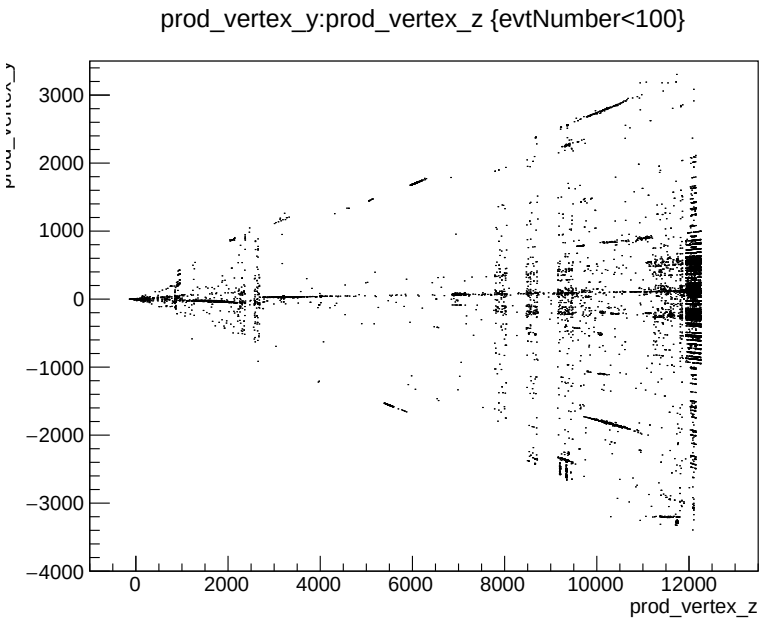


Figure 6.14: Position of production vertices of particles for 100 $Z \to ee$ events.

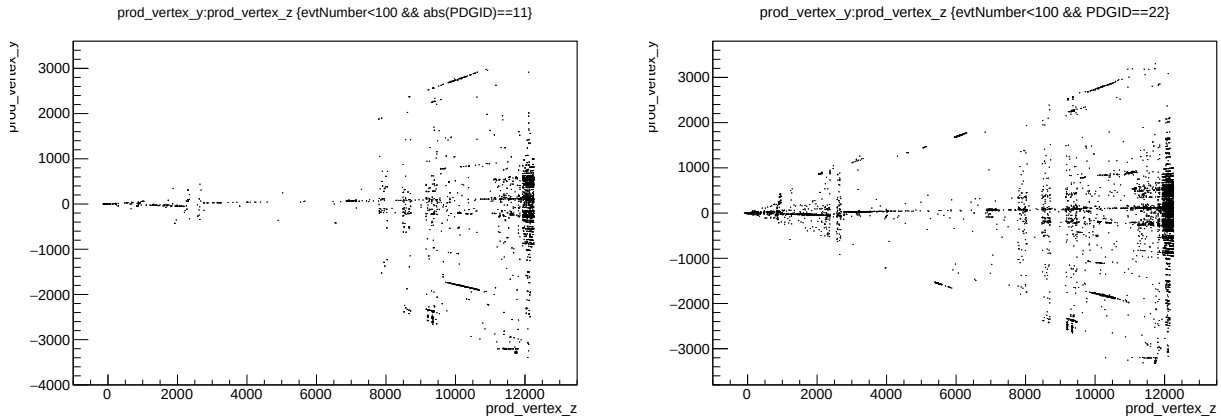


Figure 6.15: Position of production vertices of electrons (left plot) and photons (right plot) for 100 $Z \to ee$ events.

Before performing the cluster reconstruction, the ECAL response for the $Z \to ee$ signal is merged with the MB ECAL response. Once ECAL clusters have been reconstructed, electrons are reconstructed by combining information coming from tracks and ECAL clusters. In particular, here a cluster-based approach is considered to reconstruct electrons, where the electron energy is reconstructed by just looking at ECAL clusters. To reconstruct electrons, the following strategy is used:

- tracks are identified: the p_{T} of charged particles at the end of the TT (right before the magnet) is measured, and if $p_{\mathrm{T}} > 10$ GeV the track is selected;
- the track is propagated to the ECAL surface, and its position x_{ECAL}^{track} is measured;
- the ECAL cluster closest to x_{ECAL}^{track} is selected as a cluster produced by an electron or positron coming from $Z \to ee$ process;
- if more than two candidates are found, the two with the highest p_{T} track are selected.

Once the electron candidates have been selected, the Z boson candidate is reconstructed by combining the electrons' information. A preliminary validation of the simulation and reconstruction approach described so far is shown in Figure 6.16: the reconstructed Z invariant mass $m(ee)$ is shown for this simulation and reconstruction approach (left), and for the official LHCb simulation and reconstruction framework, both for $m(ee)$ reconstructed by just using tracks and for $m(ee)$ reconstructed by just using ECAL clusters. The validation is done for the Run 3 conditions, as no official LHCb detector simulation framework is available so far. For tracks, the agreement is quite good. Considering reconstruction with ECAL clusters, the shape is shifted to higher values of $m(ee)$: this effect is explained by considering that $Z \to ee$ events have been simulated with different generator level cuts. While for this simulation, the generator level cut is $p_{\mathrm{T}} > 10$ GeV, in the standard LHCb study the same cut was $p_{\mathrm{T}} > 4$ GeV: this difference, combined with the effect of ADC saturation (which will be explained later), might account for this difference.

In Figure 6.17, the reconstructed Z invariant mass $m(ee)$ is shown for Run 4, and Run 5 conditions. A few considerations are possible:

- despite the increase in occupancy between Run 4 and Run 5, the performance on the mass peak is intact. The slight shift to higher values of $m(ee)$ in Run 5 is indeed explained by the increase in occupancy;
- Run 4 performance is compatible with Run 3: as already described in Section 6.3, the innermost

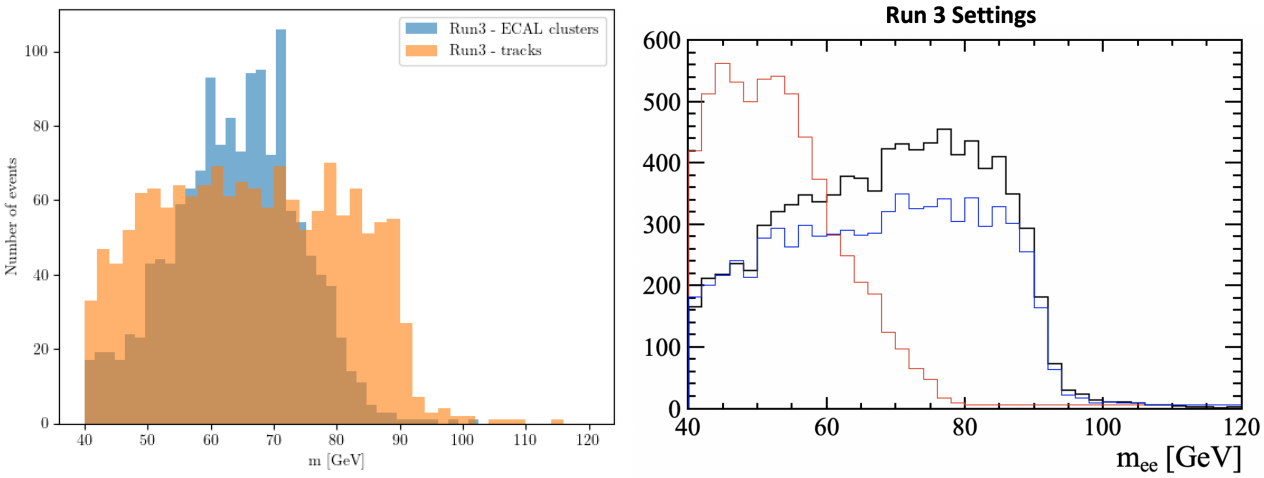


Figure 6.16: (left) Z boson invariant mass $m(ee)$ computed in Hybrid-MC framework using only track information (orange distribution) or ECAL clusters (blue distribution). (right) Z boson invariant mass $m(ee)$ computed using the standard LHCb framework usin only track information (blue distribution), ECAL clusters (red distribution) and reconstructed electrons (black distribution).

region will be replaced at the end of Run 3, but to what concerns the $Z \to ee$ reconstruction there are no visible differences.

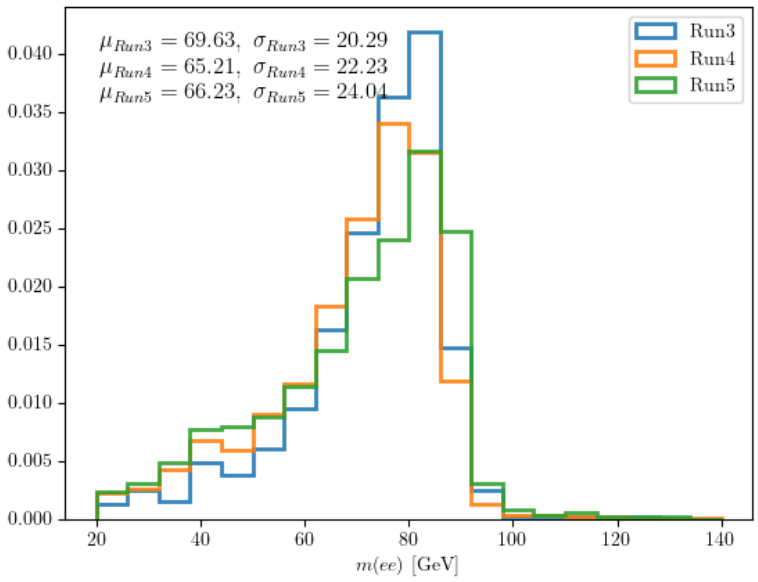


Figure 6.17: Z invariant mass $m(ee)$ for Run 3 (blue plot), Run 4 (orange plot), and Run 5 (green plot) conditions. Distributions are normalized to one. Despite the increase in occupancy, Run 5 performance are comparable to Run 3 and Run 4.

To assess the quality of the electron reconstruction, the reconstructed electron energy E_{reco} is compared to the true electron energy E_{true}. This is presented in Figure 6.18, where the electron/positron energy resolution $(E_{\text{true}} - E_{\text{reco}})/E_{\text{true}}$ evaluated in Run 5 conditions is shown.

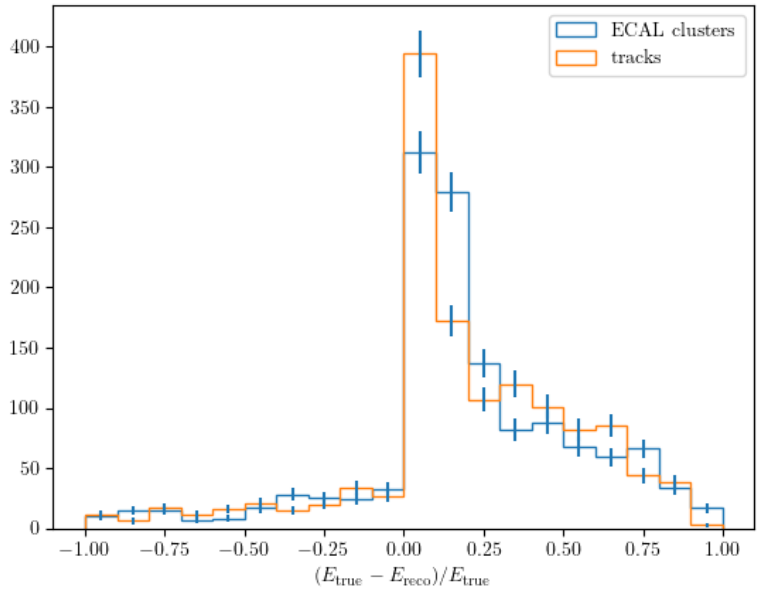


Figure 6.18: Electron energy resolution computed using only track information (orange plots) or using energy deposited in ECAL clusters (blue plots).

Bremsstrahlung recovery

The performance of the electron reconstruction presented so far does not take into account the recovery of the energy lost by Bremsstrahlung emission. This indeed might explain the right tail in Figure 6.18, since by Bremsstrahlung emission electrons lose energy. The main causes of the emission of Bremsstrahlung radiation are:

- interaction with detector material, where the charged particles scatter from a virtual photon of the Coulomb field of a nucleus of the material, with the addition of a real outgoing photon;
- interaction between the charged particle and the magnetic field.

In the LHCb experiment, the first cause is dominant: it has been estimated that an electron traveling through the LHCb magnet will lose less than 0.1% of its energy. Therefore, the emission of Bremsstrahlung radiation from the LHCb magnetic field is negligible. When considering Bremsstrahlung emission due to interaction with detector material, there are two possible scenarios, depending on where the Bremsstrahlung emission started:

- if the Bremsstrahlung emission happens after the magnet, typically the Bremsstrahlung photons are contained in the same ECAL cluster of the electron;
- if the Bremsstrahlung emission happens before the magnet, Bremsstrahlung photons might create another ECAL cluster separated from the electron one, given that electrons will bend inside the magnetic field.

These two scenarios are graphically represented in Figure 6.19.

In particular, it is possible to estimate the region where the Bremsstrahlung photons are more likely to generate ECAL clusters relative to the electron position. To do so, the distance in the ECAL surface between the signal electron/positron and the related emitted Bremsstrahlung photons is evaluated: given the electron/positron coordinates on the ECAL surface (x_e, y_e) and the Bremsstrahlung photon

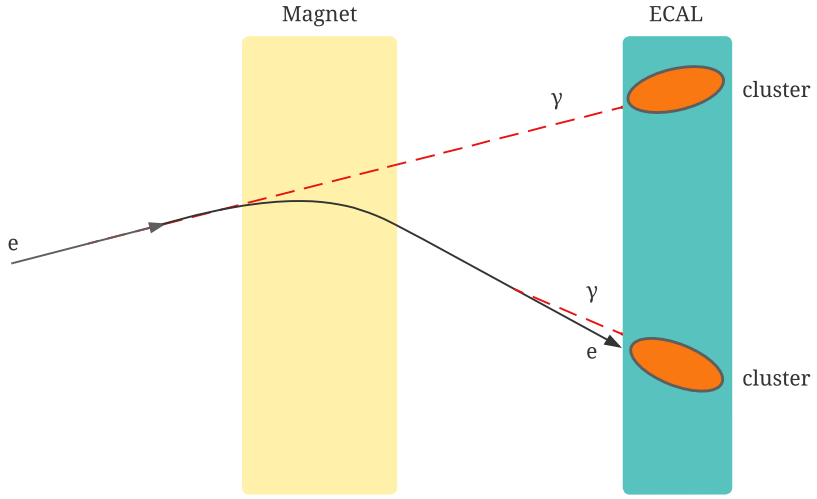


Figure 6.19: Schematic representation of the emission of Bremsstrahlung radiation before the magnet. If Bremsstrahlung radiation is emitted before the magnet, it can create an ECAL cluster spatially separated from the electron cluster.

coordinate (x_γ, y_γ), the distance is computed as

$$d = \sqrt{(x_e - x_\gamma)^2 + (y_e - y_\gamma)^2} \tag{6.2}$$

while the distances along the x and y direction are computed as

$$\begin{aligned} d_x &= |x_e - x_\gamma| \\ d_y &= |y_e - y_\gamma| \end{aligned} \tag{6.3}$$

Results are shown in Figure 6.20, where two main results are evident:

- despite being emitted before the magnet, Bremsstrahlung photons are quite close to the signal electron/positron. This means that when reconstructing the ECAL cluster, the Bremsstrahlung photon might not be resolved from the signal electron/positron, therefore contributing to the same cluster;
- $d_x > d_y$, which is expected since the bending of the charged particles is on the horizontal plane.

By measuring the true distance d, it is possible to check whether the Bremsstrahlung photons are contained inside the electron cluster: this happens 99% of the time, which confirms the fact that the electrons are so energetic that they do not bend inside the magnetic field, therefore emitting Bremsstrahlung photons that are collinear. In conclusion, for this specific case, Bremsstrahlung emission is not an explanation of the right tail of Figure 6.18. An evaluation of the energy loss by $Z \to ee$ electrons due to ionization effects has to be done.

ADC saturation

In the simulation and reconstruction framework presented so far, the effect of the ADC saturation has not been taken into consideration. Given that this is a quite limiting aspect for several EW analyses, this effect has been implemented, yet in a preliminary form, in the Hybrid-MC simulation framework. In Run 1 and Run 2, electric signals in both ECAL and HCAL have been read out by a 12-bits ADC,

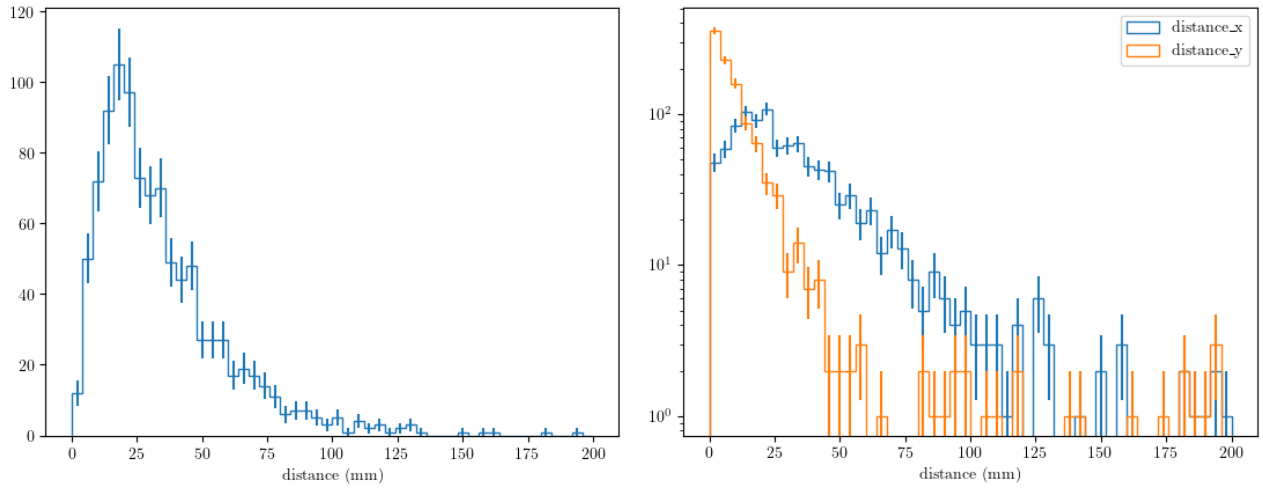


Figure 6.20: Distance d (left) and distances d_x and d_y (right) of Bremsstrahlung photons relative to the electron at the ECAL surface.

which amounts to $2^n - 1 = 2^{12} - 1$ counts. The ADC is typically calibrated in terms of transverse energy E_T deposited in a cell, defined as

$$E_T = E \cdot \sin(\theta), \tag{6.4}$$

where E is the energy deposited in the cell and θ is the angle of the cell with respect to the interaction point. A certain ADC range has to be chosen by tuning the gain of the PMTs which read the deposited energy in the cells: this aspect reflects on the maximum transverse energy E_T^{max} measurable by a cell in the following way:

$$\text{gain}_{PMT} = \frac{Q_{max} \cdot \sin(\theta)}{E_T^{max} \cdot e \cdot N_{phe}} \tag{6.5}$$

where Q_{max} is the ADC saturating charge and N_{phe} is the number of photo-electrons per GeV measured in the PMTs. It is evident from Equation 6.5 that fixing the PMT gain automatically fix the E_T^{max} measurable by a cell. During Run 1 and Run 2, the PMT gain was set to obtain $E_T^{max} = 10$ GeV, which converted to energy means that

$$\begin{aligned} E_{max} &\sim 30 \text{ GeV (outer region)} \\ E_{max} &\sim 300 \text{ GeV (inner region)} \end{aligned} \tag{6.6}$$

Looking at Figure 6.13, it is evident that this cut heavily affects EW processes. In Run 3, the ADC saturation has been relaxed to $E_T^{max} = 20$ GeV, which allowed a minimum gain in performance for the $Z \to ee$ reconstruction peak. In the Hybrid-MC simulation framework, a preliminary yet reasonable ADC saturation has been applied:

- for each cell, $E_{max}(E_T^{max})$ is evaluated for a given value of E_T^{max};
- if the energy measured by a cell $E_{cell} < E_{max}$, then the cell measures E_{cell};
- otherwise, the cell measures E_{max}.

Figure() shows the effect of the ADC saturation on E_{cell} for $E_{max} = 20$ GeV.
For both Run 4 and Run 5 configurations, the Z boson invariant mass has been evaluated for different values of the ADC saturation E_T^{max}: ideal case without ADC saturation, $E_T^{max} = 20$ GeV and $E_T^{max} = 40$ GeV. Results are shown in Figure 6.21: compared to the "no ADC" configuration, the $E_T^{max} = 40$

GeV configuration has the same performance, both in Run 4 and Run 5 configurations. Therefore, it seems there is no need to further increase the ADC threshold $E_{\mathrm{T}}^{\mathrm{max}}$. This is actually a positive aspect, in view of the proposal for Run 5: indeed the idea is to consider double gain electronics, with gains optimized for the low and the high energy regimes. As the main result of this study, $E_{\mathrm{T}}^{\mathrm{max}} = 40$ GeV can be a reasonable proposal for the high energy gain.

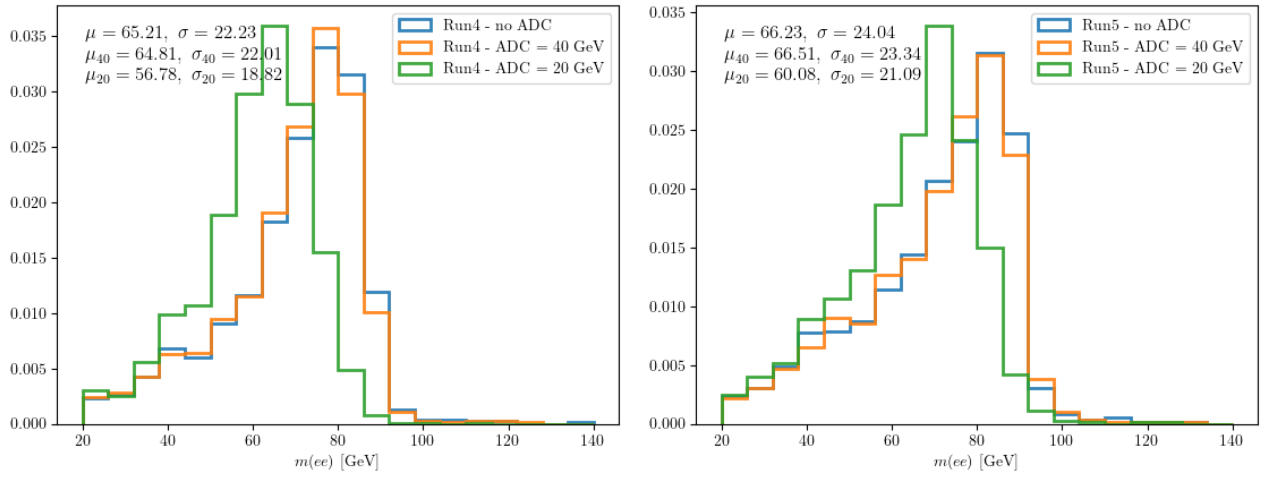


Figure 6.21: Z invariant mass $m(ee)$ for Run 4 (left) and Run 5 (right) conditions, for different values of ADC saturation $E_{\mathrm{T}}^{\mathrm{max}}$: no saturation (blue plot), $E_{\mathrm{T}}^{\mathrm{max}} = 20$ GeV (green plot), and $E_{\mathrm{T}}^{\mathrm{max}} = 40$ GeV (orange plot). Distributions are normalized to one.

6.5 HCAL removal

As presented at the beginning of this Chapter, the future upgrades of the LHCb experiment will lead whole sub-detectors to higher occupancy levels. Particularly, studies performed on the muon system have shown that shielding will be necessary to protect the muon chambers from the increase in radiation. This has led to the proposal of removing HCAL during LS3 and substituting it with a new shielding [107]. Figure. 6.22 shows a scheme of the iron+concrete shielding proposed. While HCAL major impact was in the L0 hardware trigger, which has been removed at the beginning of Run 3, its removal might affect high p_{T} physics, particularly the reconstruction of fundamental EW objects such as electrons and jets. The reconstruction of electrons in the LHCb future upgrades has been already assessed in the previous Section by measuring ECAL performance. Moreover, HCAL removal will have some impact on the PID performance: so far at LHCb, high p_{T} electrons are selected by requiring that the energy released by the electron in HCAL, normalized over the electron momentum, is under a specific threshold; it is evident that, by removing HCAL, this requirement is not available anymore, and therefore new strategies to identify high p_{T} electrons need to be found. In the context of this thesis work, two studies are presented in this Section:

- impact of HCAL removal on jet reconstruction;
- PID performance evaluated on the new proposed SpaCal [104] modules.

Throughout this Section, the iron core-iron+concrete muon shielding configuration will be referred to as the “absorber” configuration.

