UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Measurement of the strong coupling constant with the LHCb detector

Thesis supervisor

Dr. Lorenzo Sestini

Thesis co-supervisors

Prof. Donatella Lucchesi

Dr. Davide Zuliani

Candidate

Alice Moro

Academic Year 2023/2024

Contents

Notation

Throughout the whole document, the unit system called *natural units* will be adopted. According to it, the reduced Planck's constant $\hbar \simeq 1.1 \times 10^{-34}\,\mathrm{J \cdot s} \simeq 6.6 \times 10^{-16}\,\mathrm{eV \cdot s}$, i.e. the Planck's constant divided by 2π, $\hbar = h/2\pi$, becomes $\hbar = 1$ and $c \simeq 3 \times 10^{8}\,\mathrm{m/s}$ becomes $c = 1$, so that formulae get leaner and easier to read. A direct consequence of this is that energies, momenta, and masses will be all referred to in eV. Their conversion in the International System of Units is:

Energy	$1\,\mathrm{eV}$	$\xrightarrow{c=1}$	$1\,\mathrm{eV} \simeq 1.6 \times 10^{-19}\,\mathrm{J}$
Momentum	$1\,\mathrm{eV/c}$	$\xrightarrow{c=1}$	$1\,\mathrm{eV} \simeq 5.4 \times 10^{-28}\,\mathrm{kg \cdot m/s}$
Mass	$1\,\mathrm{eV/c^2}$	$\xrightarrow{c=1}$	$1\,\mathrm{eV} \simeq 1.8 \times 10^{-36}\,\mathrm{kg}$

Abstract

The goal of this thesis is to obtain a preliminary estimation of the strong coupling constant α_s and its main systematic uncertainties studying data samples collected with the LHCb experiment at LHC, CERN. For the first time at LHCb, pp data samples taken in 2016 during LHC Run 2 at a center of mass energy $\sqrt{s} = 13\,\text{TeV}$ and integrated luminosity $\mathcal{L} = 1.6\,\text{fb}^{-1}$, and containing hadronic jets are analyzed for this aim.

The final estimate of α_s is extracted from the inclusive dijet cross section, and reads:

$$\alpha_s\left(m_{Z^0}^2\right) = 0.1143 \pm 0.0007\ (\text{stat.}) \pm 0.009\ (\text{syst.}) \pm 0.002\ (\text{lumi.}) = 0.114 \pm 0.009$$

where the uncertainties come from the statistics, the corrections related to Jet Energy Scale and Jet Energy Resolution, and the integrated luminosity contribution.

Introduction

Strong interactions of quarks and gluons are described by the gauge field theory named Quantum ChromoDynamics (QCD), which is a component of the Standard Model of particle physics. The strong coupling constant α_s indicates the intensity of the strong interaction and is a so called running coupling constant, as its value depends on the energy scale of the process under study. Among the other fundamental constants, α_s is, at present, measured with the worst precision; it is crucial to improve it, both for theoretical and experimental purposes.

In this thesis, a preliminary estimation of α_s and its main systematic uncertainties is obtained studying Z+jet and dijet data samples collected with the LHCb experiment at LHC, CERN. For the first time at LHCb, data samples of pp collisions taken in 2016 during LHC Run 2 at a center of mass energy of 13 TeV and integrated luminosity $1.6\,\text{fb}^{-1}$ containing hadronic jets are analyzed for this aim. The estimate of α_s is extracted from the inclusive dijet cross section quantified with the dijet sample, while both Z+jet and dijet sets are employed as calibration samples to determine the systematic corrections required for measuring jets.

After introducing the Standard Model and the current status of α_s measurement, the LHCb experiment, the jet reconstruction techniques, and the performances of LHCb in reconstructing muons, the two main parts of this work are reported. The first part regards the development of a fast simulation to establish the proper analysis strategy, which turns out to be the determination of the dijet cross section. The second part consists in the extraction of the dijet cross section from experimental data and the comparison of it with the simulation in order to determine α_s and its uncertainty. A detector calibration is performed to study the contributions to the systematic uncertainty, which come mainly from the effects of detector response in jets identification, and are factored in with two energy correction parameters, related to the Jet Energy Scale (JES) and the Jet Energy Resolution (JER). The number of selected events, which is the product of the dijet cross section, the selection efficiency and the integrated luminosity of the data-taking, is derived from experimental data. The effects of JES, JER and integrated luminosity are propagated inside the selection efficiency, and, in turn, they have an impact on the uncertainty of the α_s measurement.

Here, the structure of the thesis is recapped:

- Chapter 1 offers an overview of the Standard Model, with a focus on the properties of QCD and the strong coupling constant α_s, together with the current status of α_s measurement techniques. Also, Parton Distribution Functions and their evolution equations are introduced.
- Chapter 2 outlines the layout of the LHCb detector during Runs 1 and 2, the now operating Upgrade Ia, and the planned Upgrades Ib and II. The role of LHCb in measuring α_s is proposed.
- Chapter 3 provides an introduction to jet physics, then reports the steps of the LHCb Run 2 jet reconstruction algorithm and the dedicated trigger lines.
- Chapter 4 concerns the LHCb muons reconstruction. They are useful for this thesis because one of the analyzed data samples contains a jet and a Z^0 boson decaying in a $\mu^+\mu^-$ pair.
- Chapter 5 describes the first stage of the analysis: the development of a fast simulation of pp events producing jets at three values of α_s, and the estimation of the relative dijet cross sections.

- Chapter 6 is dedicated to the second stage of the analysis. The employed Z+jet and dijet data sets from Run 2 are specified, then the procedures to extract the JES- and JER-related corrections are explained. The dijet cross section resulting from the dijet sample is compared with the simulated ones to get a preliminary estimate of α_s and its statistical and systematic uncertainties. The latter is due to JES, JER, and integrated luminosity.
- Chapter 7 collects some of the possible advancements of this analysis work.

Chapter 1

The strong interaction and α_s: theory and experimental status

1.1 The fundamental interactions and the Standard Model

At present, all physical phenomena can be reduced to four fundamental forces: gravity, electromagnetism, the weak interaction, and the strong interaction. Electromagnetism results from the unification of electricity and magnetism, and then electromagnetism and the weak force are combined into the electroweak theory [1–5]. The electroweak theory and the strong force together form the so called *Standard Model* (SM) of particle physics. Gravity is an attractive force between masses: classically, it is described by General Relativity, but a full quantum theory of it has not been found yet. At particle level the gravitational force is negligible compared to all other forces. One of the big challenges of today's physics is including all the forces into a single *Grand Unified Theory* describing nature [6–8].

The forces forming the SM are defined as *Quantum Field Theories* (QFTs). As the name suggests, the fundamental entities of a QFT are the fields, that permeate all the space. A particle is an excitation or a quantum of the corresponding field, and they can propagate and interact according to the rules dictated by the field equations. QFTs are based on quantum mechanics and special relativity jointly. Since SM takes into account particles that can travel with velocities close to the speed of light, the employment of special relativity is necessary. A *lagrangian* $\mathcal{L}$ is associated to each QFT, serving as Lorentz scalar and relativistic counterpart to the quantum mechanical *hamiltonian* which enters the Schrödinger equation. The lagrangian determines how to derive the *Feynman rules*, which are used to construct *Feynman diagrams* representing interactions, and to calculate properties like cross sections and decay rates. The Feynman rules translate these diagrams into mathematical expressions. Also, each interaction strength is proportional to a *charge*. When dealing with QFTs, *perturbation theory* is often employed, where solutions to complex problems are approximated by expansions in terms of a small parameter, starting from a known unperturbed solution and adding small corrections.

The SM belongs to a special category of QFTs: *gauge theories*. The lagrangian of a gauge theory is invariant under specific *local* transformations from a Lie group, known as *gauge transformations*, which are essential to determine the interactions described by the theory. In particular, the SM is equipped with the $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L \times \mathrm{SU}(3)_C$ gauge symmetry, where the groups refer to the electromagnetic, weak and strong forces, respectively. Gauge theories predict the existence of *gauge bosons*, particles with integer spin that carry the force associated with the gauge symmetries. Gauge bosons are the quanta of the force fields. Other than the gauge bosons, there are the matter particles, i.e. quarks and leptons, which are spin-1/2 fermions. Matter particles interact thanks to the mediation of gauge bosons. Below, a brief overview of the components of the SM, summarized in 1.1, is given.