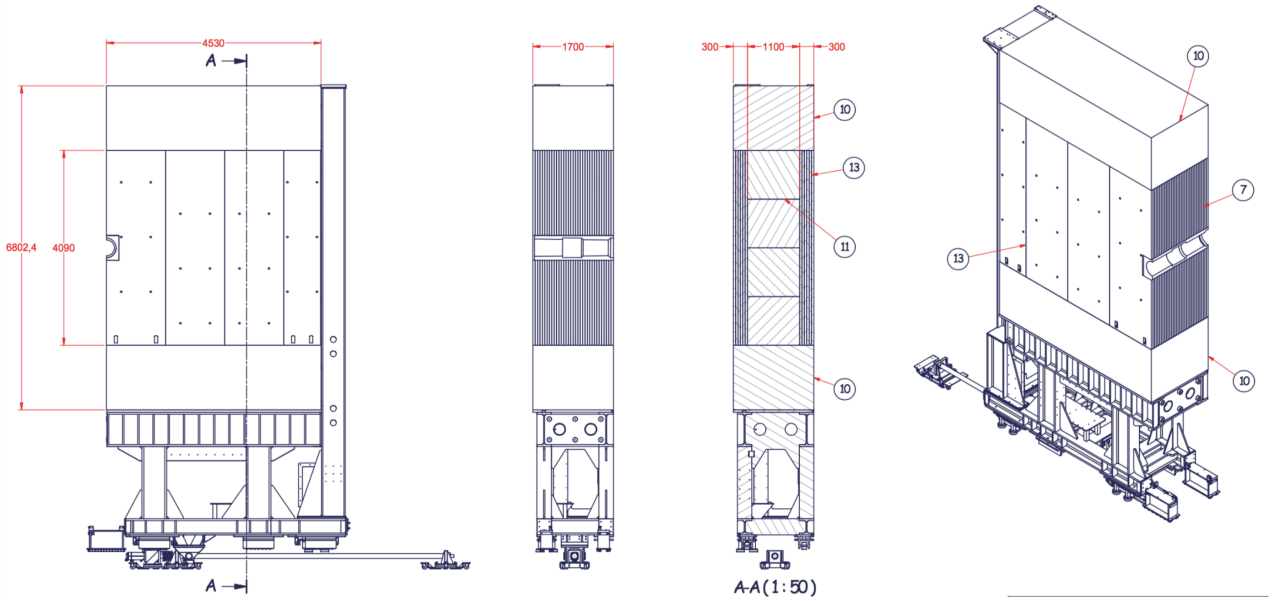


Figure 6.22: Design of the iron core/iron+concrete muon shielding proposed to replace HCAL from Run 5 [107].

6.5.1 Impact of HCAL removal on jet reconstruction

Jets reconstruction and its performance have already been presented in Chapter 3. In Section 3.3 it has been shown that almost 10% of the particle content of the jet comes from neutral hadrons: the removal of HCAL will therefore affect the performance in the jet reconstruction. To quantify the impact of HCAL removal in the jet reconstruction performance, a study using simulations has been developed, following this strategy:

- $Z \to b\bar{b}$ events have been generated with PYTHIA8 [29] in Run 4 conditions;
- the detector response has been simulated with the official LHCb simulation framework described in Section 2.2.9, considering two different conditions: one with the standard HCAL, and the other one with the absorber in place of HCAL;
- standard LHCb reconstruction tools have been used to reconstruct jets (as described in Chapter 3, adding some selection requirements:

$$\begin{aligned} p_{\mathrm{T}} &> 20 \text{ GeV} \\ 2.2 < \eta &< 4.2, \end{aligned} \tag{6.7}$$

  where p_{T} and η are respectively the transverse momentum and the pseudorapidity of the jets. These are standard cuts used in $Z \to b\bar{b}$ analysis to ensure that the QCD background is suppressed and jets are contained in the instrumented region of the calorimeters;
- once reconstruction is performed, the two jets with the highest p_{T} are taken, and a matching procedure to true MC jets is applied by requiring that the distance between reconstructed and MC jets $\Delta R = \sqrt{(\eta_{\mathrm{MC}} - \eta_{\mathrm{reco}})^2 + (\phi_{\mathrm{MC}} - \phi_{\mathrm{reco}})^2} < 0.5$.

Jet Energy Correction

Several figures of merits are used to evaluate the impact of HCAL removal on jet reconstruction performance. As already explained in Section 3.3, the Jet Energy Correction takes into account the

differences between true jets and reconstructed jets, by computing

$$\mathcal{K}(p_{\mathrm{T}}, \eta, \mathcal{N}) = \frac{E_{\mathrm{reco}}}{E_{\mathrm{MC}}} \tag{6.8}$$

where $\mathcal{N}$ is the number of PVs in the proton-proton collisions, and E_{reco} (E_{true}) is the reconstructed (true) energy of the jet. $\mathcal{K}(p_{\mathrm{T}}, \eta, \mathcal{N})$ does not have an analytical form, therefore it has to be computed for each bin of p_{T}, η and $\mathcal{N}$ so that it can be fitted and used afterward to estimate its effect on physics analyses. Results of $\mathcal{K}(p_{\mathrm{T}}, \eta, \mathcal{N})$ fit, projected separately as a function of p_{T}, η and $\mathcal{N}$ are shown in Figure 6.23: already from this quick evaluation, it is evident that $\mathcal{K}(p_{\mathrm{T}}, \eta, \mathcal{N})$ is higher when HCAL is replaced by the absorber.

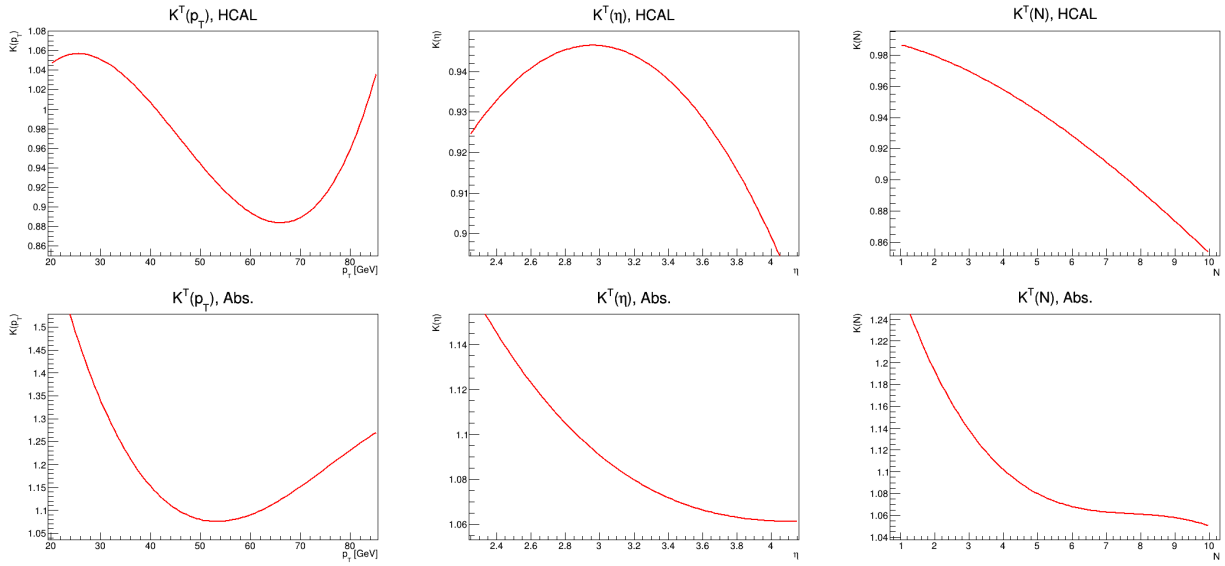


Figure 6.23: Projection of fit to $\mathcal{K}(p_{\mathrm{T}}, \eta, \mathcal{N})$ as a function of p_{T} (left column), η (central column) and $\mathcal{N}$ (right column) for the HCAL configuration (first row) and the absorber configuration (second row).

Jet Energy Resolution

Once the Jet Energy Correction has been computed, the Jet Energy Resolution can be evaluated as

$$\mathcal{R} = \frac{E_{\mathrm{true}} - \bar{E}_{\mathrm{reco}}}{E_{\mathrm{true}}} \tag{6.9}$$

where $\bar{E}_{\mathrm{reco}}$ is the reconstructed energy of the jet, corrected for the Jet Energy Correction $\mathcal{K}(p_{\mathrm{T}}, \eta, \mathcal{N})$ computed previously. If no bias is present, $\mathcal{R}$ should have a gaussian shape with a mean equal to 0. The standard deviation of $\mathcal{R}$ is the actual Jet Energy Resolution. Therefore, a first check comes from fitting the distributions of $\mathcal{R}$ with and without HCAL, as shown in Figure 6.24.

Fit results are

$$\mathcal{R}_{\mathrm{HCAL}} = (15.6 \pm 0.3)\% \quad \mathcal{R}_{\mathrm{noHCAL}} = (17.5 \pm 0.4)\%, \tag{6.10}$$

where a relative loss in jet energy resolution performance of about 10% is found. The Jet Energy Resolution can be also computed as a function of the jet p_{T}, by basically performing the same gaussian fit in every jet p_{T} bin. Results are shown in Figure 6.25 for both configurations. Interestingly, the absorber configuration seems to improve the jet energy resolution in the low p_{T} range, while for $p_{\mathrm{T}} > 40$ GeV the absorber configuration has worse performance than the HCAL configuration.

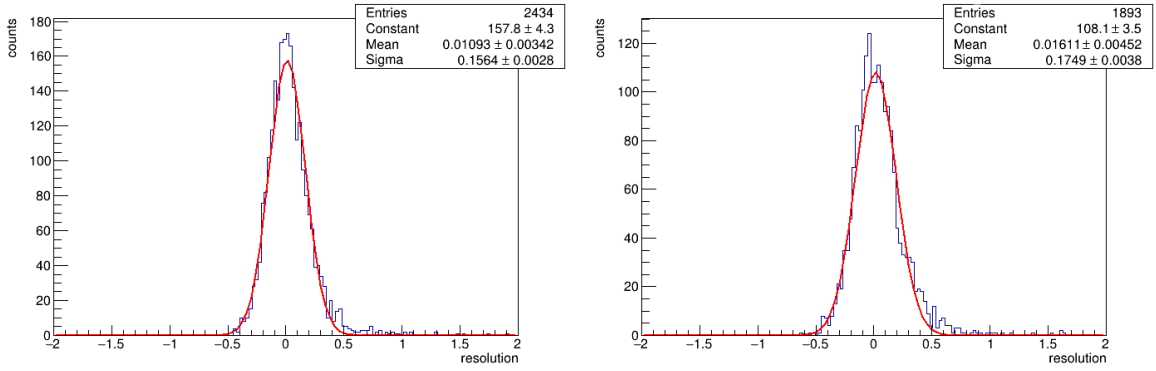


Figure 6.24: (left) Jet Energy Resolution $\mathcal{R}$ for the HCAL configuration. (right) Jet Energy Resolution $\mathcal{R}$ for the absorber configuration.

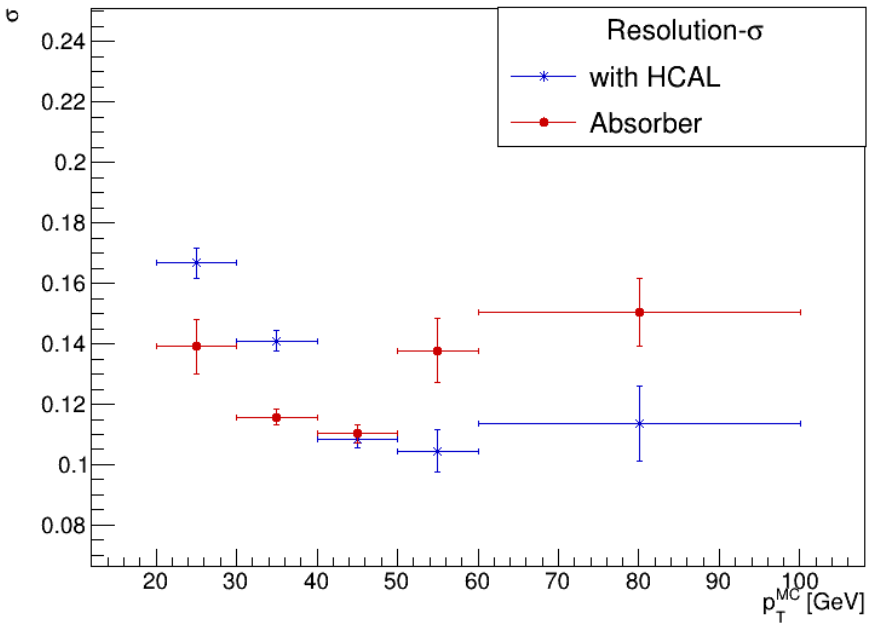


Figure 6.25: Jet Energy Resolution as a function of the jet p_{T}, for the HCAL (blue dots) and the absorber (red dots) configurations. An improvement in the jet energy resolution in the low p_{T} range is found for the absorber configuration.

Jet reconstruction efficiency and fake probability

The jet reconstruction efficiency $\varepsilon_{\mathrm{jet}}$ is defined as the number of reconstructed jets over the number of true jets in a specific jet p_{T} range; the fake probability p_{fake} is the number of fake jets (where a fake jet is a jet not matched to a true MC jet) over the number of reconstructed jets. $\varepsilon_{\mathrm{jet}}$ and p_{fake} as a function of the true jet p_{T} are shown in Figure 6.26.

From the $\varepsilon_{\mathrm{jet}}$ trend is evident that there is a loss in efficiency reconstruction when removing HCAL, up to 5% in the worst case, particularly in the low p_{T} range. Considering the fake probability p_{fake}, an increase is evident in the $20 - 30$ GeV p_{T} range, almost doubling the fake probability for the HCAL configuration. For higher p_{T} values results are compatible between HCAL and absorber configurations. The following considerations can be drawn from this study:

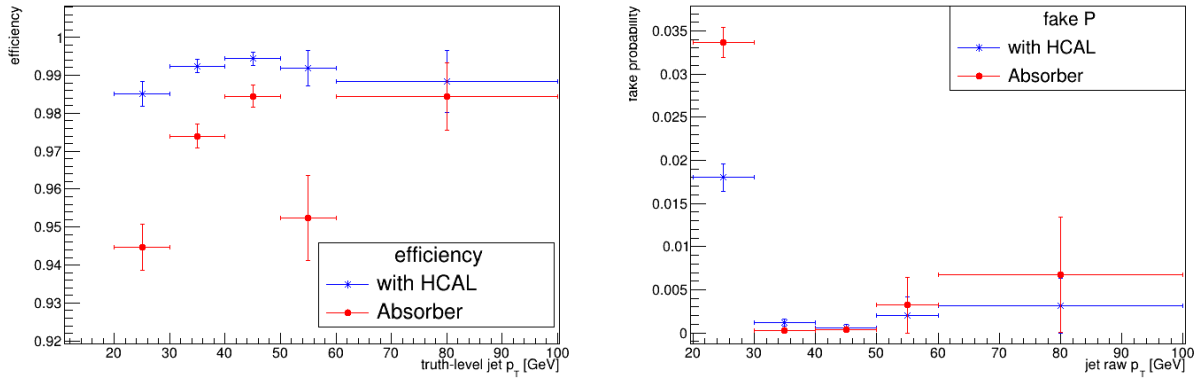


Figure 6.26: (left) Jet reconstruction efficiency ε_{jet} as a function of jet p_{T} for HCAL configuration (blue dots) and absorber configuration (red dots). (right) Fake probability p_{fake} as a function of jet p_{T} for HCAL configuration (blue dots) and absorber configuration (red dots).

- in the absorber configuration, the Jet Energy Correction factor $\mathcal{K}(p_{\text{T}}, \eta, \mathcal{N})$ is on average higher with respect to the HCAL configuration, and this is expected since the measurement of the neutral hadronic component of the jet is completely lost;
- this bigger correction factor induces on average a worse Jet Energy Resolution of about 10%;
- in the absorber configuration, the Jet Energy resolution improves in the low jet p_{T} range, but this improvement is followed by a decisive worsening of the jet reconstruction efficiency and the fake probability. A possible explanation of the better Jet Energy Resolution is the fact that the lower jet reconstruction efficiency might select jets that show better Jet Energy Resolution. Therefore, this improvement might be just a fictitious one.

6.5.2 SPACAL PID performance

While being fundamental for high p_{T} objects reconstruction, HCAL plays a role also in the PID performance. Given the Hybrid-MC simulation framework presented in Section 6.4.1, some studies have been performed to understand if some PID performance can be regained when considering solely the upgraded ECAL geometry. As already explained in Section 6.3, the new SpaCal [104] technology substitutes the Shashlik modules in the inner part of ECAL, with the possibility of longitudinal segmentation of the modules. Indeed, longitudinal segmentation might be useful in getting some PID discrimination between particles, particularly electrons versus hadrons: the possible role of longitudinal segmentation is evident in Figure 6.27, where the xz plane of the shower propagation in a SpaCal module is shown for an electron and a charged pion with energy $E = 20$ GeV. To study the PID performance of the new Spacal modules, a single SpaCal W-GAGG module has been considered, and the full Hybrid-MC simulation framework described in Section 6.4.1 has been used. Different kinds of particles (electrons, muons, charged pions, and charged kaons,) have been shot singularly into the SpaCal module, and the energy deposited in the fibers of each cell is measured. A "particle gun" approach has been used, therefore no detector material is present in front of the module: this means that particularly for electrons, no Bremsstrahlung radiation is produced (see Section 6.4.2). Particles at different energies have been simulated, namely, 1, 2, 3, 4, 5, 20, 40, 60, 80, and 100 GeV: this choice has been done to match the energies considered when collecting data during test-beam campaigns at SPS and DESY facilities[3]. The full SpaCal W-GAGG module is considered, therefore all the energies

[3]The general idea is to perform this study on SpaCal W-GAGG module simulations, and then apply it on test-beam data.

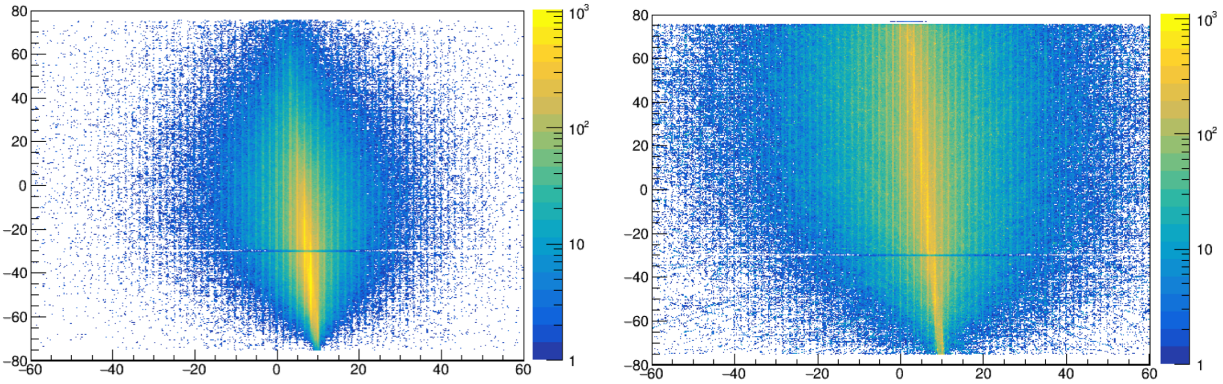


Figure 6.27: (left) Energy deposited in a SpaCal W-GAGG module by an electron of initial energy $E = 20$ GeV. (right) Energy deposited in a SpaCal W-GAGG module by a charged pion of initial energy $E = 20$ GeV.

from the 128 cells (64 cells in the front section and 64 in the back section) are considered. Particles are shot in the center of the module, to ensure that the electromagnetic shower is well contained in the module and therefore the whole energy is measured.

Figure 6.28 presents the energy measured in the front section, in the back section, and the total measured energy for particles (electrons, muons, charged pions, and charged kaons) hitting the SpaCal module with energies $E = 1, 20, 100$ GeV. A few considerations are possible:

- muons leave little energy in the SpaCal module, with little dependence on the initial energy;
- pions and kaons distributions are exactly the same, while no evident dependence on the initial energy is present;
- with increasing initial energy, the fraction of electron energy measured in the back section increases, which is expected since the shower maximum of the electron is expected to move towards the back of the module.

The combination of these plots is shown in Figure 6.29: a reasonable separation between different kinds of particles can be already obtained by eye, by simply cutting on different values of E_{front}, E_{back} and E_{total}. Here, a more refined approach is used: a Convolutional Neural Network (CNN) which uses the energy deposited in each cell is considered. CNNs are known for their ability to effectively handle data objects that have a geometrical and well-defined structure. Images are excellent candidates for CNN classification, given that they are described by pixels, where each pixel carries a property to describe its color. In this specific situation, the energy released in the SpaCal module can be thought of as pixels of an image, where the pixel structure is indeed the geometrical cell position (a square of 8×8 cells), and each pixel is described by two values: the energy in the front E_{front} and in the back section E_{back}. Therefore for this specific application, the CNN takes as input 128 features (energy released in 64×2 cells), arranged in a matrix obtained from the module geometry. The CNN structure is shown in Figure 6.30: the convolutional structure is made of 6 layers with $[64, 64, 43, 43, 16, 16]$ nodes each, and it processes information coming from the single cells; then, a dense layer combines the output of the convolutional layer together with E_{front}, E_{back} and E_{total}; in the end, a softmax activation function returns the probabilities $P(e)$, $P(\mu)$, $P(\pi)$ and $P(K)$ for that particle to be respectively an electron, muon, pion or kaon. The CNN is programmed with the TensorFlow library combined with the Keras plugin. The training procedure is performed for 200 epochs, with gradient descent optimized with the Adam optimizer. A patience checkpoint equal to 20 is considered to optimize the training procedure. A total of 10000 events per particle energy are simulated, resulting in 100000

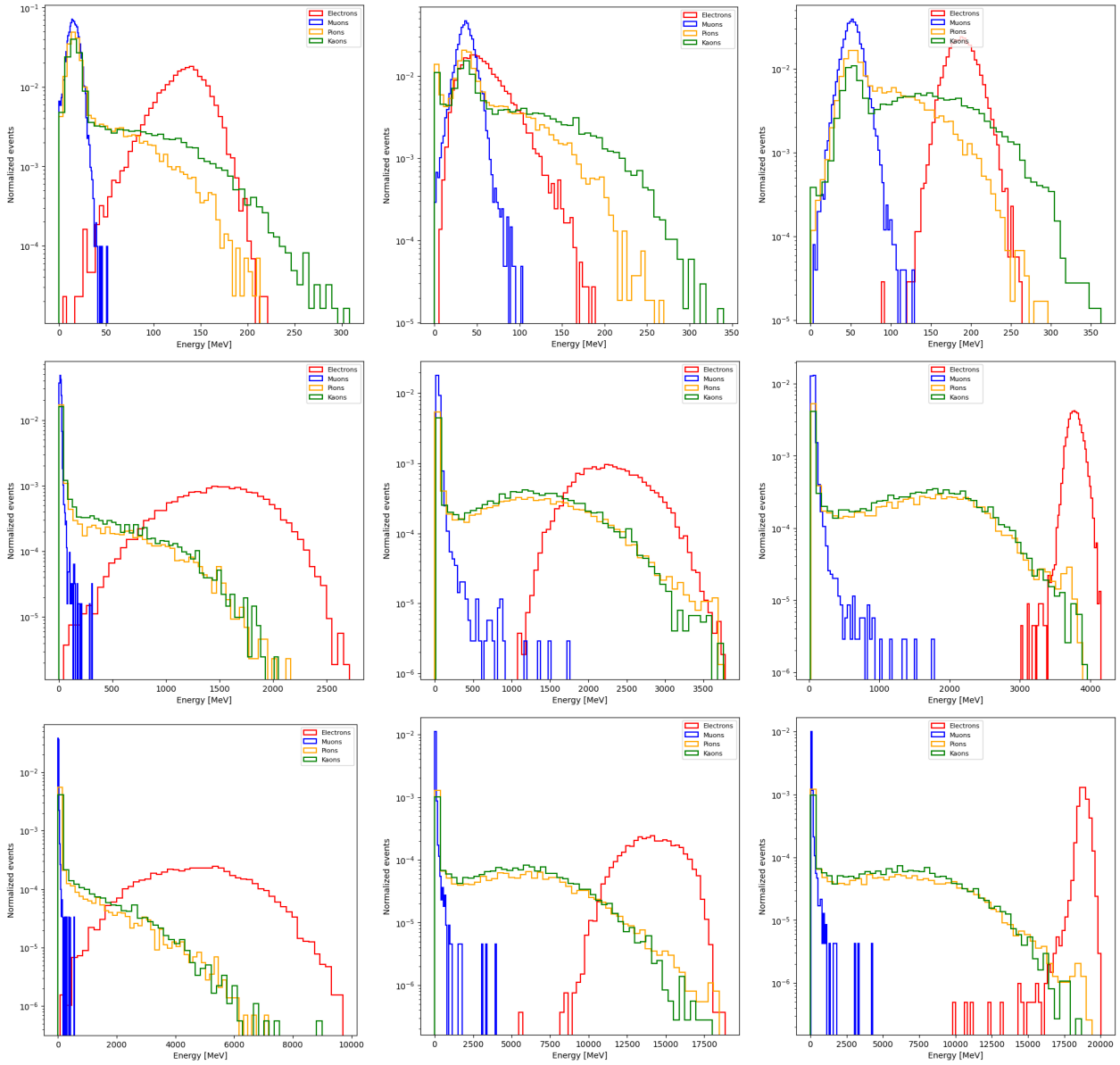


Figure 6.28: Energy measured in the front section E_{front} (left column), the energy measured in the back section E_{back} (central column), and energy measured in whole module E_{total} (right column) for electron, muons, pions and kaons with energy $E = 1$ GeV (first row), $E = 20$ GeV (second row), and $E = 100$ GeV (third row).

events per particle type. The complete dataset amounts to 400000 events, which are split into training, validation, and testing datasets.

The results of the application of the CNN to this PID problem are shown in Figure 6.31, where the ROC curve and the normalized confusion matrix are plotted. The following conclusions can be drawn:

- electrons are perfectly reconstructed, allowing a perfect separation against charged pions;
- the same applies to muons, which are well distinguished from pions and kaons;
- separation between pions and kaons gives no results, but this is expected since pion/kaon separation is mainly handled by the RICH sub-detectors.

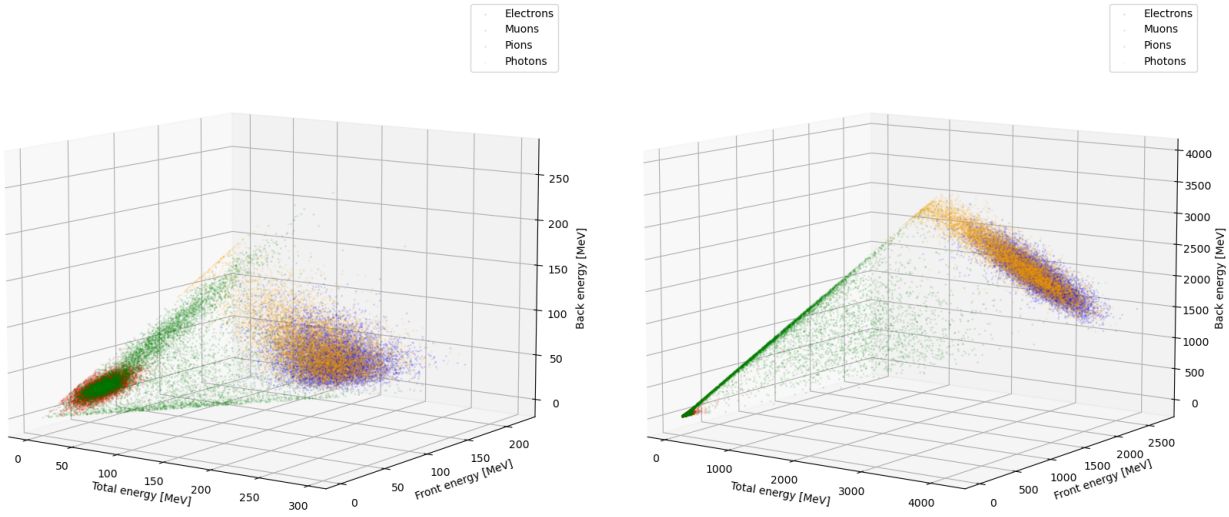


Figure 6.29: 3D scatter plot emphasizing the different distributions of E_{front}, E_{back}, and E_{total} for electrons, muons, pions, and kaons. Distributions are shown for particles with energy $E = 1$ GeV (left) and $E = 20$ GeV (right).

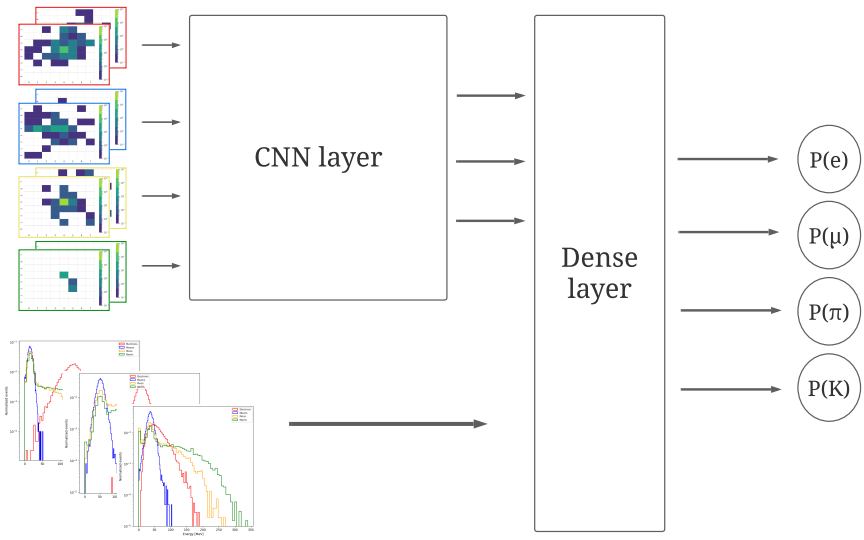


Figure 6.30: Schematical structure of the CNN used for this PID task.

6.6 Final considerations and future prospects

In this Chapter, the performance of the calorimeter system in the future LHCb upgrade has been assessed. Particularly, the upgraded ECAL performance in reconstructing high p_{T} electrons has been studied, showing that with a slightly higher requirement on ADC saturation, it is possible to increase the quality of the Z invariant mass peak. Studies on HCAL removal show that jets reconstruction will be affected, with an average loss in performance of $\sim$ 10%; concerning PID, the longitudinal segmentation of SpaCal modules will help in electron/hadron separation. There are several future studies that need to be performed in order to assess completely the upgraded ECAL performance:

- high p_{T} electron reconstruction studies: these studies still need to assess the energy loss due to

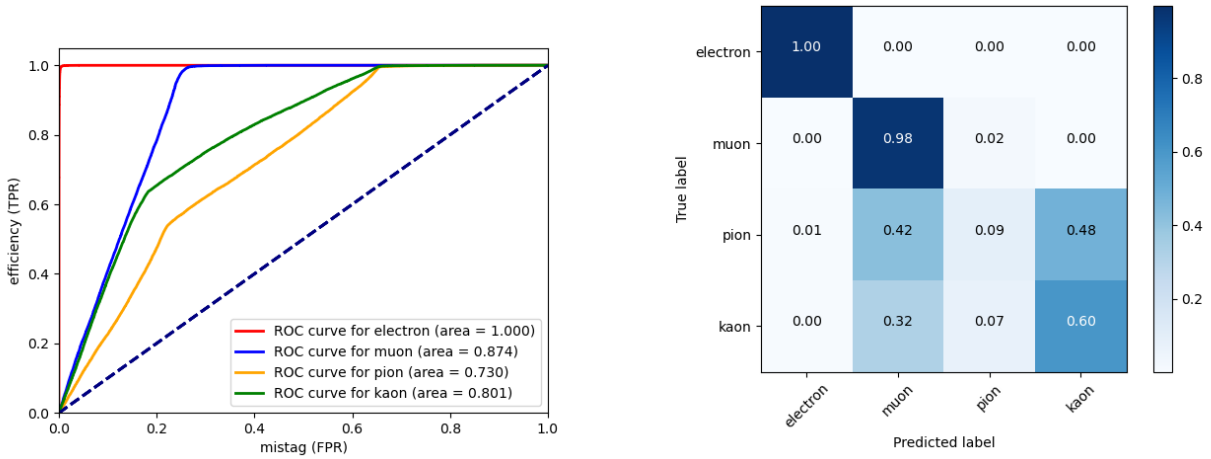


Figure 6.31: ROC curve (left) and confusion matrix (right) for the PID study on a SpaCal W-GAGG module.

ionization effects. Moreover, the impact of the proposed ADC saturation requirements needs to be evaluated also for lower energy processes, e.g. $J/\psi \to ee$;

- jets reconstruction studies: the studies presented in Section 6.5 have several limitations:
  - only the Run 4 configuration has been studied, therefore no performance at higher luminosities has been studied;
  - the considered detector geometry do not take into consideration possible improvements coming from upgrades of other sub-detectors, in particular, the upgraded ECAL studied in Section 6.4;
  - the actual jets reconstruction algorithms have been used, therefore no optimization has been studied so far.

  With these considerations in mind, new studies on jets reconstruction should be pursued;
- ML reconstruction: the studies presented in Section 6.5.2 show that indeed ML tools can be used to gain information by exploiting the higher granularity and the longitudinal segmentation of the upgraded ECAL.

Chapter 7

New classification algorithms for jet identification

Chapter 6 presented improvements from the detector side, by introducing new technologies that might help in coping with the higher occupancies of the HL-LHC phase. Another possible source of improvement comes from the algorithms used in the analysis: in Chapter 5 it has been demonstrated that the WW analysis benefits from the usage of ML techniques, which allows the separation between WW and $W+$ jets events. Better algorithms will lead to better background rejection, allowing for more and more precise measurements. In this thesis, new algorithms for jet identification based on Quantum Machine Learning (QML) techniques are studied. The structure of this Chapter is the following:

- Section 7.1 introduces the fundamentals of Quantum Computation;
- Section 7.2 gives a brief description of a typical QML algorithm, highlighting the differences and similarities with classical ML algorithms;
- Section 7.3 presents a QML application on the b-jet identification task at the LHCb experiment: a Variational Quantum Classifier (VQC) is trained to separate jets produced by b- and $\bar{b}$-quarks using information coming from the jet substructure. The VQC performance is studied and is compared with a DNN using the same input features;
- finally, Section 7.4 deals with future prospects on QML application for jet identification studies.

7.1 Fundamentals of Quantum Computation

QML is based on QC. To better understand QC, it is useful to compare it with Classical Computation. In Classical Computation, the *bit* represents the unit of information, which can assume two values, 0 (*false*) or 1 (*true*). In order to describe a larger number of states, a series of bits can be combined together. In QC, the quantum analog of the bit is called *qubit*: it is a two-level quantum system that can be measured in two states, $|0\rangle$ or $|1\rangle$, which form an orthogonal basis of a 2-dimensional Hilbert space $\mathcal{H}$. The most general system which is described by a qubit is defined as

$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle \tag{7.1}$$

with $\alpha_0, \alpha_1 \in C$ and satisfying the normalization condition

$$|\alpha_0|^2 + |\alpha_1|^2 = 1. \tag{7.2}$$

Given that the Hilbert space $\mathcal{H}$ is isomorphic to $\mathbb{C}^2$, the computational base of $\mathcal{H}$ can be easily associated with the canonical base of $\mathbb{C}^2$ using the following relations:

$$\begin{aligned} |0\rangle \in \mathcal{H} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2 \\ |1\rangle \in \mathcal{H} \rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}^2 \end{aligned} \tag{7.3}$$

With this isomorphism, quantum states can be represented as vectors. A suitable representation of the quantum state $|\psi\rangle$ leverages the usage of spherical coordinates: Equation 7.1 can be rewritten (up to irrelevant global phase) as

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + (\exp i\phi)\sin\frac{\theta}{2}|1\rangle, \tag{7.4}$$

with $0 < \theta < \pi$ and $0 < \phi < 2\pi$. Since θ and ϕ can be interpreted as spherical coordinates, a qubit can be therefore easily represented using the *Bloch sphere*: the qubit state is represented as a $\mathbb{R}^3$ vector pointing from the origin of a $\mathbb{R}^3$ coordinates system up to the surface of a sphere of radius 1, given the normalization condition expressed in Equation 7.2. A graphical representation of the Bloch Sphere is shown in Figure 7.1.

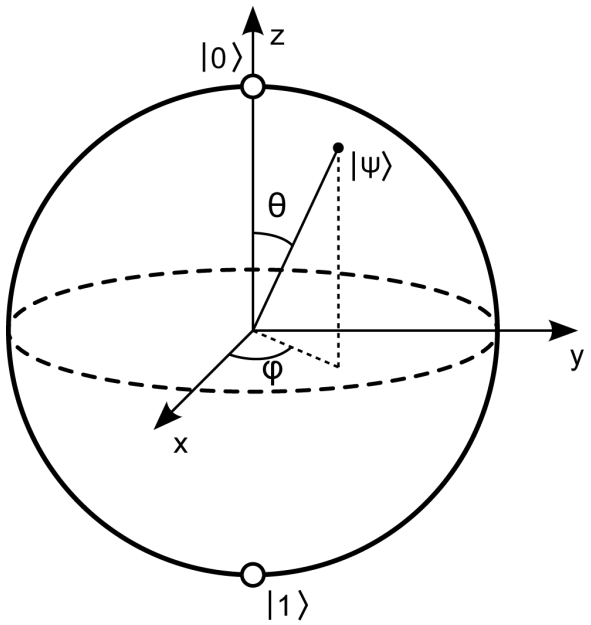


Figure 7.1: Pictorial representation of the Bloch sphere [108].

Equation 7.1 defines a so-called *pure* state, a state that does not interact with the environment. While this is a suitable approach to start working with these tools when dealing with real quantum hardware the quantum state will always interact with the environment surrounding it. States that interact with the environment, and therefore describe a combination of pure states, are called *mixed* states. It is possible to describe pure and *mixed* states using the *density matrix* description: given a pure state $|\psi\rangle$ as defined in Equation 7.1, its density matrix ρ_{pure} is

$$\rho_{\text{pure}} = |\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha_0|^2 & \alpha_0\alpha_1^* \\ \alpha_0^*\alpha_1 & |\alpha_1|^2 \end{pmatrix}. \tag{7.5}$$

ρ_{pure} is therefore a projector operator, satisfying the following properties:

- ρ_{pure} is a hermitian operator, $\rho_{\text{pure}}^{\dagger} = \rho_{\text{pure}}$
- ρ_{pure} is semi-positive defined, $|\phi\rangle : \langle\phi| \rho_{\text{pure}} |\phi\rangle$
- ρ_{pure} is a projector, $\rho_{\text{pure}}^2 = \rho_{\text{pure}}$
- the diagonal elements of ρ_{pure} are the probabilities of $|\psi\rangle$ being in a basis state, therefore $\text{Tr}\rho_{\text{pure}} = 1$

For a mixed state, the same density matrix approach can be used: if p_i is the probability associated to a pure state $|\psi_i\rangle$ (with the sum of all probabilities equal to 1), then a mixed state is described by the following density matrix:

$$\rho_{\text{mix}} = \sum_i p_i |\psi_i\rangle \langle\psi_i| . \tag{7.6}$$

This time, ρ_{mix} does not satisfy the same properties of ρ_{pure}, particularly

- $\rho_{\text{mix}}^2 \neq \rho_{\text{mix}}$
- $\text{Tr}\rho_{\text{mix}}^2 < \text{Tr}\rho_{\text{mix}}$

Once the definition of qubit is clear, in principle one would like to combine information from several qubits: in complete analogy with Classical Computation, this means to build a *quantum circuit*. The starting point is to create a *multi-qubit* state, which is based on the concept of *tensor product* between Hilbert spaces. Given n single qubits $|\psi_i\rangle \in \mathcal{H}$ with $i = 1, ..., n$, a n-qubit system $|\Psi\rangle \in \bigotimes_{i=1}^{n} \mathcal{H}$ is defined as

$$|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_n\rangle . \tag{7.7}$$

Using the isomorphism between $\mathcal{H}$ and $\mathbb{C}^n$, $|\Psi\rangle$ can be rewritten on the canonical basis of $\mathbb{C}^n$ as

$$|\Psi\rangle = \alpha_1 |0...00\rangle + \alpha_2 |0...01\rangle + ... + \alpha_{2^n} |1...11\rangle \qquad \text{with} \quad \alpha_i \in \mathbb{C} \tag{7.8}$$

Finally, the same generalization is applied to density matrices and mixed states,

$$|\psi\rangle = \sum_{i=1}^{2^n} \alpha_i |i\rangle \quad \rho_{\text{pure}} = |\psi\rangle \langle\psi| \quad \rho_{\text{mix}} = \sum_{i,j} \alpha_{ij} |i\rangle \langle j| \quad \alpha_i, \alpha_{ij} \in \mathbb{C} \tag{7.9}$$

7.1.1 Quantum Circuits

Different qubits can be combined together into quantum circuits. To properly understand how this works, the classical counterpart is taken into consideration. Indeed, classical computers operate on bit-strings by means of logical functions: a logical function can be represented as a function $f : \{0,1\}^n \to \{0,1\}^m$ that receives a n-bits string as input and gives an m-bits string as output. The relation defining the output bit-string starting from an input bit-string is completely determined by its truth table. Every logical function can be decomposed into elementary logical functions, which are called logical gates: typical logical gates are the AND gate, the OR gate, the NOT gate, or the XOR gate. Starting from these elementary logical gates, logical circuits are built from their combinations.

In analogy with classical computation, qubits can be manipulated by *quantum gates*. Quantum gates are described by unitary operators acting on the quantum state, where a unitary operator O acting on a Hilbert space $\mathcal{H}$ satisfies the requirement $OO^{\dagger} = O^{\dagger}O = \mathbb{1}$, where $\mathbb{1}$ is the identity operator. There are two types of quantum gates:

- 1-qubit gates: they are described by 2×2 complex unitary matrices. Among these, some of the most important are the *Pauli operators*:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7.10}$$

  The X gate is also called NOT gate since its action flips the qubit in the computational basis: $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$. Other typical 1-qubit gates are the *Hadamard gate* (denoted H), and the *phase gate* (denoted S), whose matrix representation is

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{7.11}$$

  Finally, there are 1-qubit gates that rotate the state vector around the x, y, and z axis on the Bloch sphere by an angle θ; these are called *rotation gates*. They are defined as:

$$R_x(\theta) = e^{-i\theta X/2} = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}X = \begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \tag{7.12}$$

$$R_y(\theta) = e^{-i\theta Y/2} = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}Y = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \tag{7.13}$$

$$R_z(\theta) = e^{-i\theta Z/2} = \cos\frac{\theta}{2}\mathbb{1} - i\sin\frac{\theta}{2}Z = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \tag{7.14}$$

- 2-qubits gates: contrary to the previous case, these gates act on two qubits. A typical example is the *controlled-U* gate: this is a 2-qubits operation that applies a U operation on a target qubit, depending on the value set on the control qubit. A controlled operation is represented as in Figure 7.2, where the control qubit (on top) is wired to the target qubit (on bottom) through the control-U gate. The most common controlled gates are the CNOT (which stays for

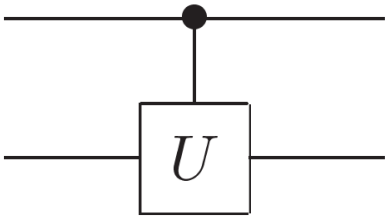


Figure 7.2: Graphic representation of a controlled-U gate.

  Controlled-NOT) and the CZ (which stands for Controlled-Z), respectively applying an X and a Z operation depending on the value of the control qubit.

7.1.2 Entanglement

Entanglement is a fundamental concept of quantum computing that has no classical counterpart. Let A and B be two quantum systems, then a state of the composite system is called *separable* if it can be written as a classical probability distribution p_i over uncorrelated states of the two subsystems ρ_A and ρ_B:

$$\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B$$

A non-separable state is called an entangled state: an example of a non-separable pure state is provided by the Bell states

$$|\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B \pm |1\rangle_A \otimes |1\rangle_B)$$
$$|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |1\rangle_B \pm |1\rangle_A \otimes |0\rangle_B)$$

These states are also called *maximally entangled* states: the outcome of a measurement on the subsystem B can be determined by measuring the subsystem A. This means that the two subsystems are maximally correlated: this kind of correlation is purely quantum and cannot be achieved by means of some sort of classical operations.

7.1.3 Measurement of a quantum system

Once the quantum circuit has been built, one would like to access a quantum state and perform some measurements. To do so, a physical observable related to the system has to be measured. For simplicity, a 1-qubit system is considered, as defined in Eq. 7.1. A measurement can occur on every basis of the quantum state, but typically this happens on the computational basis. In this specific case, the outcomes of the measurement are associated with the two projectors on the eigenspaces

$$P_0 = |0\rangle\langle 0| \qquad P_1 = |1\rangle\langle 1| \tag{7.15}$$

and the probabilities associated with the results of the measurement (0 and 1) are

$$p(0) = \langle\psi| P_0 |\psi\rangle = |\alpha_0|^2 \qquad p(1) = \langle\psi| P_1 |\psi\rangle = |\alpha_1|^2 \tag{7.16}$$

Once the measurement is performed, the quantum state collapses into the state associated to the measurement outcome, so that if the outcome of the measurement is 0 then the quantum state collapses into the $|0\rangle$ state. These kinds of measurements are called *projective measurements* on the computational basis.

7.2 Quantum Machine Learning

The combination of the main ingredients of ML, together with the fundamentals of the quantum computation just described, leads to the attempt to solve typical ML tasks using quantum circuits.

In a standard ML framework, the parameters of an algorithm are tuned through a training procedure in order to achieve a given task. In the QML context, the main ingredient is a Parametrized Quantum Circuit (PQC) [109]. PQCs (also known as variational circuits) are typically composed of a combination of fixed gates (like *CNOT*, *CZ* or *H* gates) and adjustable gates with tunable parameters (like the *RX*, *RY*, *RZ* gates) that can be used QML models. Even at low circuit complexity, PQC are able to generate non-trivial outputs that cannot be efficiently simulated by classical hardware. A PQC can be summarised in three steps, schematically shown in Figure 7.3:

1. **Pre-processing:** a feature vector x is sampled from the training data-set P_D and gets transformed following a classical pre-processing scheme by a function ϕ that maps $x \to \phi(x)$. This step can include *feature selection* and normalization techniques.

2. **PQC:** this is the core of the whole procedure. This stage of the process consists in two steps:

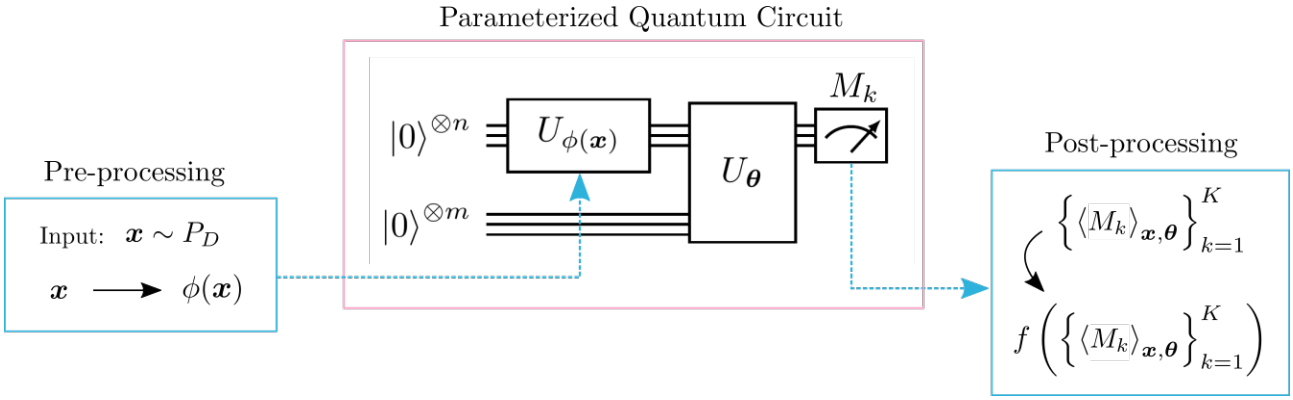


Figure 7.3: Main components of a supervised ML model based on a PQC. [109]

- data encoding: the pre-processed feature vector x is fed into the quantum circuit through an *encoding*, a circuit structure that assigns classical data to specific items of the n-qubit system. Typically this happens through an *encoder* circuit $U_{\phi(x)}$ which embeds the data into an n-qubits state. Different kinds of encodings exist, depending on the chosen circuit structure or any hardware limitations.
- variational circuit: the prepared n-qubit system is processed by a variational circuit U_θ with tunable parameters θ, which possibly act on an extended qubit register.

3. **Post-processing:** the measured expectation values $\{\langle M_k\rangle_{x,\theta}\}_{k=1}^{K}$ get mapped to a model prediction by a post-processing function $y = f(\{\langle M_k\rangle_{x,\theta}\}_{k=1}^{K})$

The tuning of the parameter of the variational circuit is obtained during the training procedure. In classical ML, this is typically obtained with a gradient descent procedure [94], which in a local optimization method that iteratively updates the tunable parameters θ of the loss function $L(\theta)$ towards the direction of the steepest descent, following the iterative rule

$$\theta^{i+1} = \theta^i - \eta \nabla L(\theta^i) \tag{7.17}$$

where η is a tunable parameter called learning rate and i is the iteration index. This is done similarly with PQCs, but this time it is possible to compute exactly the analytical gradients of PQCs using the parameter shift rule [110, 111].