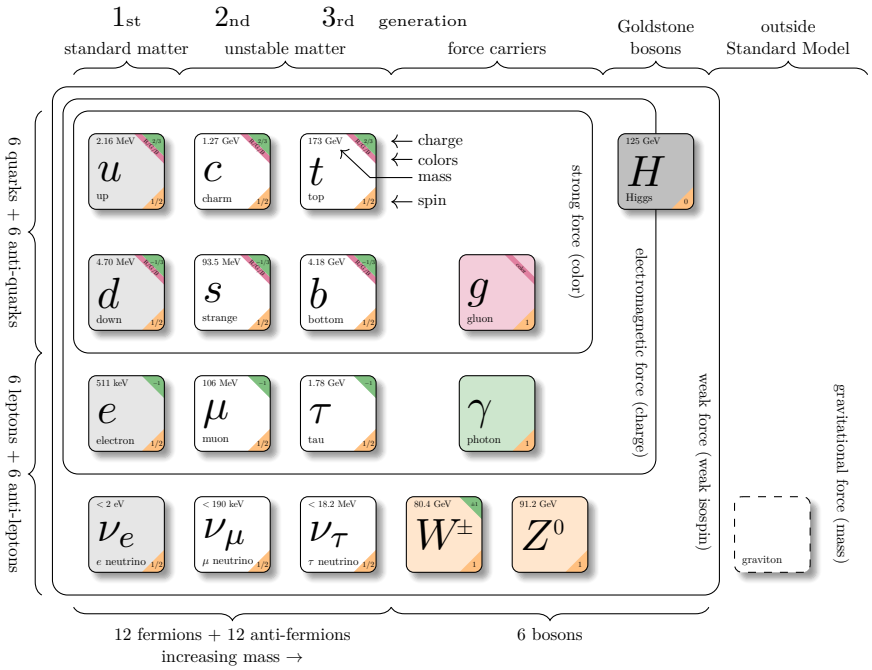

Figure 1.1: Components of the Standard Model of particle physics.

Quarks. They are the elementary particles that experience the electromagnetic, weak, and strong interactions. There are six quarks, divided into three doublets called *families* or *generations*:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}$$

where u, d, c, s, t, b stand for *up*, *down*, *charm*, *strange*, *top* and *bottom* or *beauty* respectively, and are the quark *flavors*. The quarks in the top row have electric charge 2/3 e, while the quarks in the bottom row have charge -1/3 e, where e is the absolute value of the electron's charge. The charge carried by the quarks that responds to the strong force, is called *color*. Each of the mentioned quark flavors comes in three different colors. Because of the strong force, quarks can not exist as free particles, but they are bound inside colorless *hadrons*, which can be *baryons* if they are made of three quarks, or *mesons*, if they contain two quarks. Hence, it is impossible to predict quark masses theoretically, and in fact they are free parameters of the SM that need to be determined experimentally. Their measured values are [9]:

$$m_u = 2.16\,\text{MeV} \qquad m_c = 1.2730\,\text{GeV} \qquad m_t = 172.57\,\text{GeV}$$
$$m_d = 4.70\,\text{MeV} \qquad m_s = 93.5\,\text{MeV} \qquad m_b = 4.183\,\text{GeV}$$

Leptons. Similarly, there are six leptons organized in three families:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

In the bottom row there are the electron e, the muon μ and the tau τ. They are charged with a $-e$ electric charge. In the top row there are the neutrinos, each one paired to a lepton. Neutrinos are electrically neutral. Also lepton masses are measured experimentally and are [10]:

$$m_e = 0.510\,999\,\text{MeV} \qquad m_\mu = 105.658\,\text{MeV} \qquad m_\tau = 1776.93\,\text{MeV}$$

As far as neutrino masses are concerned, they have not been measured yet. In the SM lagrangian only massless neutrinos are incorporated, even though limits on non-zero neutrino masses have been predicted, and it is known that at least two of them are different from zero.

Gauge bosons. The interaction carriers are: the photon γ, massless, mediator of the electromagnetic force; $W^{\pm}$ and Z^0 bosons, which carry the weak interaction and whose masses are [11] $m_{W^{\pm}} = 80.3692\,\text{GeV}$ and $m_{Z^0} = 91.1880\,\text{GeV}$; the gluon g, massless, mediator of the strong interaction. A massless graviton, present in figure 1.1, is not included in SM predictions, it has only been predicted by theories beyond the SM.

Higgs boson. Among gauge bosons, solely the $W^{\pm}$ and Z^0 are massive. This is the result of a key electroweak phenomenon, the *Higgs mechanism*, upon which gauge bosons in a gauge theory acquire mass through spontaneous symmetry breaking. In the SM, the Higgs field, a scalar field with a non-zero vacuum expectation value, interacts with the $W^{\pm}$ and Z^0 bosons, giving them mass while preserving the gauge symmetry. This mechanism explains how also all other particles acquire mass without violating the underlying symmetries of the SM. The Higgs boson H is the quantum excitation of the Higgs field, and it is massive with $m_H = 125.20\,\text{GeV}$ [12]. It was theoretically predicted in the 1960s and experimentally discovered at the LHC in 2012, completing all the pieces of the SM.