7.3 Quantum Machine Learning for b-jet classification

In the context of this thesis work, a study of a QML application to jet identification has been performed. To quantify the applicability of QML algorithms, b-jet charge identification has been chosen as a task for this exercise. b-jet charge identification identifies the ability of separating jets produced by b and $\bar{b}$-quarks. This is a very well known task at LHCb, which leads to important physics measurements, such as the $b\bar{b}$ forward-central asymmetry. The standard LHCb algorithm to identify the charge of the b-quark generating the jet is an *exclusive* algorithm called *muon tagging*: the algorithm searches for a muon signature inside the jet with $p_{\mathrm{T}} > 5$ GeV; given that the muon is produced through the semi-leptonic decay of a b-quark, the muon charge is used to infer the b-quark charge. While being a very pure method, its performance is limited by the branching ratio of the semi-leptonic of the b-quark, which is approximately 10%, thus limiting the statistics available for this kind of measurement. Another approach suggests to use an *inclusive* algorithm, which instead of relying on a specific physics process happening inside the jet, it is based on the information coming from the whole jet substructure:

this is particularly relevant at LHCb, given that the lower pile-up environment allows having PID on the particles inside the jet: in this way, it is possible to identify almost all the particles that combine to form the jet. Given that the inclusive algorithm is based on more information, ML tools are suited to handle this kind of situation where a lot of inputs are present. A schematic view of the different approaches is shown in Figure 7.4.

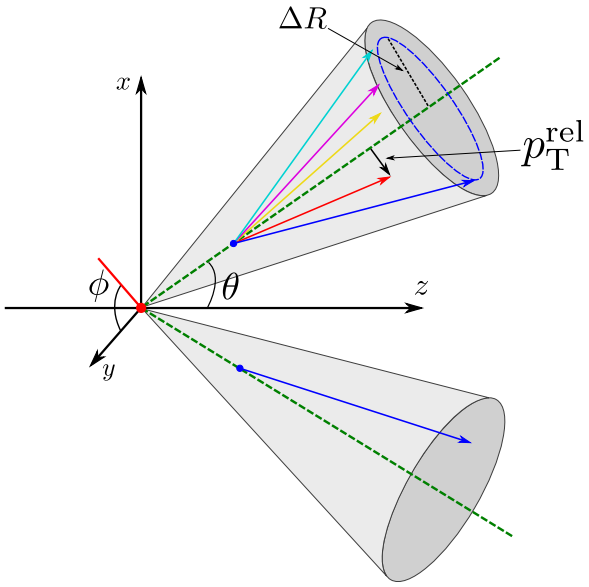


Figure 7.4: Schematic representation of possible jet tagging methods. In the exclusive method (lower jet) the information comes from a specific physics process, e.g. the muon, whose charge is correlated to the b hadron; in the inclusive method (upper jet), the information is extracted from the jet substructure [112].

7.3.1 Analysis strategy

The analysis strategy can be summarised in the following steps:

- features from the jet substructure are selected as input features for the QML algorithm and the DNN. The starting dataset is divided into training, validation, and testing datasets;
- both DNN and QML algorithms are trained on the training dataset and their performance is evaluated;
- DNN and QML algorithm are used to infer the charge of b-jets on the testing dataset. The performance is optimized to maximize the tagging power $\varepsilon_{\mathrm{tag}}$ defined in Equation 7.18.

The analysis steps are schematically represented in Figure 7.5.

7.3.2 Tagging power and dataset description

The performance of the b-jet charge identification is measured with a figure of merit called *tagging power* $\varepsilon_{\mathrm{tag}}$, defined as

$$\varepsilon_{\mathrm{tag}} = \varepsilon(1-2\omega)^2 \tag{7.18}$$

where ε is the efficiency and ω is the mistag, defined as

$$\varepsilon = \frac{\#\text{ identified jets}}{\#\text{ total jets}} \qquad \omega = \frac{\#\text{ wrongly identified jets}}{\#\text{ identified jets}} \tag{7.19}$$

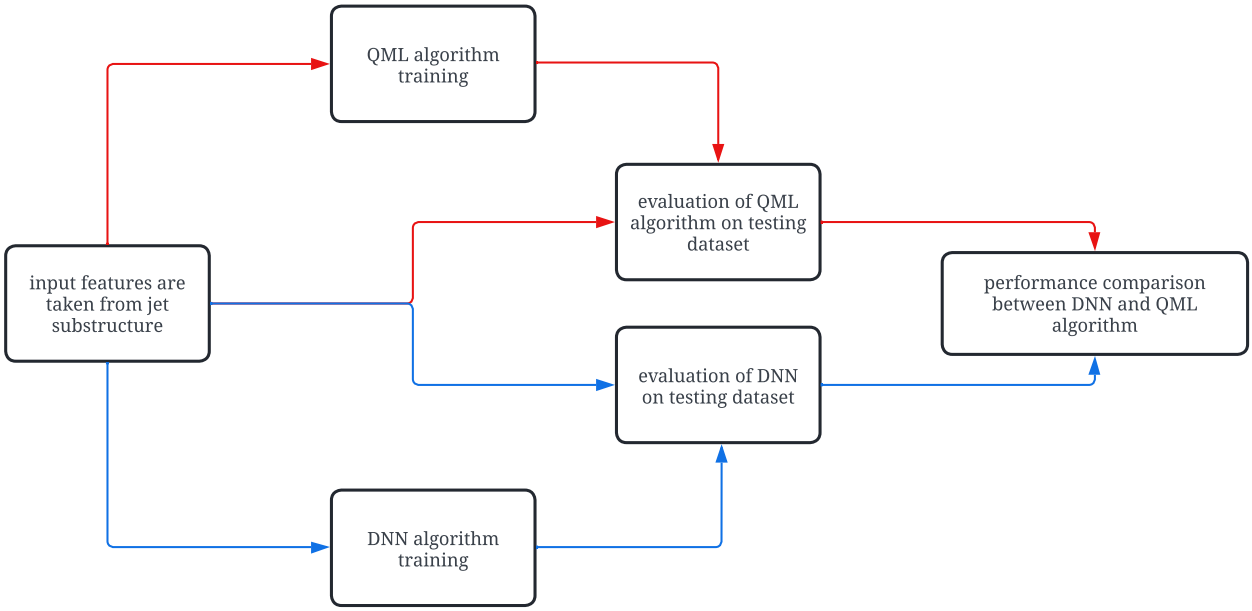


Figure 7.5: Flowchart of the analysis described in this Chapter. Red (blue) lines identify the steps where the QML algorithms (DNN) are used.

ω ranges from 0 (perfect classifier) to 0.5 (random classifier). In this way, the tagging power takes into account the fact that the algorithm might wrongly identify the jet charge.

In this study, LHCb simulated samples are used. The $b\bar{b}$ di-jets samples are produced within the LHCb simulation framework [70], which uses PYTHIA8.1 [29] with a specific LHCb configuration [73], to generate proton-proton interactions and jet fragmentation and hadronization at center-of-mass energy $\sqrt{s} = 13$ TeV. Pairs of b- and $\bar{b}$-jets are selected by requiring the jet transverse momentum $p_{\mathrm{T}} > 20$ GeV and jet pseudorapidity in the range $2.2 < \eta < 4.2$, to ensure that they are well inside the instrumented part of the detector. After the pre-selection, a fixed number of 16 different features related to the jet substructure are used as input to the classifiers. Among the reconstructed particles inside a jet the muon, kaon, pion, electron, and proton with the highest p_{T} are selected. For each particle three physical variables are considered: the magnitude of the transverse momentum to the jet axis ($p_{\mathrm{T}}^{\mathrm{rel}}$), the charge ($q$), and the distance, measured in the (η,ϕ) space, between the particle and the jet axis (ΔR). If a particle type is missing, the relative features are set to 0. The last feature is the weighted jet charge Q_{tot}, defined as the sum of the charges of the particles inside the jet weighted with the particles $p_{\mathrm{T}}^{\mathrm{rel}}$ [113–116]:

$$Q_{\mathrm{tot}} = \frac{\sum_i (p_{\mathrm{T}}^{\mathrm{rel}})_i q_i}{\sum_i (p_{\mathrm{T}}^{\mathrm{rel}})_i} \ . \tag{7.20}$$

The analysis is performed using two datasets. The *complete dataset* includes the events selected with the 16 features described above. The *muon dataset* contains jets with at least one muon and only four features: $p_{\mathrm{T}}^{\mathrm{rel}}$, ΔR, q of the muon and the weighted jet charge Q. Table 7.1 summarises the characteristics of the data samples.

7.3.3 Algorithm description

The QML algorithm used here is a VQC, already described in Section 7.2. Two types of data embeddings are considered:

Dataset	Muon			Kaon			Pion			Electron			Proton			
	$p_{\mathrm{T}}^{\mathrm{rel}}$	q	ΔR	$p_{\mathrm{T}}^{\mathrm{rel}}$	q	ΔR	$p_{\mathrm{T}}^{\mathrm{rel}}$	q	ΔR	$p_{\mathrm{T}}^{\mathrm{rel}}$	q	ΔR	$p_{\mathrm{T}}^{\mathrm{rel}}$	q	ΔR	Q_{tot}
Complete	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Muon	✓	✓	✓													✓

Table 7.1: Summary of the jet substructure features contained in the two datasets.

- *Angle Embedding*: given a n-qubit state, data are embedded in n angles of n rotational gates, with a one-to-one correspondence between the input feature and the qubit, therefore requiring n qubits for n input features.
- *Amplitude Embedding*: given a n-qubit state, data are embedded in the 2^n amplitudes x_i of the state $|x\rangle$

$$|x\rangle = \sum_{i=1}^{2^n} x_i \, |x_i\rangle \, . \tag{7.21}$$

  Therefore, given n input features, this circuit structure requires only $\log_2 n$ qubits.

After the data embedding, the variational part is considered. In this study, a full entanglement structure has been chosen, where all the qubits are entangled with each other. In the end, only the first qubit is measured: the Pauli operator $< \sigma_z > \in [-1, 1]$ is measured and its value is mapped to the probability for a jet to be generated by a b or $\bar{b}$-quark:

$$\begin{aligned} P_b &= \frac{1}{2}(< \sigma_z > +1) \\ P_{\bar{b}} &= 1 - P_b \end{aligned} \tag{7.22}$$

A scheme of the two circuit structures considered is given in Figure 7.6.

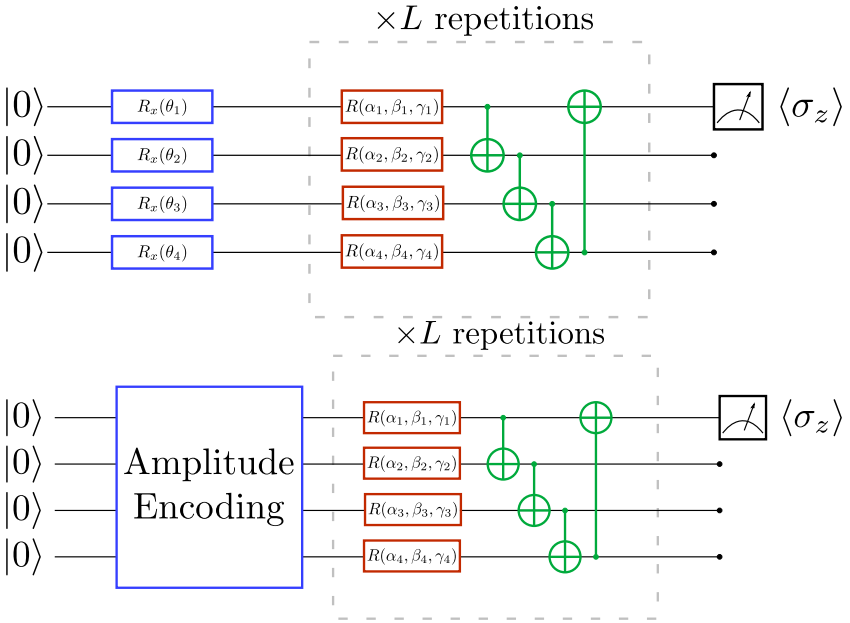


Figure 7.6: Schematic representations of the two circuit structures used in this study: Angle Embedding (upper circuit) and Amplitude Embedding (lower circuit). Both embeddings are followed by L repetitions of a variational circuit [112].

The classical optimization procedure of the QML algorithm is based on a mini-batch gradient descent algorithm using the ADAM [95] optimizer. The loss function to be minimized during the training procedure is the Mean Squared Error, defined as

$$\mathcal{L}(\{\theta\}) = \frac{1}{N}\sum_{i=1}^{N}(P_b^i(\{\theta\}) - T^i)^2, \tag{7.23}$$

where N is the number of training jets, $\{\theta\}$ defines the set of optimized parameters of the circuits, $P_b^i(\{\theta\})$ is the probability for the i-jet to be generated by a b-quark (as defined in Equation 7.22) and T^i is the true charge of the b-quark.

Quantum circuits are simulated by means of noiseless simulators (noise impact is studied in Section 7.3.7) using Pennylane [117], a Python framework designed specifically for QML applications. The quantum circuit is embedded into a classical optimization algorithm, using the Jax [**ref:jax**] Python library. Since the idea is to compare the quantum algorithms results and the results obtained with a classical DNN, the same analysis is performed with a standard feed-forward DNN, implemented using the Keras [118] framework with the TensorFlow [119] back-end. The DNN layout is represented in Figure 7.7.

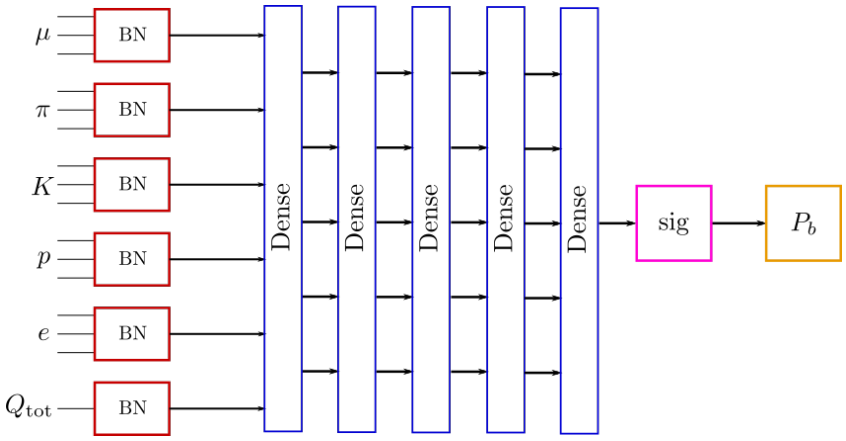


Figure 7.7: Schematic representation of the DNN layout used as a comparison with the QML algorithms [112].

7.3.4 Training and testing phases

The *muon* and *complete* datasets are both split into training and testing sub-datasets: about 60% of the samples are used in the training process that includes also the validation and the remaining 40% are used to test, evaluate and compare the classifiers. In the *muon dataset* analysis, 60000 jets are used for training and 40000 jets are used for testing. The *complete dataset* training is performed on 400000 jets and the remaining 290000 are used for testing and assessing performance. Figure 7.8 shows the probability distribution P_b for the complete dataset. Two results come out of Figure 7.8:

- the P_b distributions for b- and $\bar{b}$-jets are separated, showing that the QML algorithm is able to perform some discrimination;
- jets that are wrongly classified tend to accumulate close to $P_b = 0.5$, while correctly classified jets tend to stay at the edges of the distributions. This is going to be relevant in Section 7.3.5 when discussing the tagging power optimization.

The QML algorithm is compared to the DNN by looking at the ROC curve, as shown in Figure 7.9 for the muon and the complete dataset:

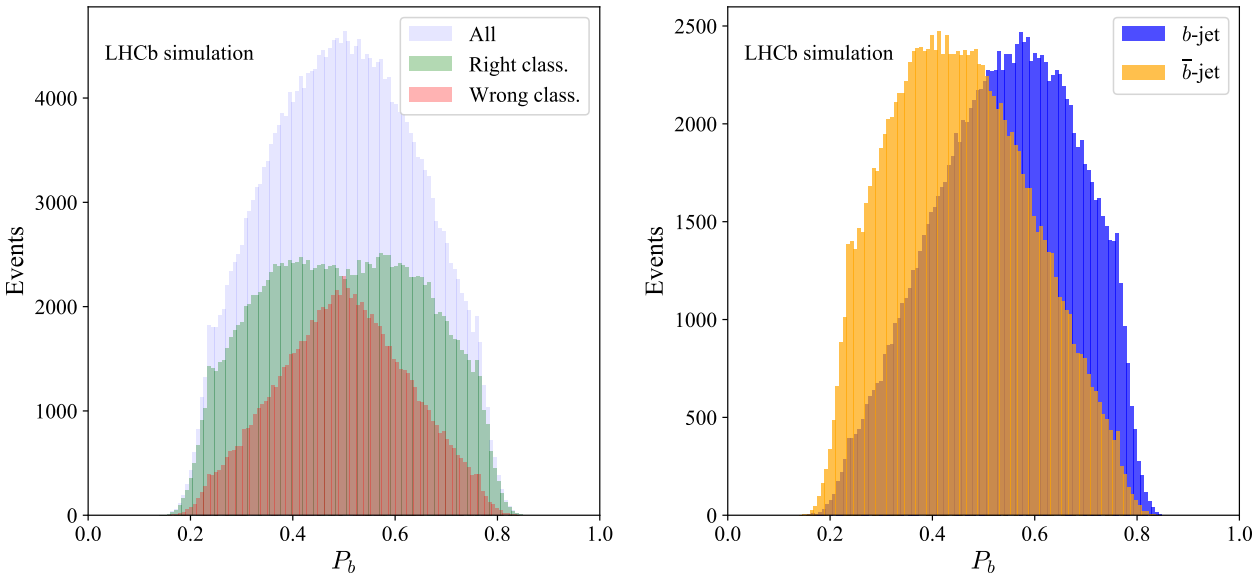


Figure 7.8: (Left) Probability distribution for the complete dataset evaluated with the Angle Embedding structure. Red (green) distributions show the wrongly (correctly) tagged jets. (Right) Probability distributions for jet tagged to (blue) b and (yellow) $\bar{b}$ quarks [112].

- the Angle Embedding structure performs almost as good as the DNN, particularly for the muon dataset;
- for the muon dataset, the Amplitude Embedding does not perform as good as the Amplitude Embedding, while the gap between the two approaches seems to reduce in the complete dataset. A possible explanation for this behavior is that when classifying the muon dataset, the Amplitude Embedding is evaluated on a 2-qubit circuit, whose structure might not be enough complex to perform classification.

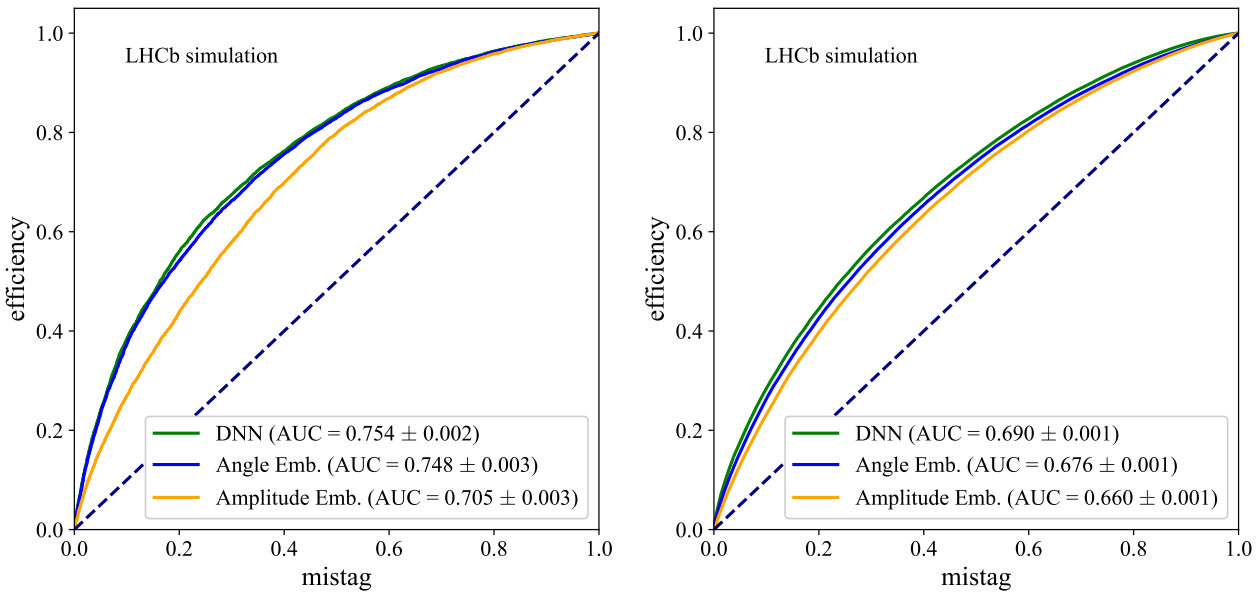


Figure 7.9: ROC curves for the muon (left) and the complete (right) dataset. QML algorithms are compared to the classical DNN [112].

7.3.5 Results and algorithm performance

Following a physics-driven approach, the performance of the QML algorithm and the DNN is evaluated using the jet tagging power, defined in Eq. 7.18. The tagging power is computed as a function of the jet p_{T} and η for both the quantum and the classical classifiers. In order to optimize the tagging power, a region symmetric with respect to 0.5 is defined, where no classification is performed, as already demonstrated in Figure 7.8. The width Δ_{cut} of the excluded region is defined for each classifier by maximizing the tagging power evaluated using all the jets in the testing dataset. Such an exclusion region reduces the tagging efficiency because fewer jets are tagged, but enhances the identification probability by reducing the mistag ω. The probability distributions and the excluded region are shown in Figure 7.10: indeed the region where the prediction power is minimum is excluded. The width Δ_{cut} of the excluded region for each classifier and for *muon* and *complete dataset* are summarised in Table 7.2.

	Classifier		
Dataset	DNN	Angle Embedding	Amplitude Embedding
Muon	0.30	0.25	0.16
Complete	0.21	0.19	0.12

Table 7.2: Width Δ_{cut} for different classifiers and dataset [112].

7.3.6 Dependence of the results on the number of training events and circuit depth

To further understand the performance of the quantum circuits studied, the dependence with respect to the number of training events and the circuit complexity are evaluated: this is necessary in order to have a concrete understanding of the feasibility of near-term applications on quantum hardware. These parameters have an impact on the execution times and therefore on the possibility to use it. The performance dependence on the number of training samples is an interesting parameter to compare QML and DNN methods, in order to assess the differences between the two approaches. Given the high computational efforts of simulating complex circuits with several qubits, only the *muon dataset* is used. For QML, the Angle Embedding structure is considered with a different number of strongly entangled layers and a different number of training events. The results are compared with the same DNN considered in the previous section. The metric used to quantify the goodness of the quantum classifier is the accuracy on a test subset of 40000 jets. The performance is calculated by averaging over 10 training rounds. In Figure 7.13 (a) the accuracy of the Angle Embedding circuit is shown as a function of the number of layers of the circuit. As expected, by increasing the depth of the circuit, and therefore its complexity, the accuracy increases. This behavior stops at around 5 layers, where the accuracy is saturating and no further improvement is evident. It is clear that, for a given number of features and training data, the Angle Embedding model does not profit from an arbitrarily large number of layers, therefore it is possible to keep a low number of layers, and subsequently a lower complexity of the circuit, to obtain the best performance. This would reduce also the computing time and resources needed for the simulation.

The accuracy as a function of the number of training events for the Angle Embedding circuit and the DNN is shown in Figure 7.13 (b). Increasing the number of training events the performance of the quantum algorithm is similar to the DNN, but when the number of training events decreases the quantum algorithm keeps a very high performance, while the DNN is not able to perform a good classification. This means that, with respect to the DNN, the QML method reaches optimal performance with a lower number of events. Considering the fact that state-of-the-art ML algorithms require very large data sets to get meaningful performance, this unique feature of QML algorithms

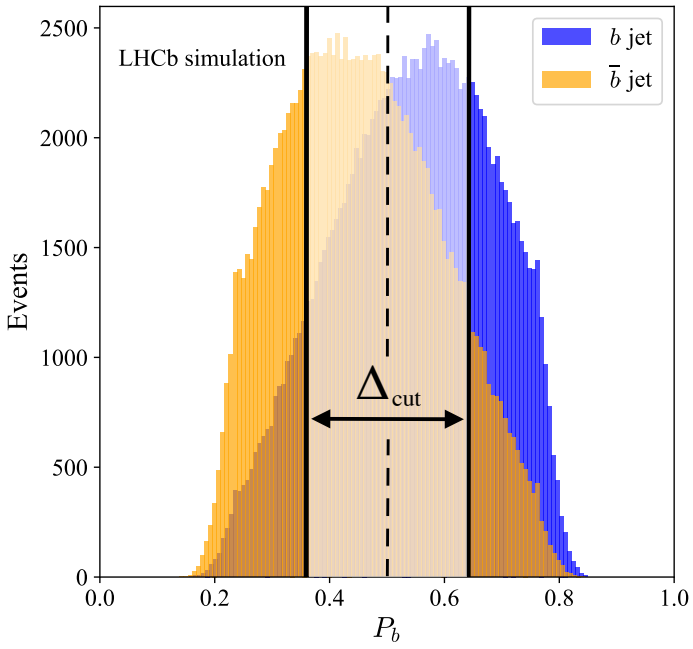


Figure 7.10: Probability distributions for jet tagged to (blue) b and (yellow) $\bar{b}$ quarks. The Δ_{cut} region to optimize the tagging power is shown [112].

needs further investigation, which could lead to a better understanding of how these algorithms are using the input features.

7.3.7 Noise contribution to circuit performance

All results presented so far have been obtained in a "noiseless" scenario, which means that the interactions between the circuit and the environment surrounding it have not been considered. Instead, in a real scenario, *i.e.* when dealing with real quantum hardware, it is important to understand the impact of noise on quantum circuits. Two kinds of noise can affect quantum algorithms:

- coherent noise: it originates from unitary errors in the application of quantum gates. This lead to the construction of a different quantum state with respect to the desired one. A typical source of this kind of noise is non-ideal calibrations of the quantum hardware;
- incoherent noise: this noise results from the interaction between the quantum hardware and the environment. This noise gives quantum states that are not pure anymore and are described by mixed states, *i.e.* probability distributions over different states.

While coherent noise can be easily described by unitary matrices combined with the circuit, incoherent noise needs the language of density matrix and mixed states. Indeed, incoherent noise is modeled by

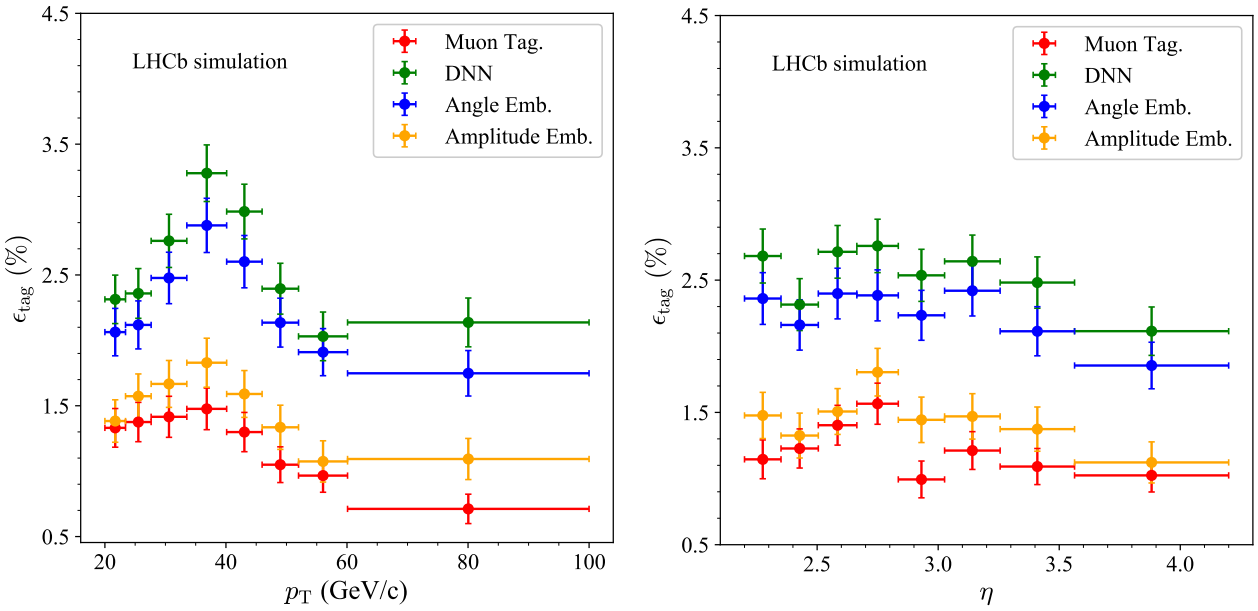


Figure 7.11: Tagging power as a function of jet p_{T} (left) and jet η (right) for the muon dataset. QML algorithms are compared to DNN and the muon tagging approach.

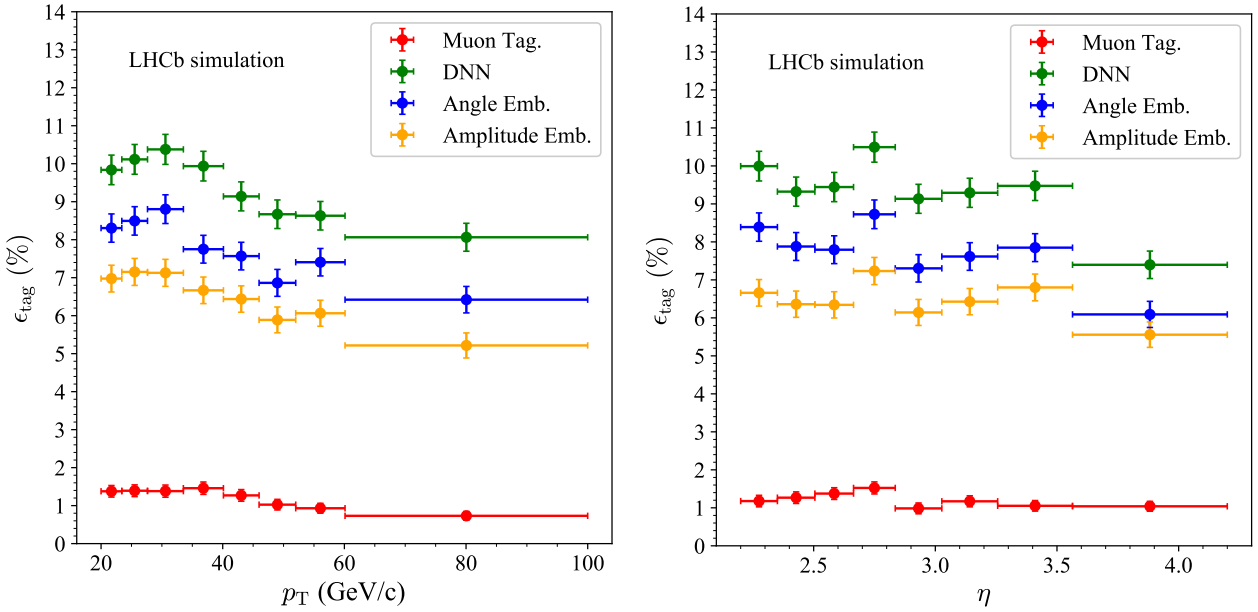
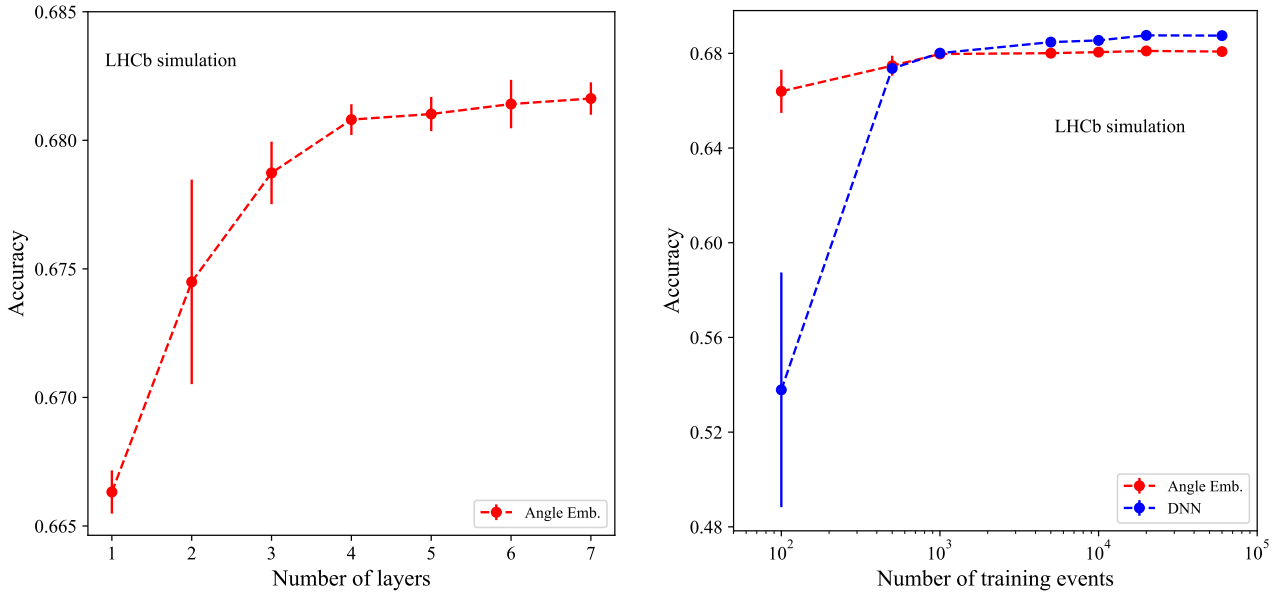


Figure 7.12: Tagging power as a function of jet p_{T} (left) and jet η (right) for the complete dataset. QML algorithms are compared to DNN and the muon tagging approach [112].

quantum channels. Mathematically, a quantum channel is a linear, completely positive, and trace-preserving (CPTP) map. Given an initial state ρ, the output state $\Phi(\rho)$ after applying the channel Φ to the state ρ can be described as

$$\Phi(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{7.24}$$

where K_i are called Kraus operators, satisfying the condition $\sum_i K_i K_i^\dagger = I$, where I is the identity matrix. Therefore, the action of a quantum channel (and therefore the contribution of the environment



(a) (b)

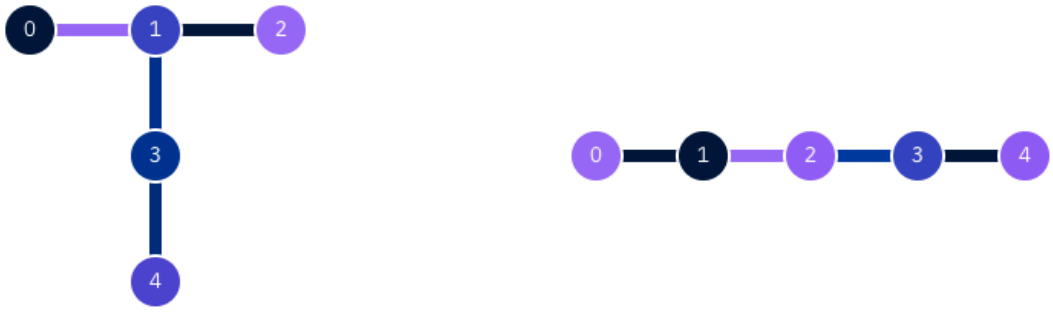
Figure 7.13: (a) Accuracy of the Angle Embedding structure on the *muon dataset* versus the number of layers. (b) Accuracy of the (red) Angle Embedding structure and (blue) DNN on the *muon dataset* versus the number of training events [112].

to the circuit performance) can be represented as an application of a transformation corresponding to the Kraus operator K_i with a given probability; in particular, the quantum channel applies the transformation $\frac{1}{p_i} K_i \rho K_i^\dagger$ with a probability $p_i = \mathrm{Tr}[K_i \rho K_i^\dagger]$. Quantum channels, therefore, represent a probability distribution over different possible transformations on a quantum state.

For this specific study, the simulations of noise contribution taking into account both sources of noise in quantum circuit measurements have been performed using the *pennylane-qiskit* plugin [117, 120]. This plugin allows the simulation of noise models coming from different real IBM quantum computers [121], including state preparation and readout errors and keeping the Pennylane syntax. The result is a simulation of a quantum algorithm on a real device structure. It is therefore possible to study the performance of the considered algorithms on real hardware, without the need to deal with all the technical difficulties arising when using a real quantum computer. Four IBM quantum computers are considered: `ibmq-belem`, `ibmq-santiago`, `ibmq-jakarta` and `ibmq-toronto`, which have different numbers of qubits (respectively 5, 5, 7 and 27 qubits), different quantum volumes[1] (respectively 16, 32, 16, and 32) and different qubits structure, as shown in Figure 7.14 for the `ibmq-santiago` and `ibmq-belem` which have the same number of qubits.

Studies are performed on the Angle Embedding circuit structure with three strongly entangled layers. A small subset of the *muon dataset* is used because simulating circuits including noise contribution is more time and computationally consuming; on the other hand, with a low number of events the quantum algorithm performance is sufficiently high, as shown in Figure7.13. In this way, a subset of 1000 jets of the *muon dataset* is selected for training while validation is performed on a subset of 10000 jets. For each noise model, the training is performed for 50 epochs using ADAM [95] with a learning rate $\xi = 0.01$ and batch size of 10 jets. The results are averaged over five rounds of training, using five independent training subsets. The relevant figure of merit to assess noise models performance is the

[1]The quantum volume is the maximum size of a quantum circuit that can be effectively implemented on a noisy intermediate-scale quantum device. In this paper, the definition from Ref. [**ref:quantumvolume**] is adopted.



(a) (b)

Figure 7.14: Qubit structure of `ibmq-belem` (a) and `ibmq-santiago` (b) [112].

accuracy on the validation test. The results are shown in Figure 7.15 and summarised in Table 7.3. Models including noise need more epochs to reach convergence, but in the end the results are consistent with those of noiseless simulations within error. Such a result demonstrates that the proposed circuit model for the *muon dataset* is robust to noise.

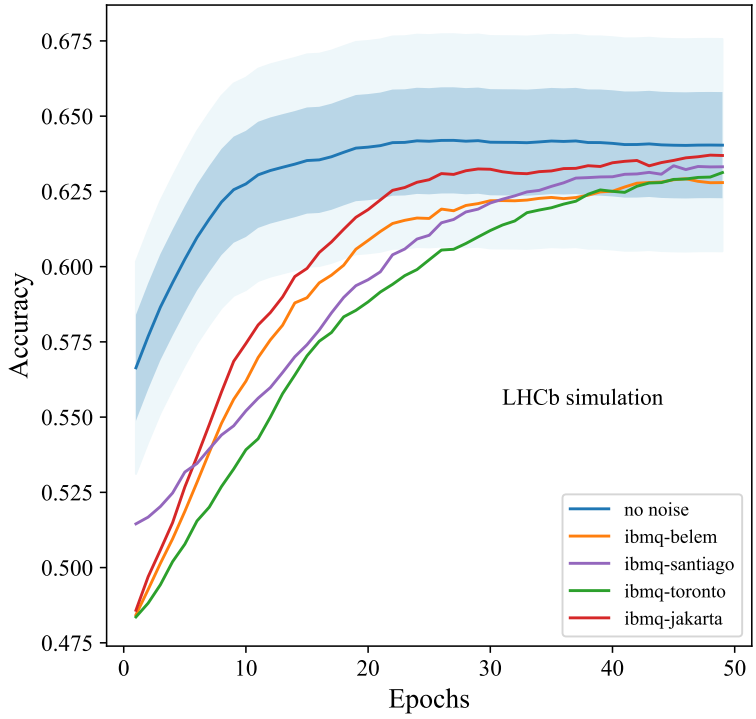


Figure 7.15: Validation accuracy for noise models as a function of the number of epochs. Blue (light-blue) band represent 1σ (2σ) uncertainty bounds for the noiseless model [112].

7.4 Future prospects

The work described in this Chapter has been proposed as an exploratory approach of QML algorithms applied to a real physics problem, possibly using a physics-driven approach suitable at the analysis level. The main limitations of this approach come from the application of these tools using real

Noise model	accuracy
no noise	0.640 ± 0.017
`ibmq-belem`	0.629 ± 0.047
`ibmq-santiago`	0.633 ± 0.038
`ibmq-jakarta`	0.637 ± 0.042
`ibmq-toronto`	0.631 ± 0.044

Table 7.3: Accuracy for noisy circuits obtained with the *muon dataset* [112].

hardware:

- for this specific study, the performance seems to improve by using circuit structures with more qubits. So far, quantum computers have limited numbers of qubits, possibly limiting the performance of QML algorithms. On the other hand, the performance of state-of-the-art classifiers such as DNN is seen to have reached their maximum performance, therefore an effort in using more and more qubits has to be pursued in order to fully exploit the potential of QML algorithms;
- while Figure 7.15 shows that the 4-qubit circuit structure is robust to noise, and therefore to applications on real quantum hardware, this is definitely not the case for more complex circuit structures, where it is harder to mitigate the interactions between the quantum hardware and the environment. This results in higher errors during the circuit evaluation, which degrade the performance of the classifier. To this extent, an effort towards error mitigation techniques and optimized circuit design must be taken;
- finally, these QML algorithms can be applied to a variety of tasks. Chapter 5 proved that better classification performance between WW and $W+$ jets will definitely result in more precise measurement. Therefore, QML algorithms might help in improving the classification performance and consequently the precision of several physics measurements.

Conclusions

In this thesis, the measurements of W boson production in association with jets and $W(\to \mu\nu)W(\to$ jets) pair production performed at the LHCb experiment are presented. Both analyses are based on data taken by the LHCb experiment during the Run 2 data-taking campaign, for a total integrated luminosity of $\mathcal{L} = 5.4$ fb^{-1}.

In the $W+$ jets measurement a high p_{T} muon together with at least one or two jets with $p_{\mathrm{T}} > 20$ GeV are required. The cross section defined in the fiducial region determined by kinematic requirements on the muon and the jets, is measured differentially as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet} and η^{μ}, independently of the W charge and for the two different charges of the muon. The values of the cross section in each bin of η and p_{T} will be made available for the inclusion in the next round of the PDFs determination. Results for the total cross section in the fiducial region are

$$\begin{aligned}
\sigma(W + 1 \text{ jet}) &= 462.29 \pm 0.48 \pm 20.34 \pm 9.24 \text{ pb} \\
\sigma(W^+ + 1 \text{ jet}) &= 287.73 \pm 0.37 \pm 12.66 \pm 5.75 \text{ pb} \\
\sigma(W^- + 1 \text{ jet}) &= 85.87 \pm 0.30 \pm 8.17 \pm 3.71 \text{ pb} \\
\sigma(W + 2 \text{ jets}) &= 47.69 \pm 0.25 \pm 2.95 \pm 0.95 \text{ pb} \\
\sigma(W^+ + 2 \text{ jets}) &= 29.87 \pm 0.21 \pm 1.85 \pm 0.59 \text{ pb} \\
\sigma(W^- + 2 \text{ jets}) &= 17.88 \pm 0.16 \pm 1.14 \pm 0.35 \text{ pb}
\end{aligned} \tag{7.25}$$

where the first uncertainty is statistical, the second one is systematic, and the third one is related to luminosity determination. These results, obtained for the first time at a center-of-mass energy $\sqrt{s} = 13$ TeV in the forward region, are in agreement with the SM expectations.

The WW pair production measurement requires one W boson to decay semileptonically into a muon and a neutrino, and the other one to decay hadronically in two jets. The same fiducial region used in the $W+$ jets analysis is used as well as the same requirements for the physical objects reconstruction. In this case the $W+$ jets events constitute an irreducible background of the searched signal. To increase the separation between WW and $W+$ jets events a ML techniques is exploited. A DNN has been optimised for the classification tasks, with input features coming from the kinematic observables of the muon and the jets. The result for the $W(\to \mu\nu)W(\text{jets})$ cross section in the fiducial region is

$$\sigma\left(W(\to \mu\nu)W(\to \text{jets})\right) = 0.64 \pm 0.30 \pm 0.05 \pm 0.01 \text{ pb} \tag{7.26}$$

where the first uncertainty is statistical, the second one is systematic, and the third one is related to luminosity determination. The measurement is statistically limited, but it could achieve much higher precision in the future with the next LHCb data-taking campaigns.

This is the first time that di-bosons production is studied in the forward region in proton-proton collisions at $\sqrt{s} = 13$ TeV. Given the limited precision, it is difficult to go more deeply in the comparison with the theory but it opens a new field of study for LHCb.

LHCb would greatly improve its potentials on electroweak measurements if the electron could be included. The performance of electron reconstruction for the upgraded ECAL has been studied and presented using $Z \to ee$ events. Different luminosities and pile-up conditions have been considered. A Hybrid-MC simulation framework has been developed to study different ECAL geometry configurations, in order to assess the quality of the Z invariant mass peak reconstruction. This study shows that requiring an ADC saturation threshold equal to 40 GeV will allow a better reconstruction of high p_{T} electrons. Secondly, the impact of the proposed HCAL removal on jets reconstruction has been studied: results show that this will involve a 10% degradation in the jets reconstruction performance, especially for the jet energy resolution and the jet reconstruction efficiency. More studies are needed to understand if this worsening in performance can be mitigated or recovered at all.

An other important improvement could arise from the implementation of Quantum Machine Learning algorithms on jet identification. In this thesis, the study of jets originated by b- and $\bar{b}$-quark is taken as a proof-of-principle to test the performance of these algorithms and to investigate if new insights are possible in the jet sub-structure. While no improvements have been found with respect to state-of-the-art ML techniques for the moment with the available quantum computing resources, this study sets the first stepping stone in understanding the full potential of these new tools in future experiments.

Appendix A

Reweighting procedure templates

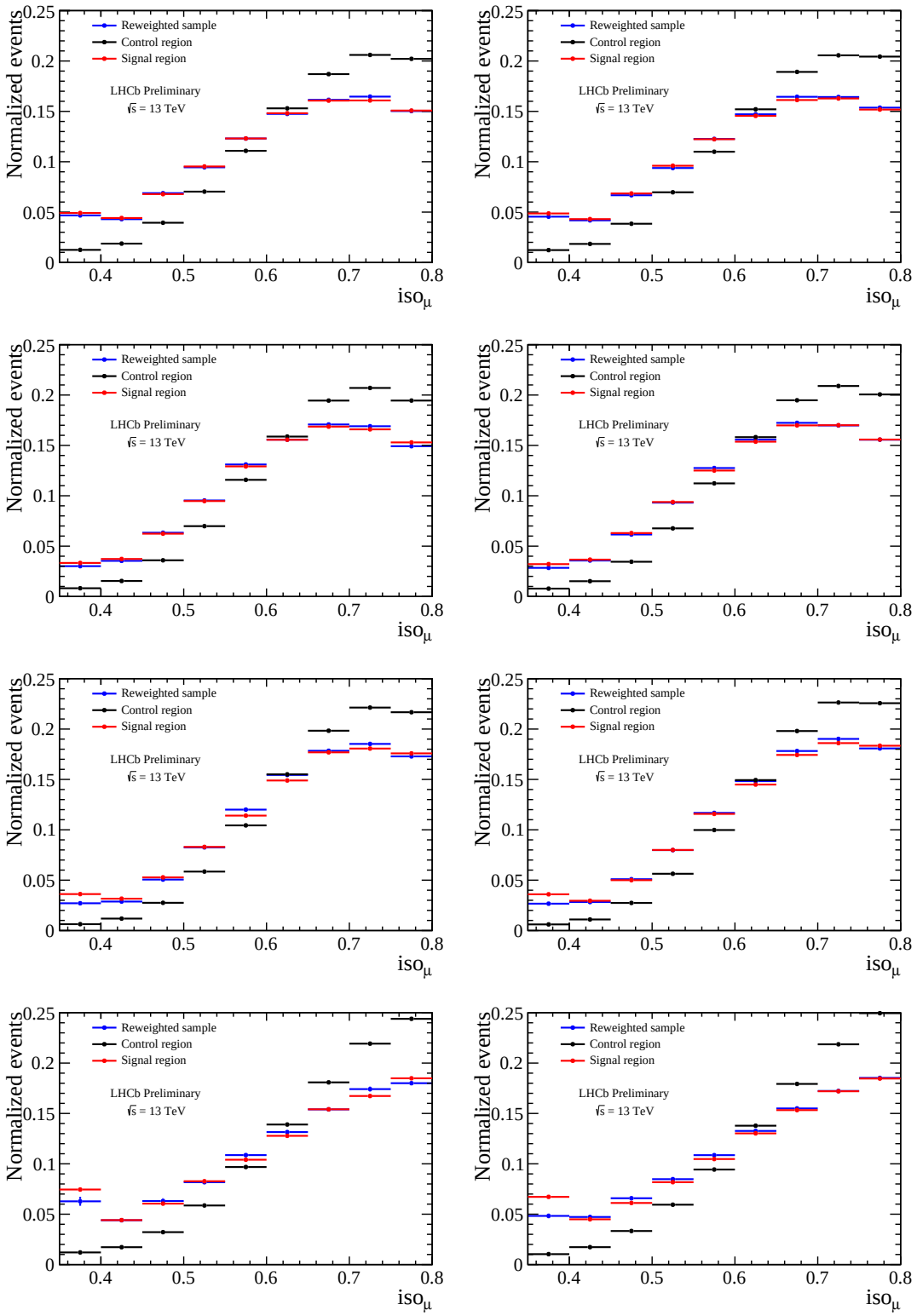


Figure A.1: I_μ distributions for SR (red dots), CR (black dots) and the reweighted distribution (blue dots) for $I_\mu < 0.8$ as a function of η^μ. Distributions are normalized to one and show the evident matching between SR and reweighted distributions.