Each particle has its own *antiparticle*, characterized by opposite values of electric charge, color charge and flavor charge than the corresponding particle, but with same mass and same spin.

1.2 Quantum ChromoDynamics and the strong coupling constant

The strong interaction is responsible for holding atomic nuclei together. Its name stems from the fact that it is much more intense than the electromagnetic force, which would make the protons inside a nucleus repel. As already stated, quarks are the particles that experience the strong force and gluons are the bosons which carry it. *Quantum ChromoDynamics* (QCD) is the part of the SM that models strong interactions, and it is a non-abelian[1] gauge theory with $\text{SU}(3)_C$ gauge symmetry. The lagrangian of QCD [13] is the following:

$$\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{ab})\psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\,\mu\nu} \tag{1.1}$$

where $\psi_{q,a}$ are field spinors for a quark of flavor q, mass m_q, and color a, with a running from 1 to 3, as quarks can be of three colors; a *spinor* is a mathematical entity used in the QFT formalism to describe the quantum state of particles that have half-integer spin, such as fermions; it has four components subdivided into two two-component objects, representing particles and antiparticles, the Weyl spinors, classified as *right-handed* or *left-handed*, right and left being the *helicities*. γ^μ are the 4×4 Dirac γ-matrices. $\mathcal{A}^C_\mu$, instead, are the gluon fields, and the index C runs from 1 to 8, since there can be eight types of gluons. t^C_{ab} are eight 3×3 matrices, and are the generators of the $\text{SU}(3)_C$ group; they specify the effect of the interaction of a gluon with a quark, i.e. how the gluon causes the rotation of the quark color in the $\text{SU}(3)_C$ space. $F^A_{\mu\nu}$ is a field tensor given by $F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu$, where f_{ABC} are the structure constants of the $\text{SU}(3)_C$ group, $\left[t^A, t^B\right] = i f_{ABC} t^C$.

The quantity g_s is the QCD coupling constant, and it regulates the strength of the interaction between quarks and gluons. The Feynman rules of QCD involve a quark-antiquark-gluon vertex ($\propto g_s$), a 3-gluon vertex ($\propto g_s$), and a 4-gluon vertex ($\propto g_s^2$). A vertex is a point in space-time where at least three particles converge. The QCD vertices are depicted in figure 1.2.

In most experimental and phenomenological contexts, it is customary to consider α_s as the *strong coupling constant*, defined as:

$$\alpha_s = \frac{g_s^2}{4\pi} \tag{1.2}$$

In fact, it provides a more convenient scale for comparing the strength of the strong interaction at different energies. Besides quark masses, which have electroweak origin, α_s is the only free parameter of the QCD lagrangian, so precisely measuring it is a key requirement.

[1] In a non-abelian transformation, applying two successive transformations in different orders gives different results. In other words, the gauge symmetry group is non-commutative, meaning that the group elements do not commute.

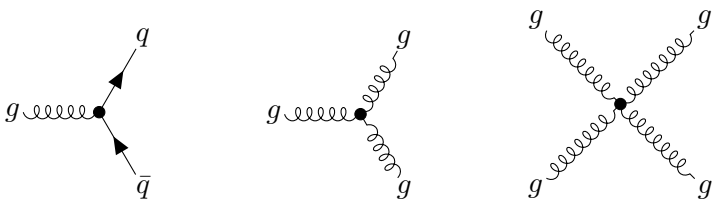

Figure 1.2: Possible QCD vertices in Feynman diagrams notation: quark-antiquark-gluon vertex, 3-gluon vertex, 4-gluon vertex.

Despite being called constants, all couplings of the fundamental interactions are a function of the scale of the momentum transfer Q of a given process. This occurrence is due to the influence of spontaneously created particle-antiparticle pairs, an effect that indeed depends on the energy scale, and is explained by the procedure known as *renormalization*. It follows that $\alpha_s\left(\left|Q^2\right|\right)$ indicates the effective intensity of the strong interaction at play in the process. α_s and the other energy-dependent couplings are thus named *running coupling constants*. For QCD, it holds [14, 15]:

$$\alpha_s\left(\left|Q^2\right|\right) = \frac{\alpha_s\left(\mu_R^2\right)}{1 + \frac{\alpha_s(\mu_R^2)}{12\pi}\left(11n_C - 2n_f\right)\ln\left(\frac{|Q^2|}{\mu_R^2}\right)} \qquad \left(\left|Q^2\right| = -Q^2 \gg \mu_R^2\right) \tag{1.3}$$