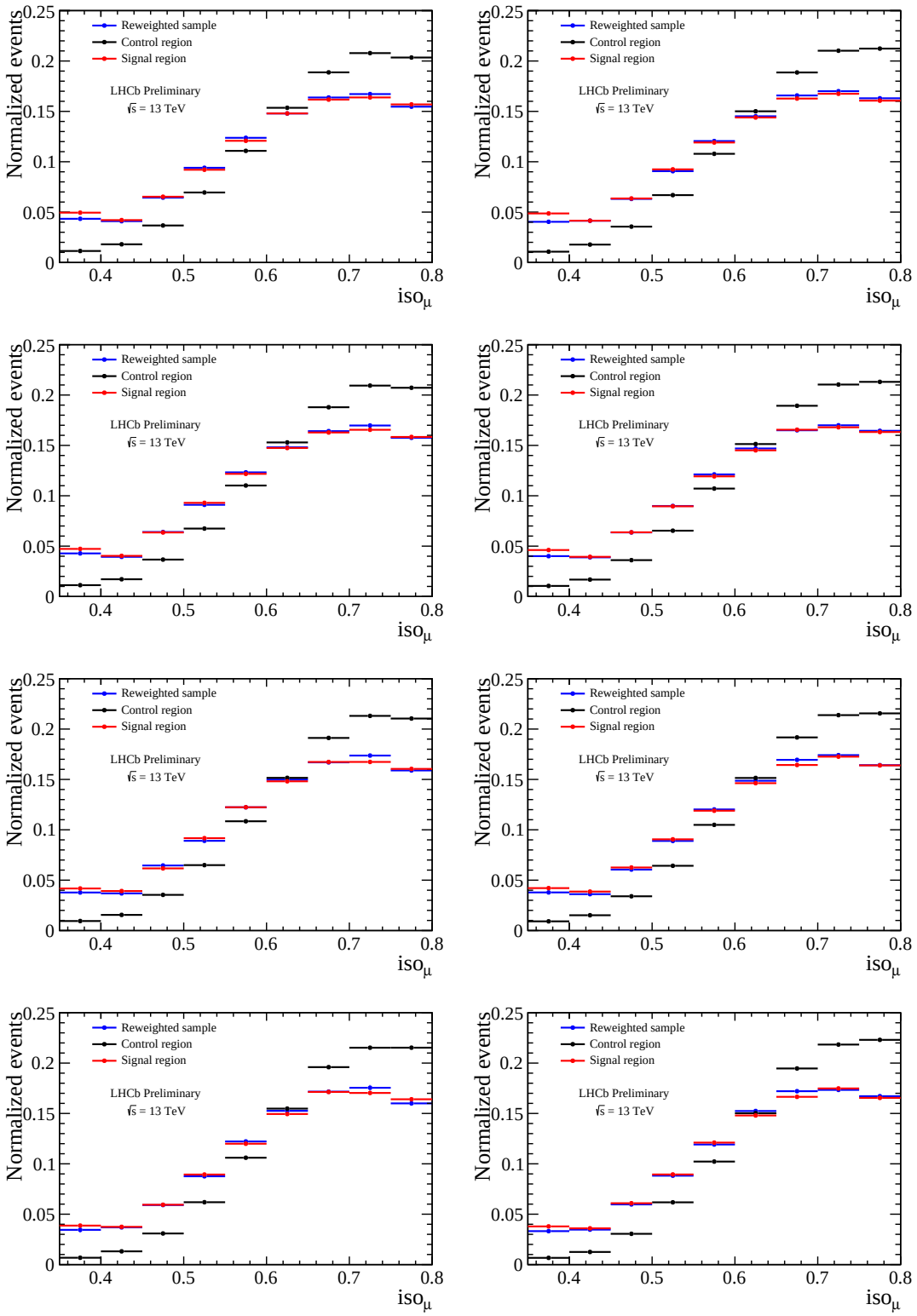


Figure A.2: I_μ distributions for SR (red dots), CR (black dots) and the reweighted distribution (blue dots) for $I_\mu < 0.8$ as a function of η^{jet}. Distributions are normalized to one and show the evident matching between SR and reweighted distributions.

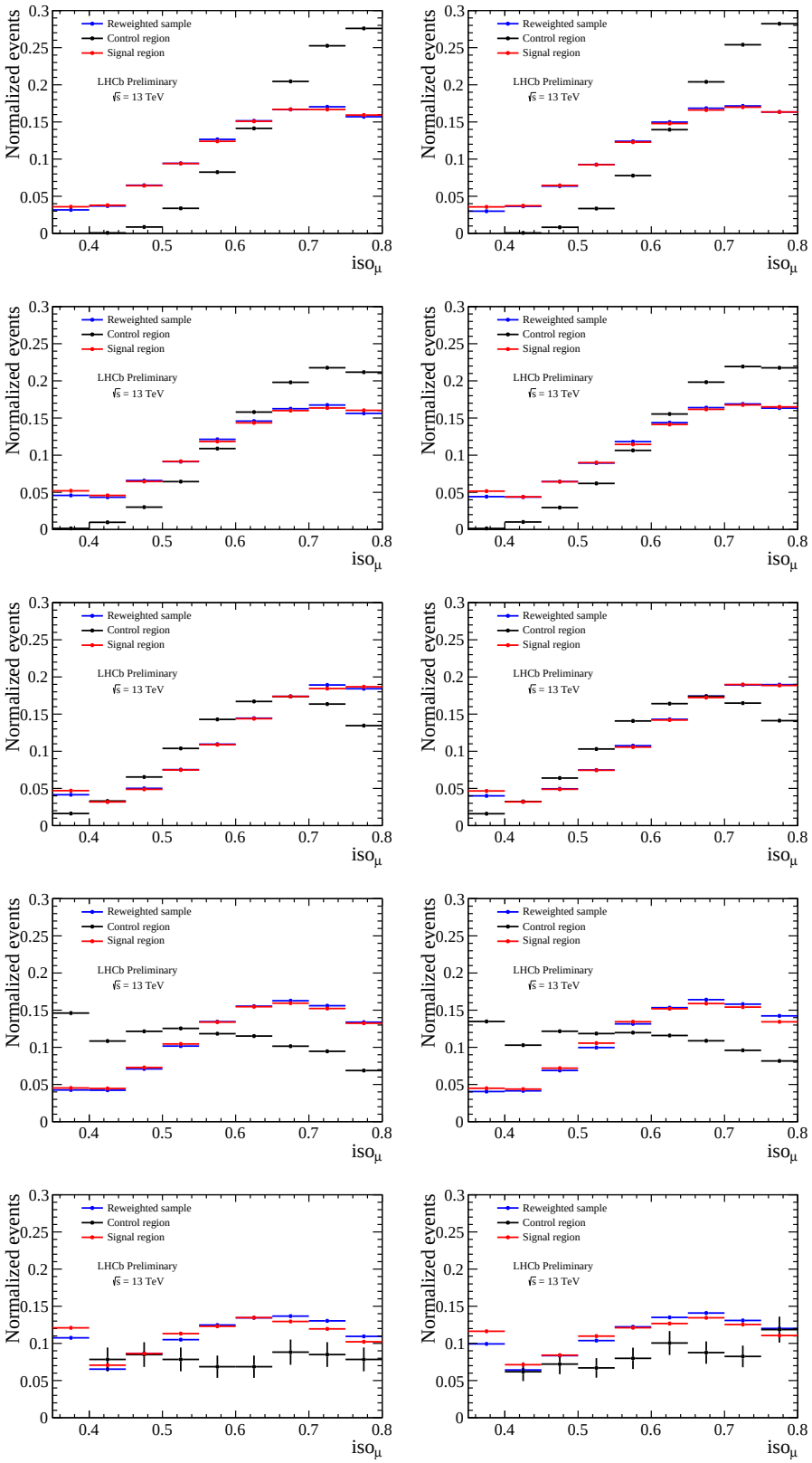


Figure A.3: I_μ distributions for SR (red dots), CR (black dots) and the reweighted distribution (blue dots) for $I_\mu < 0.8$ as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$. Distributions are normalized to one and show the evident matching between SR and reweighted distributions.

Appendix B

Template fits for different bins

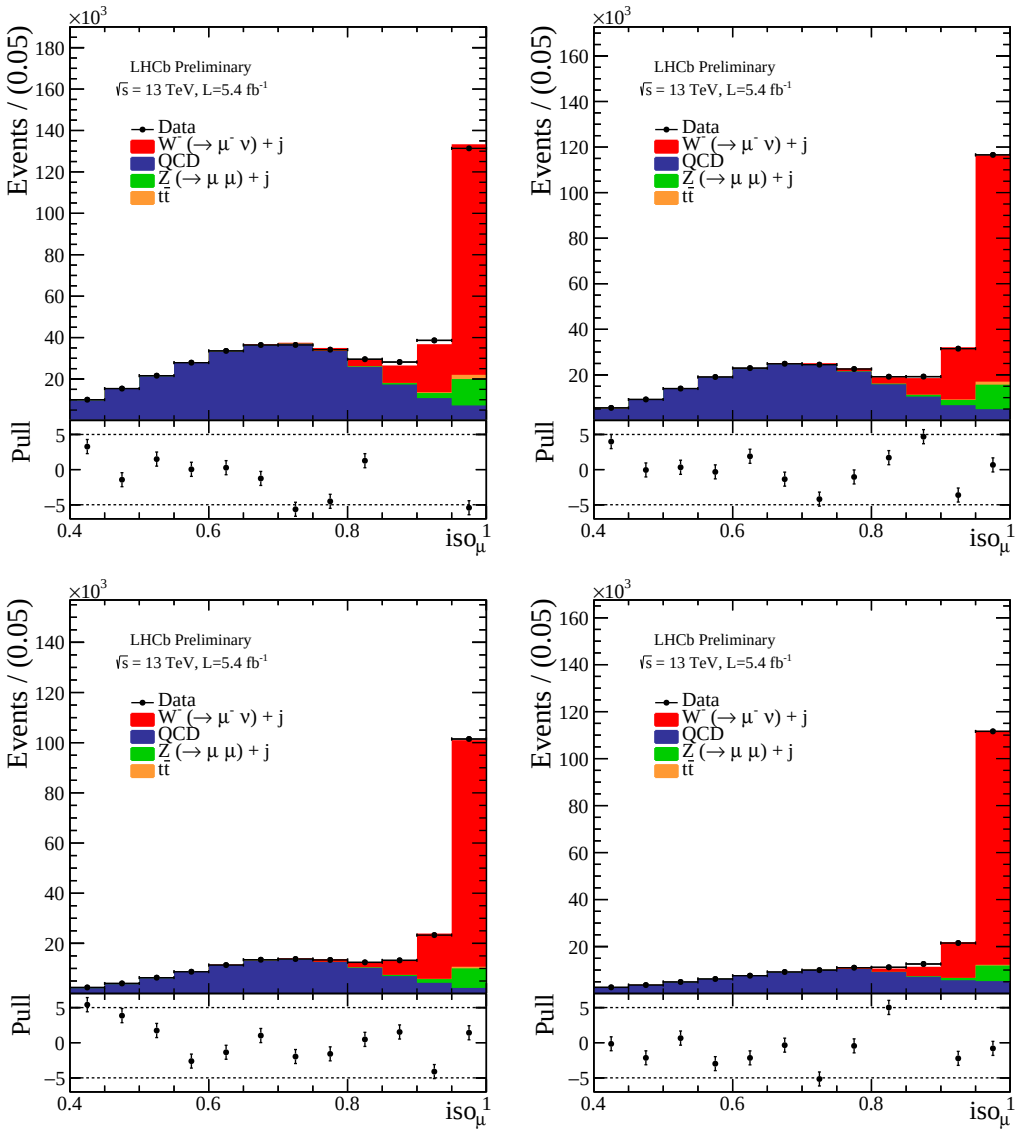


Figure B.1: I_μ template fit as a function of η^μ for W^-+ 1 jet selection.

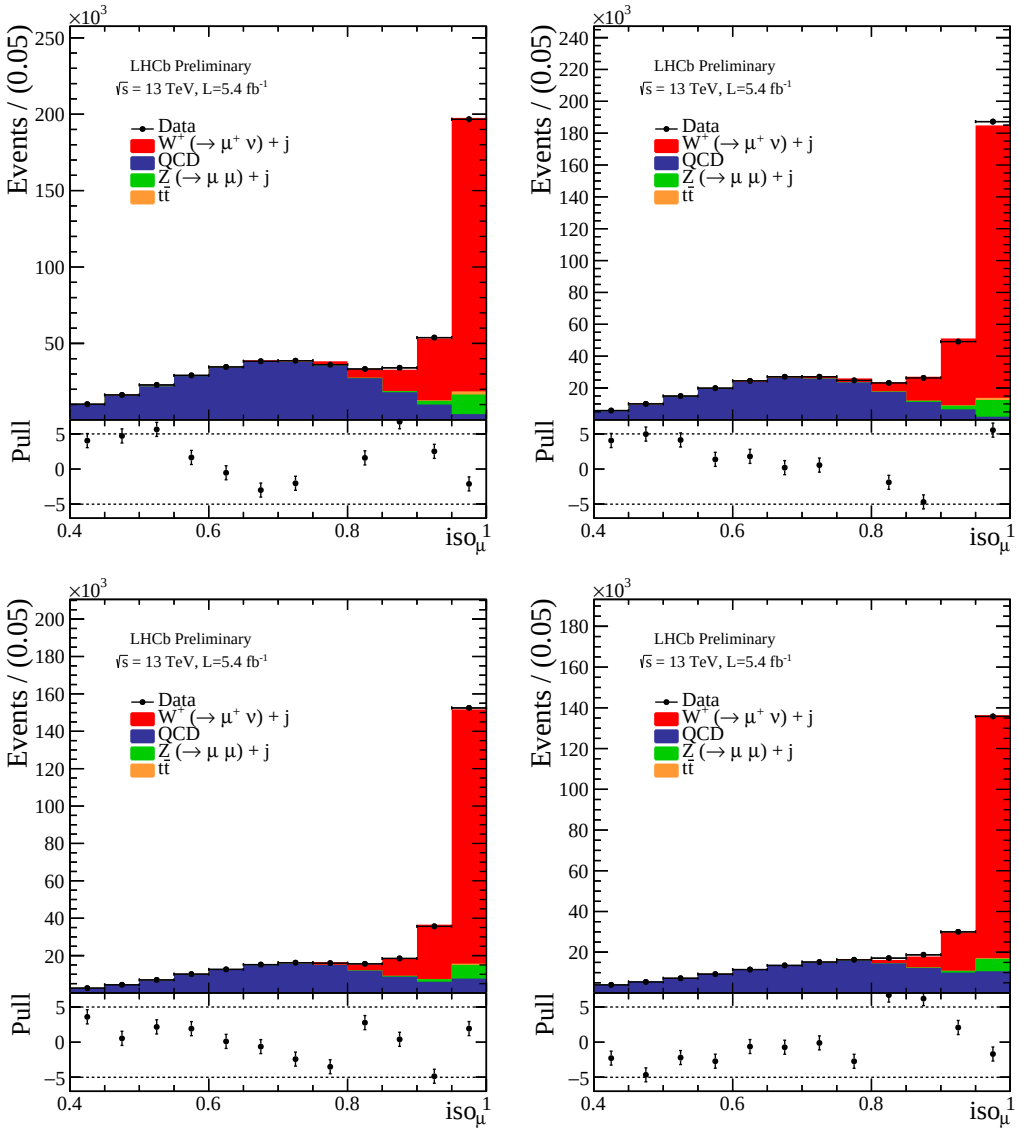


Figure B.2: I_μ template fit as a function of η^μ for W^++ 1 jet selection.

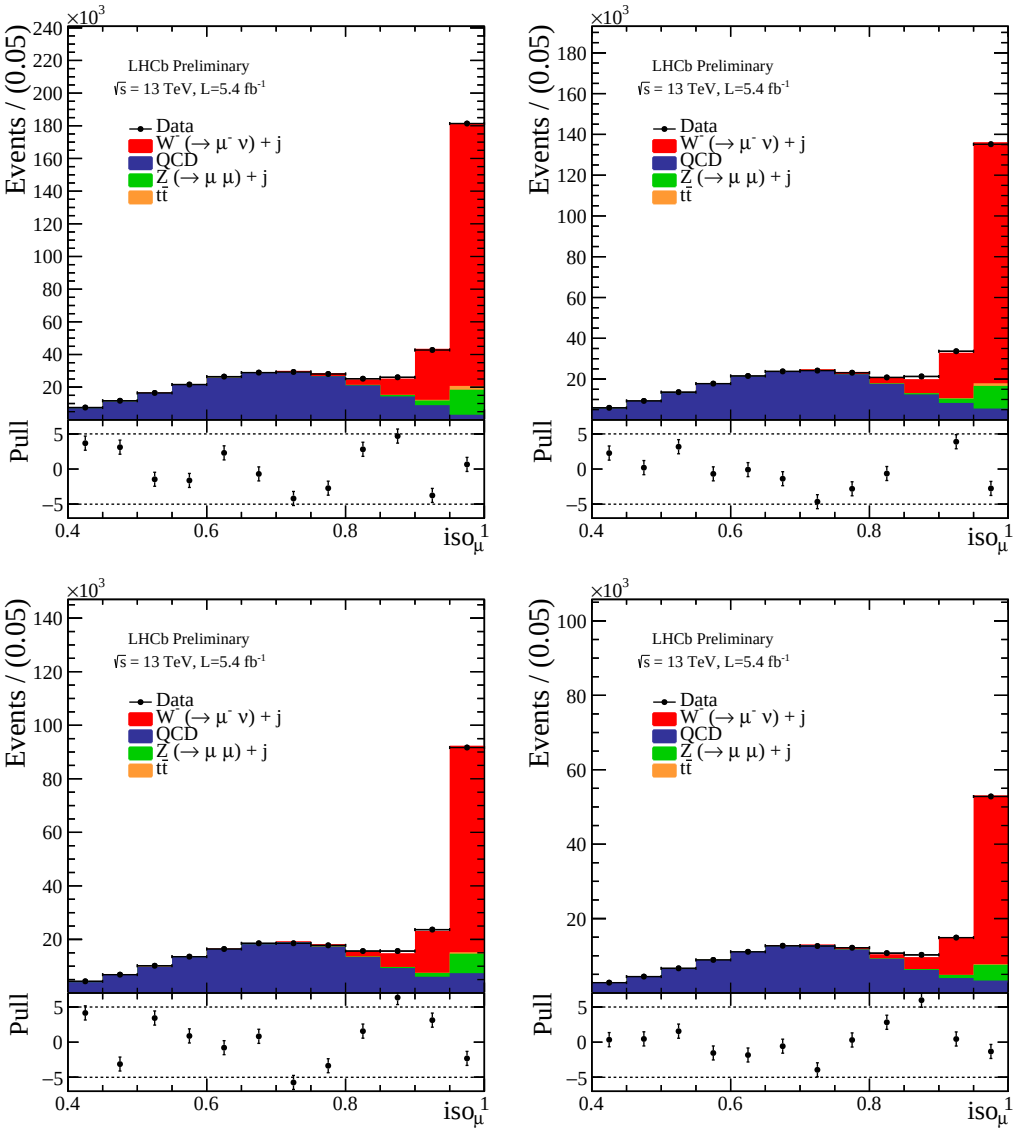


Figure B.3: I_μ template fit as a function of η^{jet} for W^-+ 1 jet selection.

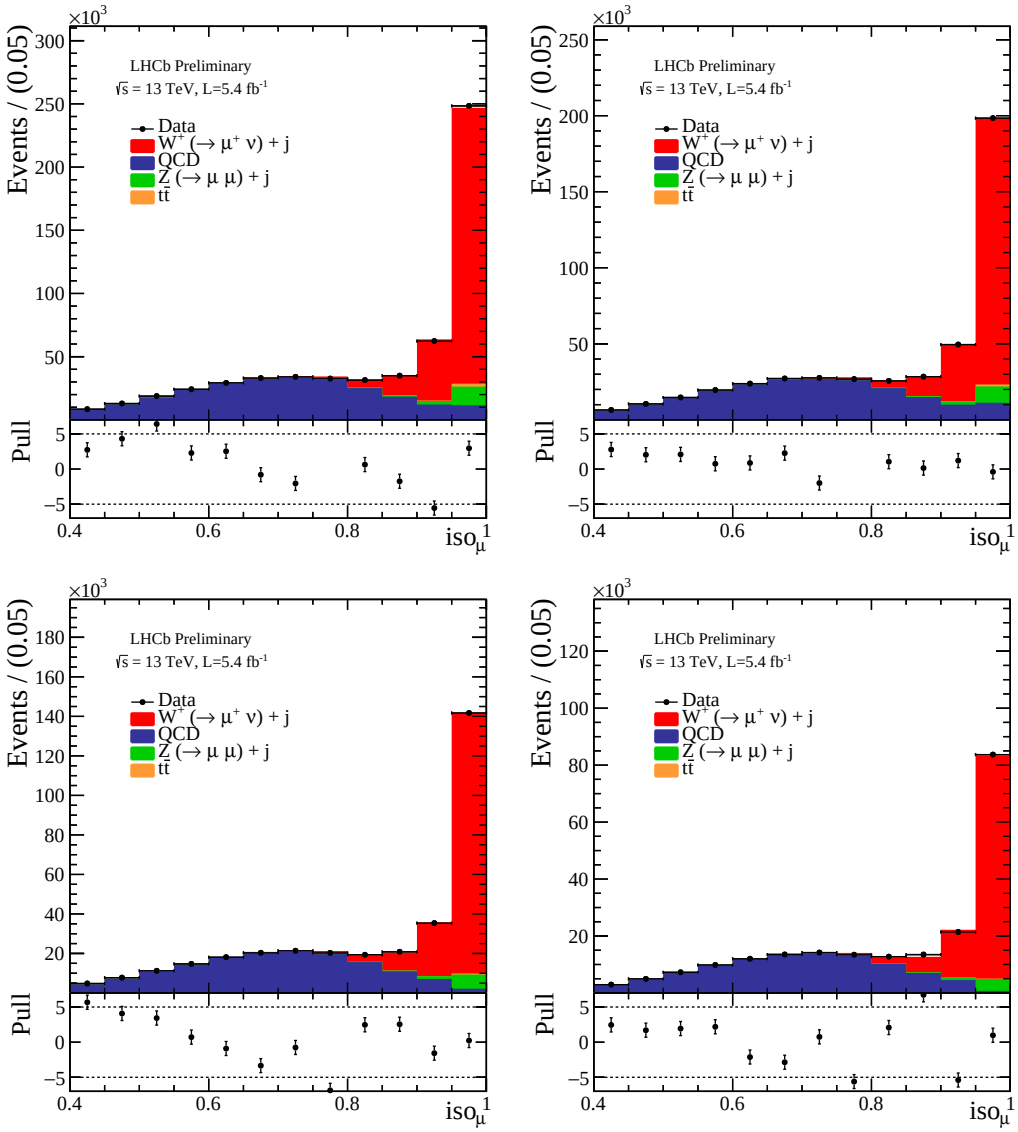


Figure B.4: I_μ template fit as a function of η^{jet} for W^++ 1 jet selection.

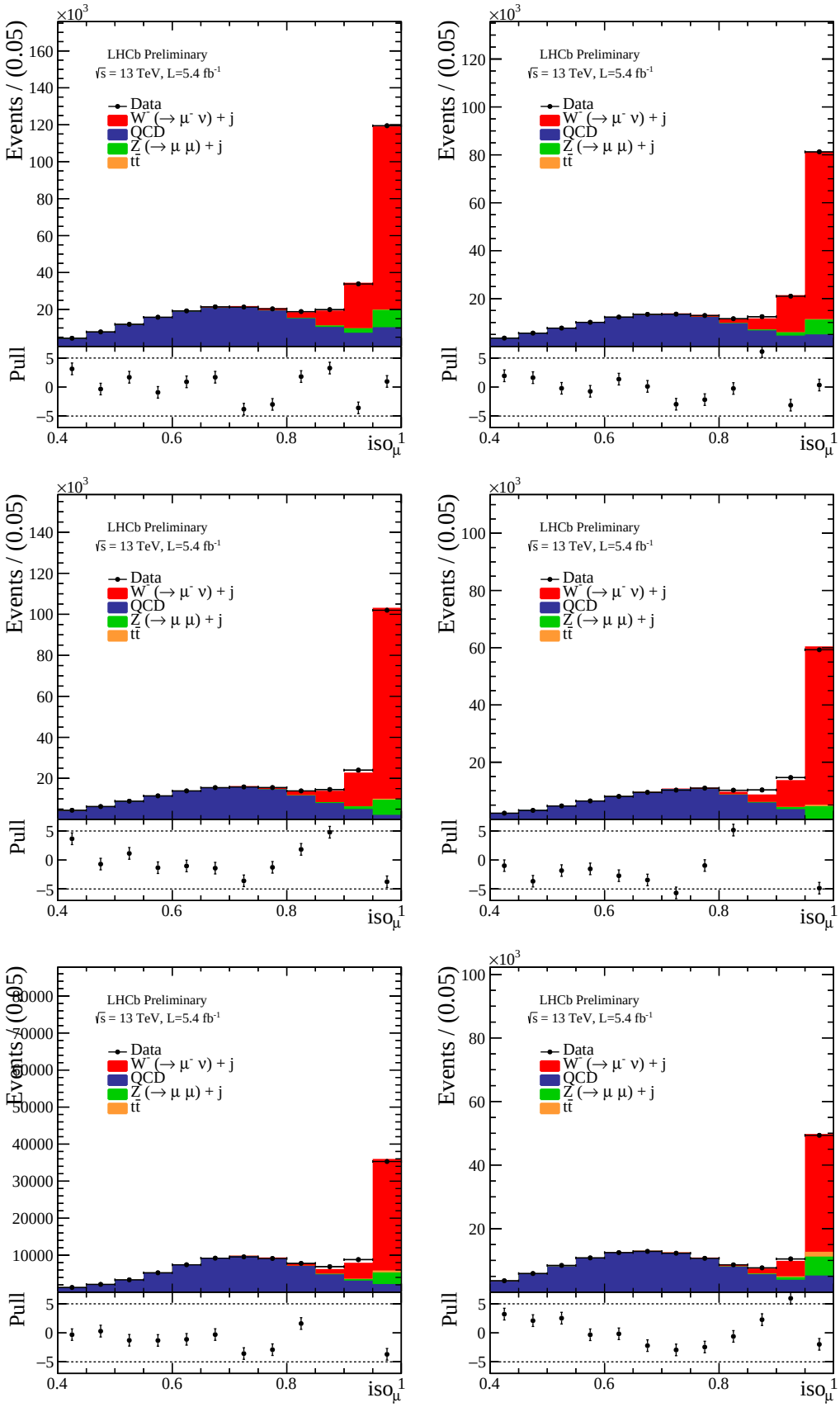


Figure B.5: I_μ template fit as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ for W^-+ 1 jet selection.

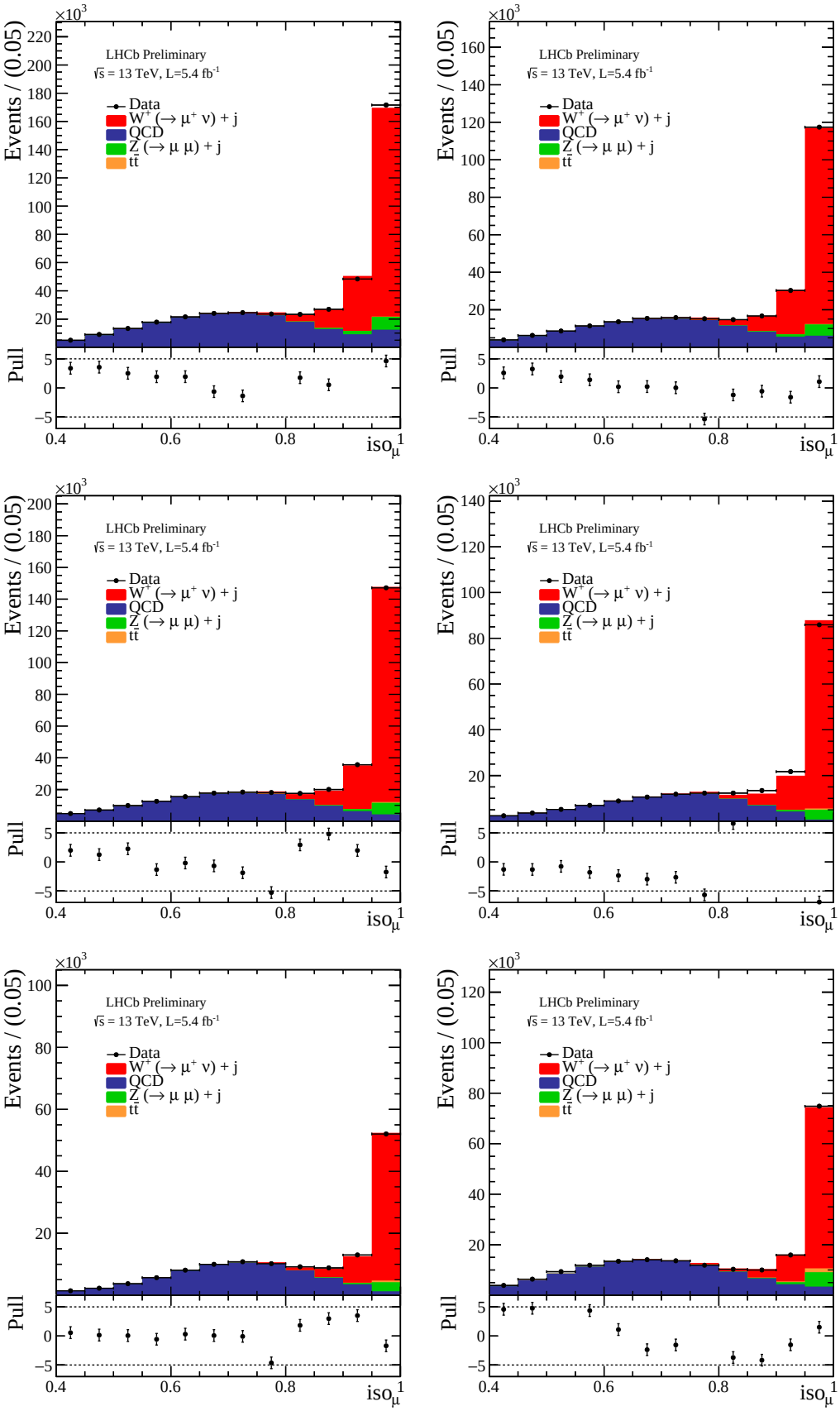


Figure B.6: I_μ template fit as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$ for W^++ 1 jet selection.

Appendix C

Jet reconstruction and identification

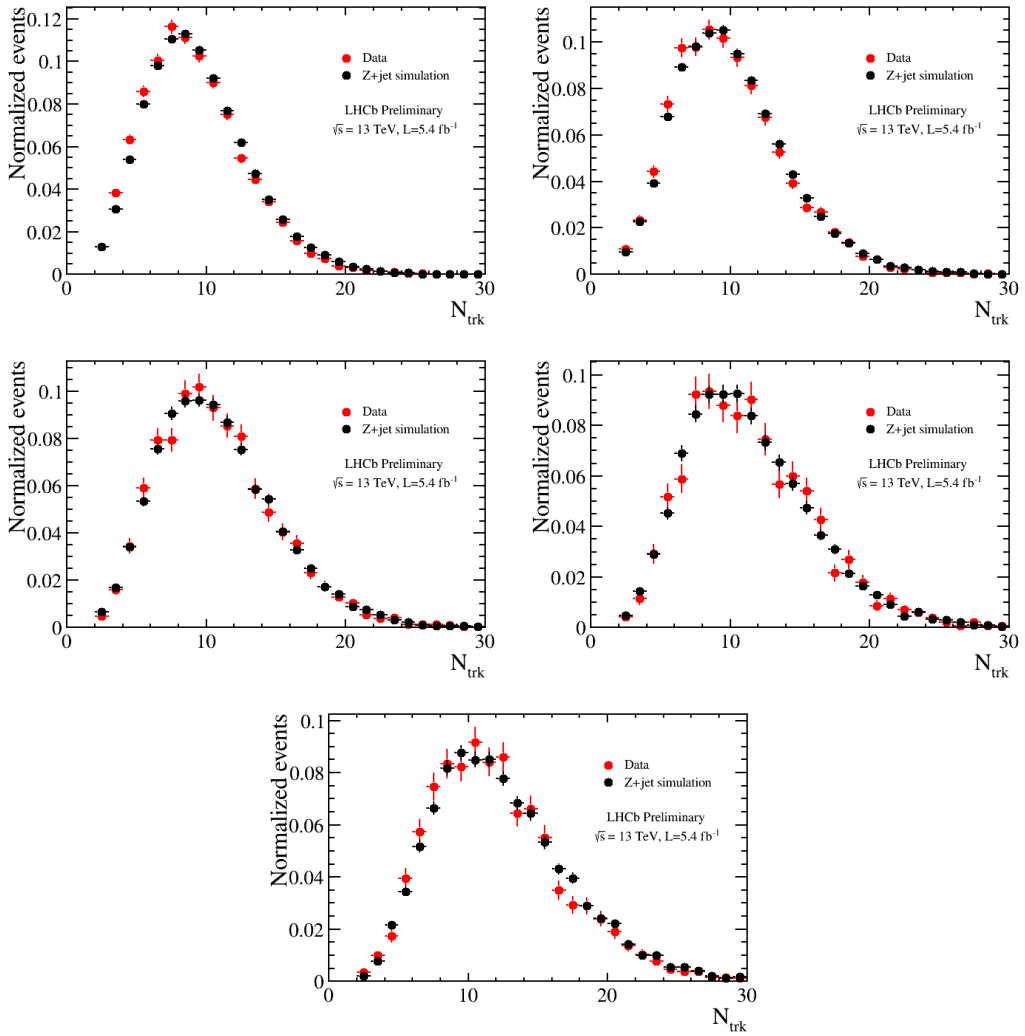


Figure C.1: Distribution of N_{trk} for $Z \to \mu\mu$+jet in data and simulations for different $p_{\mathrm{T}}^{\mathrm{jet}}$ bins

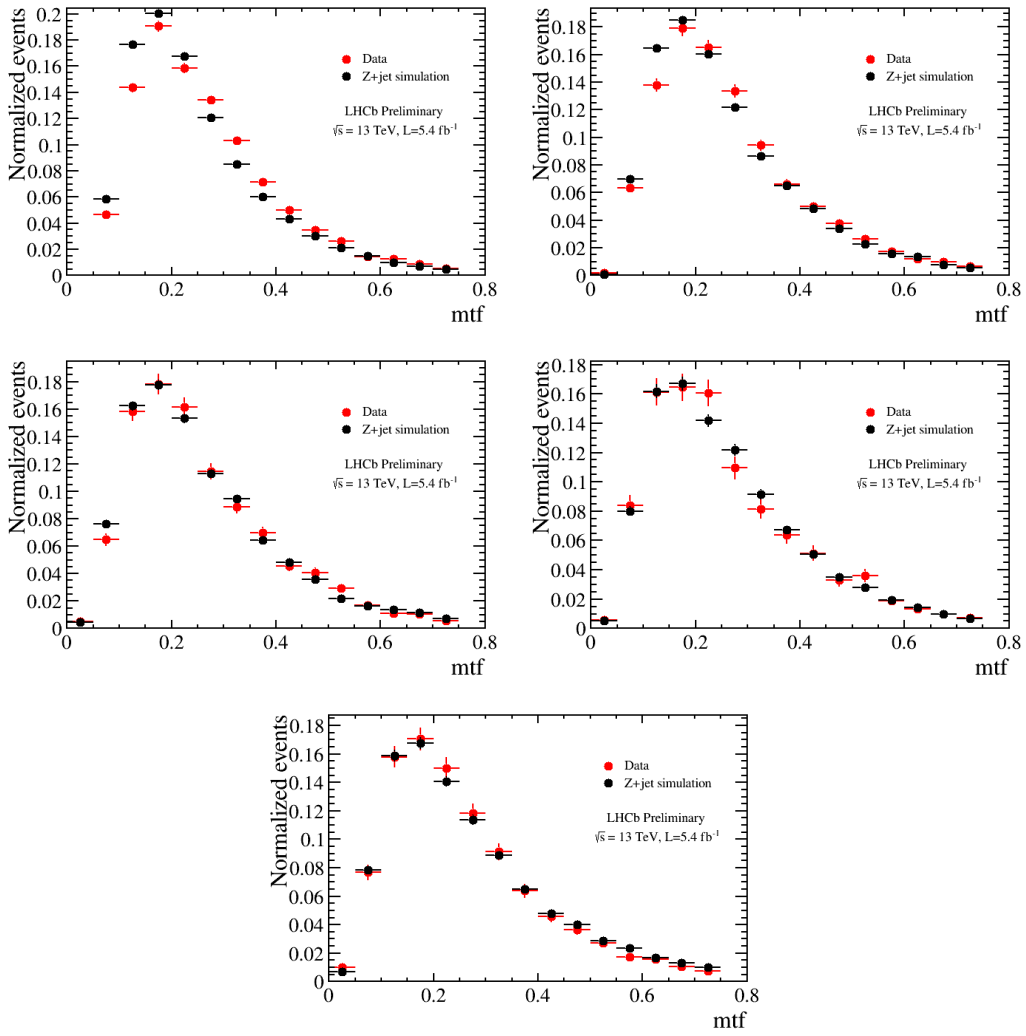


Figure C.2: Distribution of mtf for $Z \to \mu\mu$+jet in data and simulations for different $p_{\mathrm{T}}^{\mathrm{jet}}$ bins

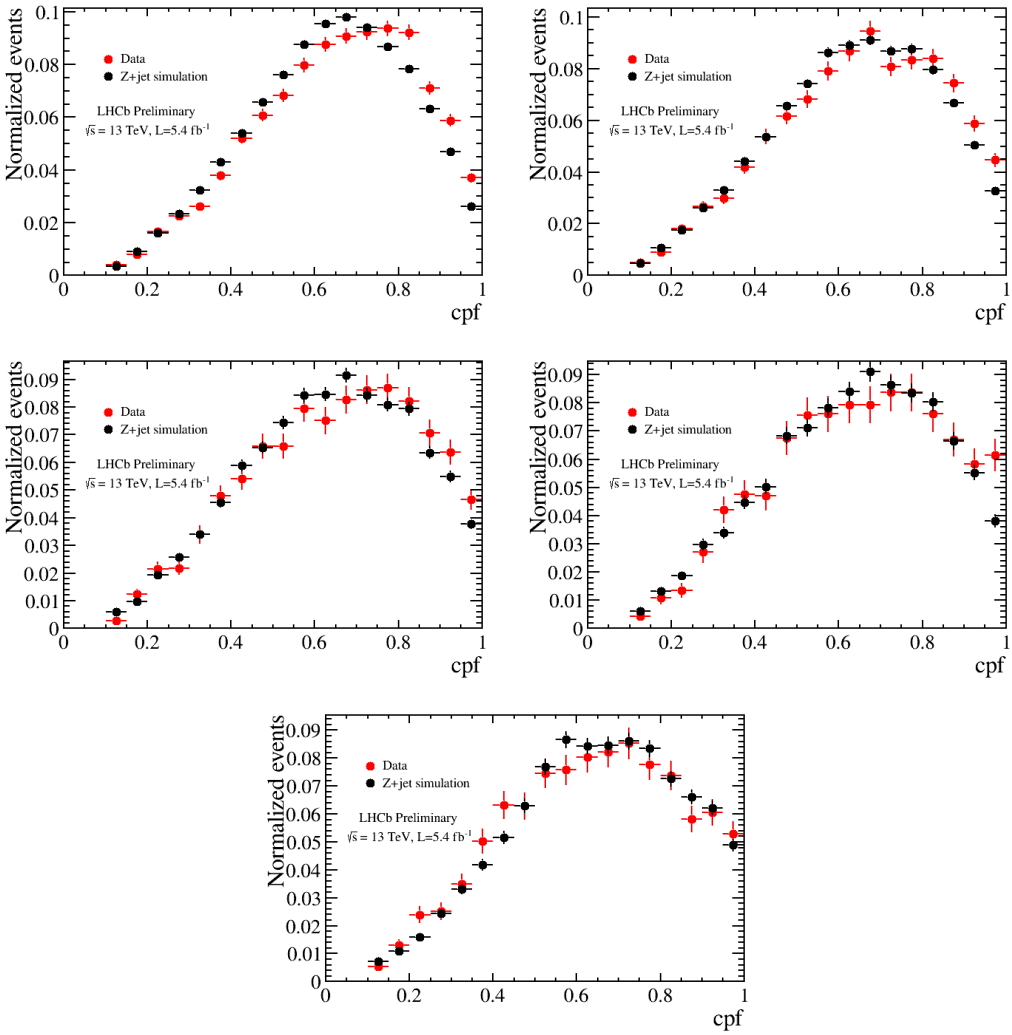


Figure C.3: Distributions of cpf for $Z \to \mu\mu$+jet in data and simulations for different $p_{\mathrm{T}}^{\mathrm{jet}}$ bins.

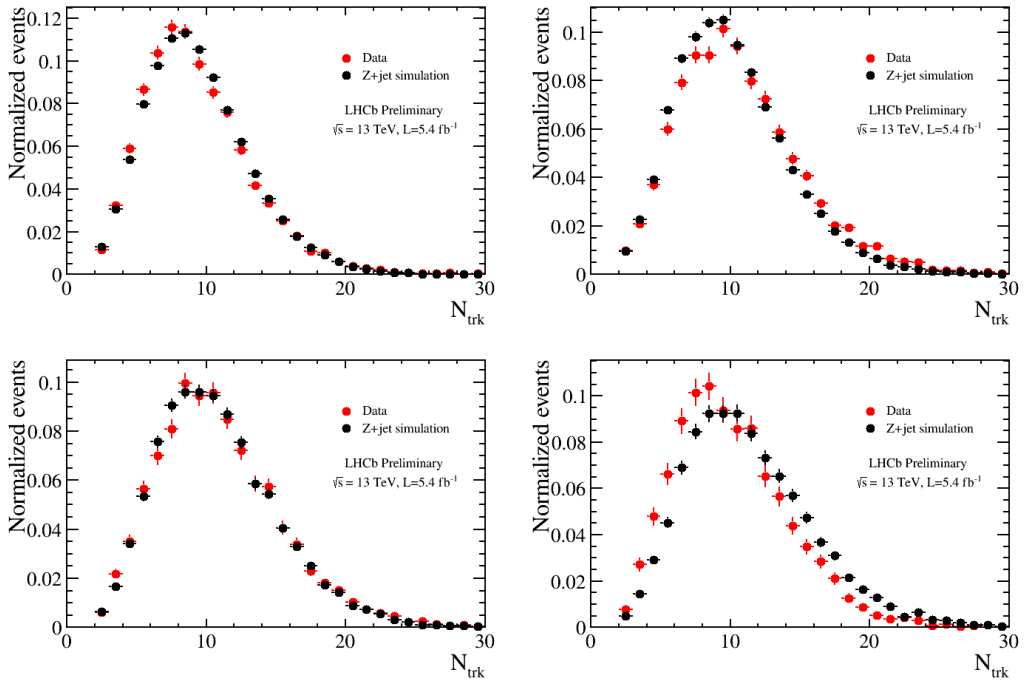


Figure C.4: Distributions of N_{trk} for $Z \to \mu\mu$+jet in data and simulations for different η^{jet} bins.

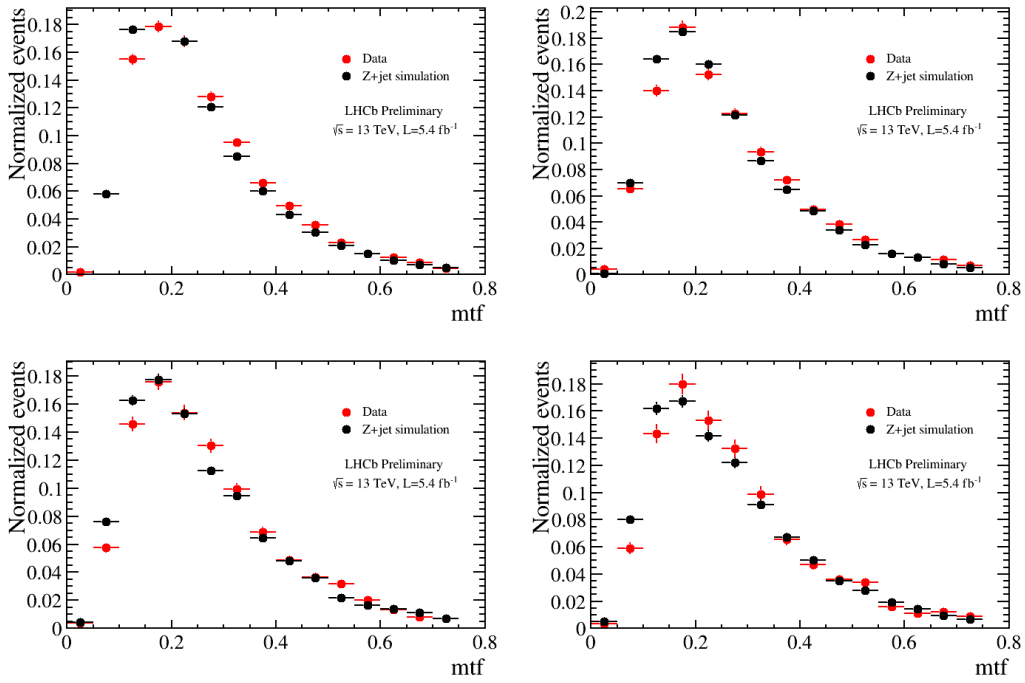


Figure C.5: Distributions of mtf for $Z \to \mu\mu$+jet in data and simulations for different η^{jet} bins.

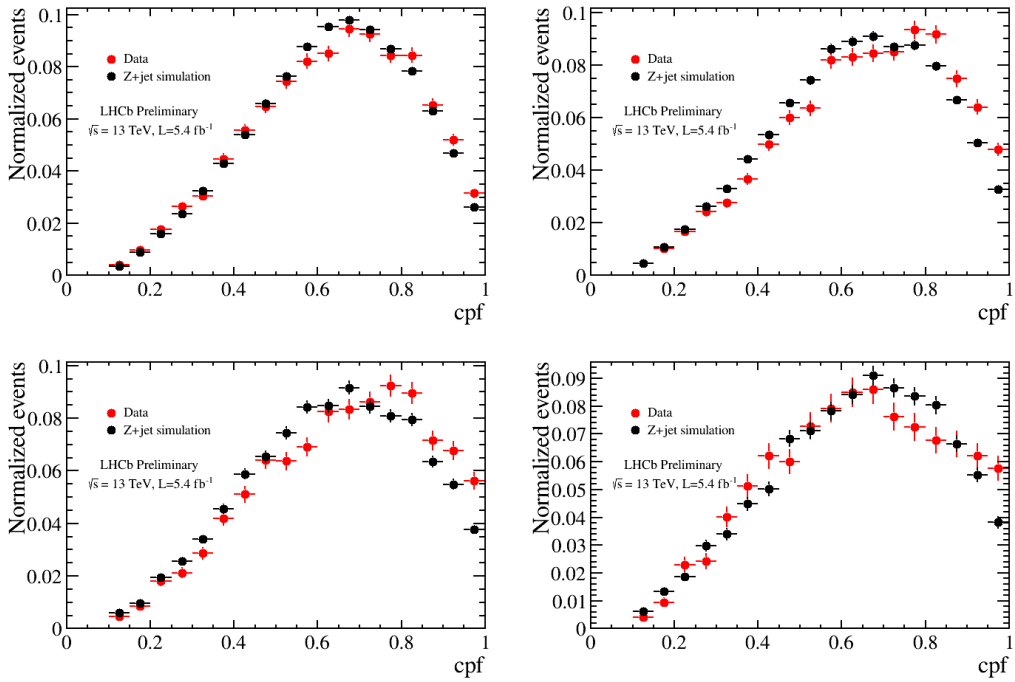


Figure C.6: Distributions of cpf for $Z \to \mu\mu$+jet in data and simulations for different η^{jet} bins.

Appendix D

PID performance with SPACAL module

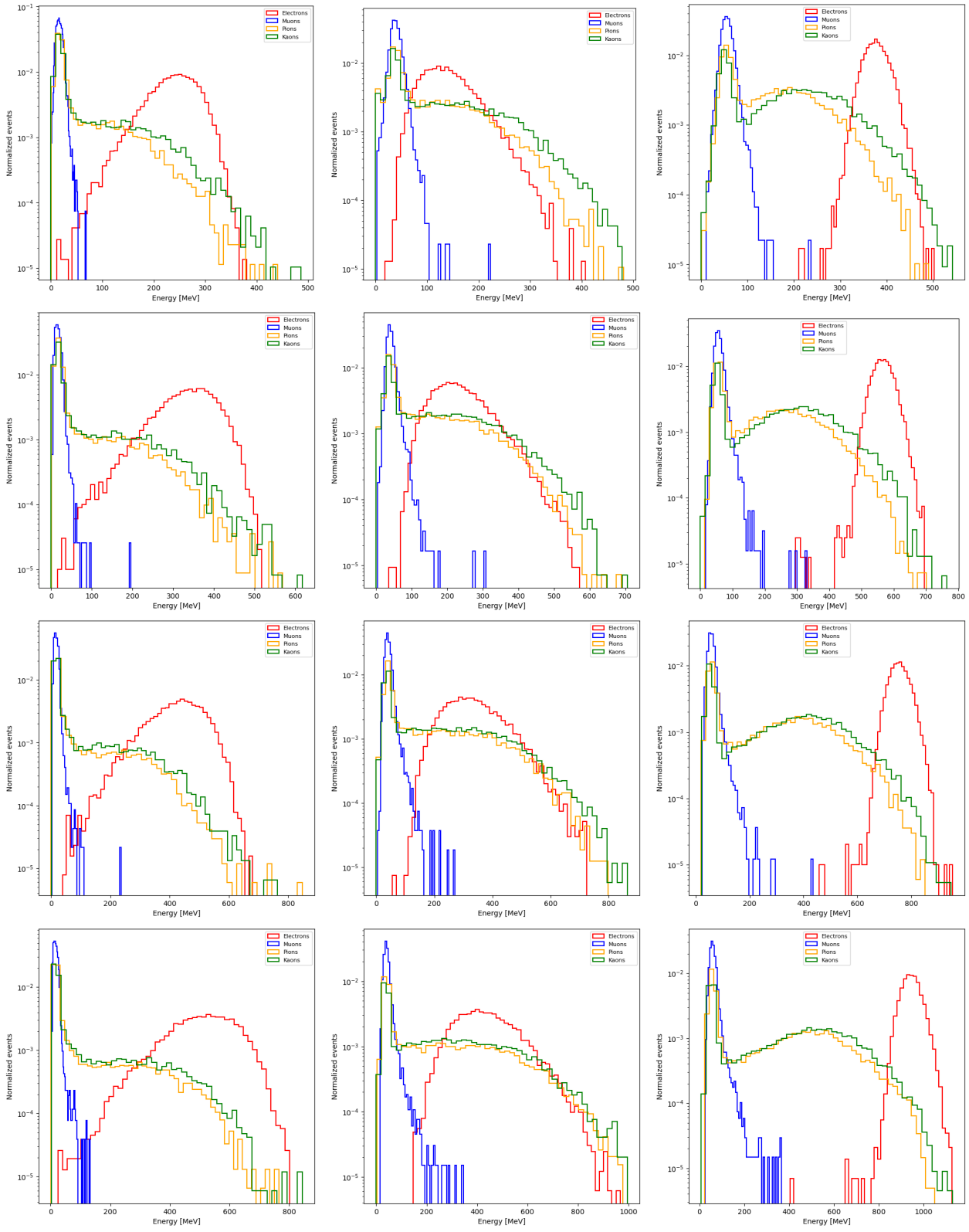


Figure D.1: Energy measured in the front section E_{front} (left column), the energy measured in the back section E_{back} (central column), and energy measured in whole module E_{total} (right column) for electron, muons, pions and kaons with energy from $E = 2$ GeV (first row) to $E = 5$ GeV (fourth row).

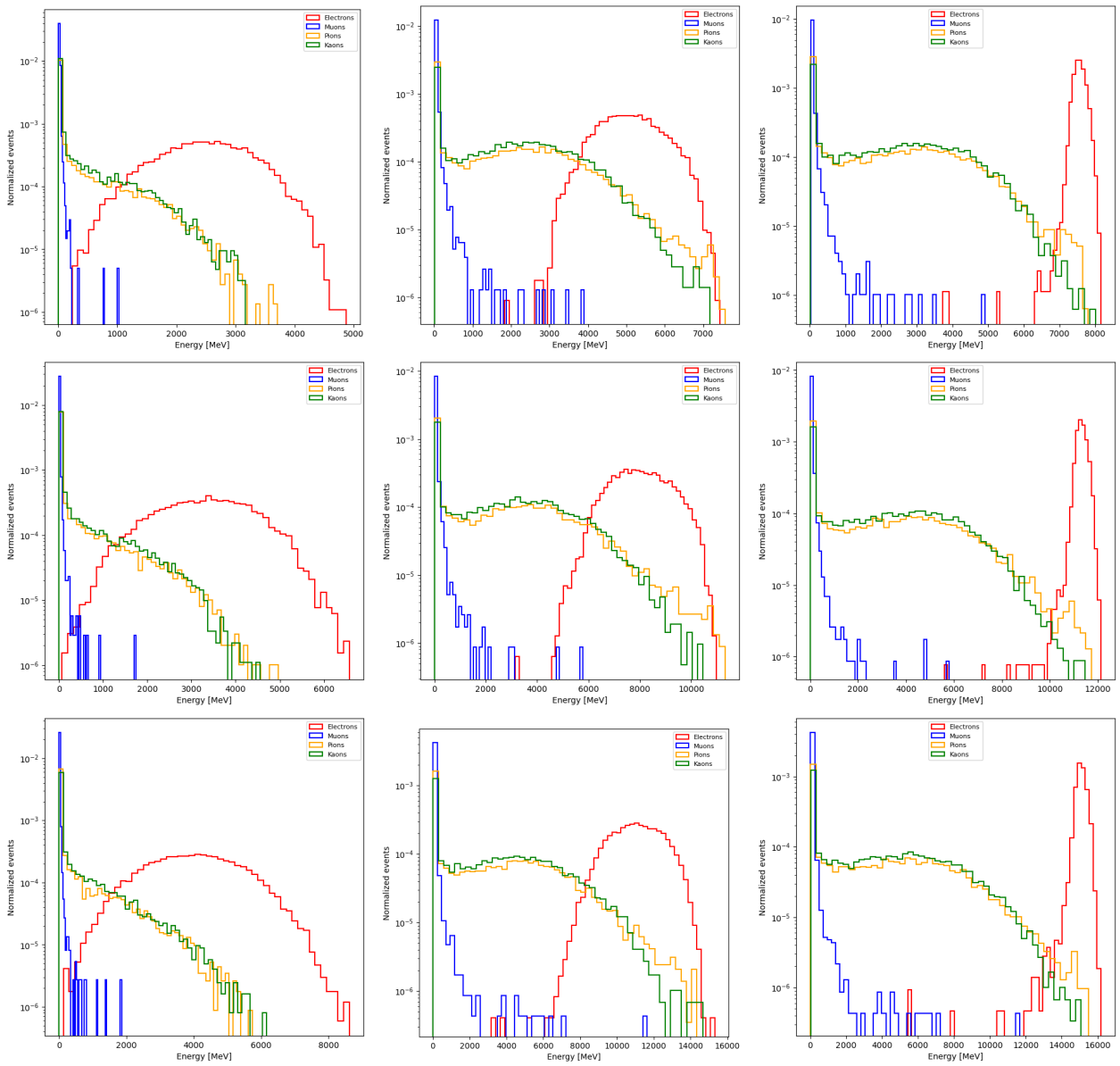


Figure D.2: Energy measured in the front section E_{front} (left column), the energy measured in the back section E_{back} (central column), and energy measured in whole module E_{total} (right column) for electron, muons, pions and kaons with energy from $E = 40$ GeV (first row) to $E = 80$ GeV (third row).

Bibliography

[1] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[2] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[3] G. 't Hooft and M. Veltman. "Regularization and renormalization of gauge fields". In: *Nuclear Physics B* 44.1 (1972), pp. 189–213. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(72)90279-9. URL: https://www.sciencedirect.com/science/article/pii/0550321372902799.

[4] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[5] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[6] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[7] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[8] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[9] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[10] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[12] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[13] T. W. B. Kibble. "Symmetry Breaking in Non-Abelian Gauge Theories". In: *Phys. Rev.* 155 (5 Mar. 1967), pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554. URL: https://link.aps.org/doi/10.1103/PhysRev.155.1554.

[14] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[15] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[16] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Nonabelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (1973). Ed. by J. C. Taylor, pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[17] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (1973). Ed. by J. C. Taylor, pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346.

[18] Shaw G. Mandl F. *Quantum Field Theory*. Wiley, 2009.

[19] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996. DOI: 10.1017/CBO9780511628788.

[20] J. D. Bjorken. "Current algebra at small distances". In: *Conf. Proc. C* 670717 (1967), pp. 55–81.

[21] Thomson M. A. *Modern Particle Physics*. Cambridge University Press, 2013.

[22] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. URL: https://doi.org/10.1140/epjc/s10052-017-5199-5.

[23] Guido Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[24] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[25] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[26] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[27] R. Frederix et al. "The automation of next-to-leading order electroweak calculations". In: *JHEP* 07 (2018). [Erratum: JHEP 11, 085 (2021)], p. 185. DOI: 10.1007/JHEP11(2021)085. arXiv: 1804.10017 [hep-ph].

[28] Simone Alioli et al. "Vector boson plus one jet production in POWHEG". In: *Journal of High Energy Physics* 2011.1 (2011), p. 95. DOI: 10.1007/JHEP01(2011)095. URL: https://doi.org/10.1007/JHEP01(2011)095.

[29] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[30] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[31] R. Blair et al. *The CDF-II detector: Technical design report*. Tech. rep. 1996. URL: https://cds.cern.ch/record/1478626.

[32] T. Aaltonen et al. "Measurement of the cross section for W-boson production in association with jets in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 77 (1 Jan. 2008), p. 011108. DOI: 10.1103/PhysRevD.77.011108. URL: https://link.aps.org/doi/10.1103/PhysRevD.77.011108.

[33] V. M. Abazov et al. "D0 Run IIB Upgrade Technical Design Report". In: (Dec. 2002).

[34] V.M. Abazov et al. "Measurements of inclusive W+jets production rates as a function of jet transverse momentum in pp$^-$ collisions at s=1.96 TeV". In: *Physics Letters B* 705.3 (2011), pp. 200–207. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2011.10.011. URL: https://www.sciencedirect.com/science/article/pii/S0370269311012433.

[35] T. Aaltonen et al. "Measurement of the W^+W^- Production Cross Section and Search for Anomalous $WW\gamma$ and WWZ Couplings in $p\overline{p}$ Collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. Lett.* 104 (20 May 2010), p. 201801. DOI: 10.1103/PhysRevLett.104.201801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.201801.

[36] T. Aaltonen et al. "Measurement of the $WW + WZ$ Production Cross Section Using the lepton + jets Final State at CDF II". In: *Phys. Rev. Lett.* 104 (10 Mar. 2010), p. 101801. DOI: 10.1103/PhysRevLett.104.101801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.101801.

[37] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003.

[38] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[39] Morad Aaboud et al. "Measurement of differential cross sections and W^+/W^- cross-section ratios for W boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *JHEP* 05 (2018). [Erratum: JHEP 10, 048 (2020)], p. 077. DOI: 10.1007/JHEP05(2018)077. arXiv: 1711.03296 [hep-ex].

[40] Morad Aaboud et al. "Measurement of fiducial and differential W^+W^- production cross-sections at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 79.10 (2019), p. 884. DOI: 10.1140/epjc/s10052-019-7371-6. arXiv: 1905.04242 [hep-ex].

[41] Albert M Sirunyan et al. "Measurement of the differential cross sections for the associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 96.7 (2017), p. 072005. DOI: 10.1103/PhysRevD.96.072005. arXiv: 1707.05979 [hep-ex].

[42] Vardan Khachatryan et al. "Measurements of differential cross sections for associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Rev. D* 95 (2017), p. 052002. DOI: 10.1103/PhysRevD.95.052002. arXiv: 1610.04222 [hep-ex].

[43] Albert M Sirunyan et al. "W$^+$W$^-$ boson pair production in proton-proton collisions at $\sqrt{s} =$ 13 TeV". In: *Phys. Rev. D* 102.9 (2020), p. 092001. DOI: 10.1103/PhysRevD.102.092001. arXiv: 2009.00119 [hep-ex].

[44] Roel Aaij et al. "Study of W boson production in association with beauty and charm". In: *Phys. Rev. D* 92.5 (2015), p. 052001. DOI: 10.1103/PhysRevD.92.052001. arXiv: 1505.04051 [hep-ex].

[45] Roel Aaij et al. "Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 08 (2015), p. 039. DOI: 10.1007/JHEP08(2015)039. arXiv: 1505.07024 [hep-ex].

[46] Roel Aaij et al. "Measurement of the forward W boson cross-section in pp collisions at $\sqrt{s} = 7\,\text{TeV}$". In: *JHEP* 12 (2014), p. 079. DOI: 10.1007/JHEP12(2014)079. arXiv: 1408.4354 [hep-ex].

[47] Roel Aaij et al. "Measurement of forward $t\bar{t}$, $W + b\bar{b}$ and $W + c\bar{c}$ production in pp collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Lett. B* 767 (2017), pp. 110–120. DOI: 10.1016/j.physletb.2017.01.044. arXiv: 1610.08142 [hep-ex].

[48] Roel Aaij et al. "Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 09 (2016), p. 136. DOI: 10.1007/JHEP09(2016)136. arXiv: 1607.06495 [hep-ex].

[49] Roel Aaij et al. "Measurement of forward W and Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *JHEP* 05 (2016), p. 131. DOI: 10.1007/JHEP05(2016)131. arXiv: 1605.00951 [hep-ex].

[50] Heather McKenzie Wark. "Measurement of the top quark pair production cross-section in the μeb final state at $\sqrt{s}$ =13 TeV with the LHCb detector". PhD thesis. U. Liverpool (main), 2020. DOI: 10.17638/03084026.

[51] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *Eur. Phys. J. C* 75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[52] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088/1748-0221/3/08/s08001.

[53] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088/1748-0221/3/08/s08005.

[54] F. Follin and D. Jacquet. "Implementation and experience with luminosity levelling with offset beam". In: *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*. 2014, pp. 183–187. DOI: 10.5170/CERN-2014-004.183. arXiv: 1410.3667 [physics.acc-ph].

[55] P R Barbosa-Marinho et al. *LHCb VELO (VErtex LOcator): Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/504321.

[56] "LHCb detector performance". In: *International Journal of Modern Physics A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227.

[57] S Amato et al. *LHCb magnet: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/424338.

[58] R Antunes-Nobrega et al. *LHCb reoptimized detector design and performance: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/630827.

[59] "Performance of the LHCb Outer Tracker". In: *Journal of Instrumentation* 9.01 (Jan. 2014), P01002. DOI: 10.1088/1748-0221/9/01/P01002. URL: https://dx.doi.org/10.1088/1748-0221/9/01/P01002.

[60] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC". In: *The European Physical Journal C* 73.5 (2013), p. 2431. DOI: 10.1140/epjc/s10052-013-2431-9. URL: https://doi.org/10.1140/epjc/s10052-013-2431-9.

[61] Neville Harnew. "Particle identification with the LHCb RICH system". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 563.2 (2006). TRDs for the Third Millenium, pp. 326–332. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2006.02.167. URL: https://www.sciencedirect.com/science/article/pii/S0168900206004475.

[62] Irina Machikhiliyan. "First years of running for the LHCb Calorimeter System". In: (2013). URL: https://cds.cern.ch/record/1543420.

[63] Carlos Abellán Beteta et al. *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*. Tech. rep. All figures and tables, along with any supplementary material and additional information, are available at http://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbPro DP-2020-001.html (LHCb public pages). 2020. arXiv: 2008.11556. URL: https://cds.cern.ch/record/2729028.

[64] Irina Machikhiliyan and LHCb calorimeter group. "The LHCb electromagnetic calorimeter". In: *Journal of Physics: Conference Series* 160.1 (Apr. 2009), p. 012047. DOI: 10.1088/1742-6596/160/1/012047. URL: https://dx.doi.org/10.1088/1742-6596/160/1/012047.

[65] LHCb Collaboration. *LHCb PID Upgrade Technical Design Report*. Tech. rep. 2013. URL: https://cds.cern.ch/record/1624074.

[66] A A Alves Jr et al. "Performance of the LHCb muon system". In: *Journal of Instrumentation* 8.02 (Feb. 2013), P02022. DOI: 10.1088/1748-0221/8/02/P02022. URL: https://dx.doi.org/10.1088/1748-0221/8/02/P02022.

[67] R Aaij et al. "The LHCb trigger and its performance in 2011". In: *Journal of Instrumentation* 8.04 (Apr. 2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. URL: https://dx.doi.org/10.1088/1748-0221/8/04/P04022.

[68] Roel Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *JINST* 14.04 (2019). 46 pages, 35 figures, 1 table. All figures and tables are available at https://cern.ch/lhcbproject/Publications/LHCbProjectPublic/LHCb-DP-2019-001.html, P04013. DOI: 10.1088/1748-0221/14/04/P04013. arXiv: 1812.10790. URL: https://cds.cern.ch/record/2652801.

[69] G. Barrand et al. "GAUDI — A software architecture and framework for building HEP data processing applications". In: *Computer Physics Communications* 140.1 (2001). CHEP2000, pp. 45–55. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(01)00254-5. URL: https://www.sciencedirect.com/science/article/pii/S0010465501002545.

[70] M Clemencic et al. "The LHCb Simulation Application, Gauss: Design, Evolution and Experience". In: *Journal of Physics: Conference Series* 331.3 (Dec. 2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023. URL: https://dx.doi.org/10.1088/1742-6596/331/3/032023.

[71] S. Agostinelli et al. "GEANT4–a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[72] J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278. DOI: 10.1109/TNS.2006.869826.

[73] Gloria Corti et al. "Software for the LHCb experiment". In: *IEEE Trans. Nucl. Sci.* 53 (2006), pp. 1323–1328. URL: https://cds.cern.ch/record/913834.

[74] Rene Brun and Fons Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X. URL: https://www.sciencedirect.com/science/article/pii/S016890029700048X.

[75] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[76] The LHCb collaboration. "Measurement of the track reconstruction efficiency at LHCb". In: *Journal of Instrumentation* 10.02 (Feb. 2015), P02007. DOI: 10.1088/1748-0221/10/02/P02007. URL: https://dx.doi.org/10.1088/1748-0221/10/02/P02007.

[77] Marcin Kucharczyk, Piotr Morawski, and Mariusz Witek. *Primary Vertex Reconstruction at LHCb*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1756296.

[78] Roel Aaij et al. *Optimization of the muon reconstruction algorithms for LHCb Run 2*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2253050.

[79] Menglin Xu, Hang Yin, and Stephen Farry. *Measurement of high-p_T muon reconstruction efficiencies for studies of γ^*/Z production and angular coefficients with LHCb RunII data-sets*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2747827.

[80] Florian Beaudette. "The CMS Particle Flow Algorithm". In: (2013). Comments: 10 pages, 12 figures, proceedings for CHEF2013, Paris, France, April 2013. arXiv: 1401.8155. URL: https://cds.cern.ch/record/1645993.

[81] William Barter. "Z boson and associated jet production at the LHCb experiment". Presented 27 May 2014. 2014. URL: https://cds.cern.ch/record/1970903.

[82] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[83] Rodrigues, Eduardo. "The Scikit-HEP Project". In: *EPJ Web Conf.* 214 (2019), p. 06005. DOI: 10.1051/epjconf/201921406005. URL: https://doi.org/10.1051/epjconf/201921406005.

[84] The LHCb collaboration. "Identification of beauty and charm quark jets at LHCb". In: *Journal of Instrumentation* 10.06 (June 2015), P06013. DOI: 10.1088/1748-0221/10/06/P06013. URL: https://dx.doi.org/10.1088/1748-0221/10/06/P06013.

[85] R. Aaij et al. "First Measurement of the Charge Asymmetry in Beauty-Quark Pair Production". In: *Physical Review Letters* 113.8 (Aug. 2014). ISSN: 1079-7114. DOI: 10.1103/physrevlett.113.082003. URL: http://dx.doi.org/10.1103/PhysRevLett.113.082003.

[86] R. Aaij et al. "Measurement of the W boson mass". In: *Journal of High Energy Physics* 2022.1 (2022), p. 36. DOI: 10.1007/JHEP01(2022)036. URL: https://doi.org/10.1007/JHEP01(2022)036.

[87] R. Aaij et al. "First Observation of Top Quark Production in the Forward Region". In: *Phys. Rev. Lett.* 115 (11 Sept. 2015), p. 112001. DOI: 10.1103/PhysRevLett.115.112001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.112001.

[88] Z. Was. "TAUOLA the library for τ lepton decay, and KKMC/KORALB/KORALZ/... status report". In: *Nuclear Physics B - Proceedings Supplements* 98.1 (2001), pp. 96–102. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(01)01200-2. URL: https://www.sciencedirect.com/science/article/pii/S0920563201012002.

[89] Andreas Hocker and Vakhtang Kartvelishvili. "SVD approach to data unfolding". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 372.3 (1996), pp. 469–481. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(95)01478-0. URL: https://www.sciencedirect.com/science/article/pii/0168900295014780.

[90] Tim Adye. "Unfolding algorithms and tests using RooUnfold". In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 313–318. DOI: 10.5170/CERN-2011-006.313. arXiv: 1105.1160 [physics.data-an].

[91] The LHCb collaboration. "Precision luminosity measurements at LHCb". In: *Journal of Instrumentation* 9.12 (Dec. 2014), P12005. DOI: 10.1088/1748-0221/9/12/P12005. URL: https://dx.doi.org/10.1088/1748-0221/9/12/P12005.

[92] S. Dittmaier et al. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN Yellow Reports: Monographs. The authors dedicate this Report to the memory of Nicola Cabibbo and Georges Charpak. Geneva: CERN, 2011. DOI: 10.5170/CERN-2011-002. URL: https://cds.cern.ch/record/1318996.

[93] Michelangelo L. Mangano et al. "ALPGEN, a generator for hard multiparton processes in hadronic collisions". In: *Journal of High Energy Physics* 2003.07 (July 2003), p. 001. DOI: 10.1088/1126-6708/2003/07/001. URL: https://dx.doi.org/10.1088/1126-6708/2003/07/001.

[94] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[95] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[96] James P. Egan, Gordon Z. Greenberg, and Arthur I. Schulman. "Operating Characteristics, Signal Detectability, and the Method of Free Response". In: *The Journal of the Acoustical Society of America* 33.8 (1961), pp. 993–1007. DOI: 10.1121/1.1908935. eprint: https://doi.org/10.1121/1.1908935. URL: https://doi.org/10.1121/1.1908935.

[97] Roel Aaij et al. *Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era*. Tech. rep. ISBN 978-92-9083-494-6. Geneva: CERN, 2016. DOI: 10.17181/CERN.QZRZ.R4S6. arXiv: 1808.08865. URL: https://cds.cern.ch/record/2636441.

[98] A.M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011. URL: http://dx.doi.org/10.1088/1748-0221/13/05/P05011.

[99] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012. URL: http://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[100] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012. URL: http://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[101] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[102] Alessio Piucci. "The LHCb Upgrade". In: *Journal of Physics: Conference Series* 878.1 (July 2017), p. 012012. DOI: 10.1088/1742-6596/878/1/012012. URL: https://dx.doi.org/10.1088/1742-6596/878/1/012012.

[103] CERN (Meyrin) LHCb Collaboration. *Framework TDR for the LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era* . Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2776420.

[104] L. An et al. "Performance of a spaghetti calorimeter prototype with tungsten absorber and garnet crystal fibres". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1045 (2023), p. 167629. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2022.167629. URL: https://www.sciencedirect.com/science/article/pii/S0168900222009214.

[105] O Deschamps et al. *Photon and neutral pion reconstruction.* Tech. rep. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/691634.

[106] Miriam Calvo Gomez et al. *Studies of a neutron shielding for the upgraded LHCb detector.* Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2268537.

[107] Pietro Albicocco et al. *Considerations on Muon detector upgrade II.* Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2714057.

[108] Wikimedia Commons. *The Bloch sphere, a geometric representation of a two-level quantum system.* 2009. URL: https://commons.wikimedia.org/wiki/File:Bloch%5C_sphere.svg.

[109] Marcello Benedetti et al. "Parameterized quantum circuits as machine learning models". In: *Quantum Science and Technology* 4.4 (Nov. 2019), p. 043001. ISSN: 2058-9565. DOI: 10.1088/2058-9565/ab4eb5. URL: http://dx.doi.org/10.1088/2058-9565/ab4eb5.

[110] K. Mitarai et al. "Quantum circuit learning". In: *Physical Review A* 98.3 (Sept. 2018). ISSN: 2469-9934. DOI: 10.1103/physreva.98.032309. URL: http://dx.doi.org/10.1103/PhysRevA.98.032309.

[111] Maria Schuld et al. "Evaluating analytic gradients on quantum hardware". In: *Physical Review A* 99.3 (Mar. 2019). ISSN: 2469-9934. DOI: 10.1103/physreva.99.032331. URL: http://dx.doi.org/10.1103/PhysRevA.99.032331.

[112] Alessio Gianelle et al. "Quantum Machine Learning for b-jet charge identification". In: *JHEP* 08 (2022), p. 014. DOI: 10.1007/JHEP08(2022)014. arXiv: 2202.13943 [hep-ex].

[113] David Krohn et al. "Jet Charge at the LHC". In: *Phys. Rev. Lett.* 110 (21 May 2013), p. 212001. DOI: 10.1103/PhysRevLett.110.212001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.212001.

[114] R.D. Field and R.P. Feynman. "A parametrization of the properties of quark jets". In: *Nuclear Physics B* 136.1 (1978), pp. 1–76. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(78)90015-9. URL: https://www.sciencedirect.com/science/article/pii/0550321378900159.

[115] *Jet Charge Studies with the ATLAS Detector Using $\sqrt{s} = 8$ TeV Proton-Proton Collision Data.* Tech. rep. ATLAS-CONF-2013-086. Aug. 2013. URL: https://cds.cern.ch/record/1572980.

[116] *Measurement of the Jet Vertex Charge algorithm performance for identified b-jets in $t\bar{t}$ events in pp collisions with the ATLAS detector.* Tech. rep. ATLAS-CONF-2018-0220. June 2018. URL: https://cds.cern.ch/record/2622370.

[117] Ville Bergholm et al. *PennyLane: Automatic differentiation of hybrid quantum-classical computations.* 2020. arXiv: 1811.04968 [quant-ph].

[118] François Chollet et al. *Keras.* https://keras.io. 2015.

[119] Martin Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[120] Md Sajid Anis et al. *Qiskit: An Open-source Framework for Quantum Computing*. 2021. DOI: 10.5281/zenodo.2573505.

[121] IBM Quantum. https://quantum-computing.ibm.com/. 2021.