where n_C is the number of colors (3 in the SM), and n_f is the number of flavors effectively contributing, namely those which mass satisfies $m_f < |Q|$ (in total there are 6 possible flavors in the SM). The underlying mechanisms are quark-antiquark and gluon-gluon pairs production and annihilation. The gluon-gluon pairs dominate the scene, and produce an *antiscreening* of the charge, i.e. the color, causing α_s to decrease with increasing $\left|Q^2\right|$. This is valid in general for any theory for which $11n_C > 2n_f$. The peculiar behavior of $\alpha_s\left(\left|Q^2\right|\right)$ leads to two consequences, that are the two faces of the same coin. The first one is *asymptotic freedom*: at short ranges, i.e. at large $\left|Q^2\right|$ values, the strong force weakens; in such high energy regimes, quarks and gluons are free, no more bound inside hadrons; this configuration allows to use Feynman diagrams to study QCD systems. The second one is *confinement*: at long distances, or equivalently at daily energy scales, quarks can only exist in the form of colorless combinations.

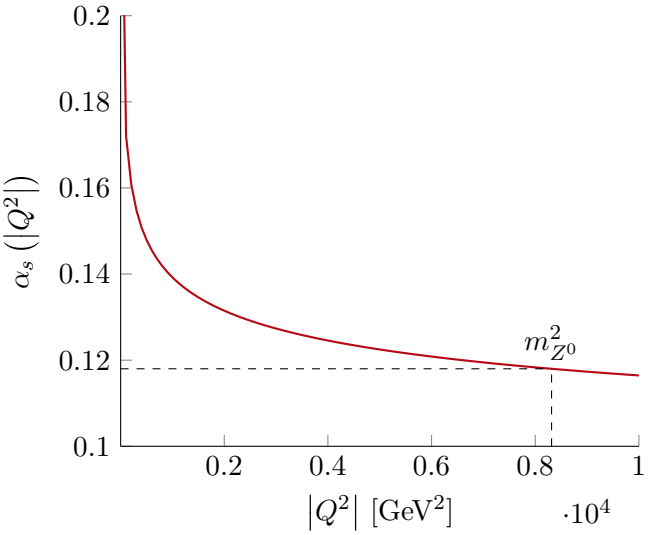

Figure 1.3: Behavior of $\alpha_s\left(\left|Q^2\right|\right)$ as a function of $\left|Q^2\right|$ at the Z^0 mass pole. To compute it, $n_f = 5$ was set, since there are five active quarks at m_{Z^0} energy scale ($m_t > m_{Z^0}$). It results $\Lambda_{QCD}\left(\mu_R = m_{Z^0}, n_f = 5\right) \simeq 88\,\text{MeV}$.

In equation 1.3, the parameter μ_R is the *renormalization scale*. It is an unphysical reference value with respect to which the predictions for observables are expressed. The choice is arbitrary, provided that $\alpha_s(\mu_R^2)$ is small enough to justify a perturbation expansion of $\alpha_s\left(\left|Q^2\right|\right)$, and so it must be $\alpha_s(\mu_R^2) < 1$.

To be more quantitative, a new variable, Λ_{QCD}, with the dimension of a mass, can be introduced:

$$\ln \Lambda^2_{QCD} = \ln \mu_R^2 - \frac{12\pi}{(11 n_C - 2 n_f)\, \alpha_s\left(\mu_R^2\right)} \tag{1.4}$$

so that the strong coupling constant becomes:

$$\alpha_s\left(\left|Q^2\right|\right) = \frac{12\pi}{(11 n_C - 2 n_f) \ln\left(\frac{|Q^2|}{\Lambda^2_{QCD}}\right)} \qquad \left(\left|Q^2\right| \gg \Lambda^2_{QCD}\right) \tag{1.5}$$

The constant Λ_{QCD} approximately lies in the range $100\,\text{MeV} < \Lambda_{QCD} < 500\,\text{MeV}$. Actually, different values for Λ_{QCD} can be defined depending on the number of active quarks considered.

The common scale to determine α_s is the Z^0 boson mass, $\mu_R^2 = m^2_{Z^0}$, at which the current world average [16] gives $\alpha_s\left(m^2_{Z^0}\right) = 0.1180 \pm 0.0009$. The behavior of $\alpha_s\left(\left|Q^2\right|\right)$ as a function of $\left|Q^2\right|$ at the Z^0 mass pole can be found in figure 1.3.

Recalling the quick recap on perturbation theory done in section 1.1, it can be now understood that also QCD is studied perturbatively, with perturbation parameter α_s. At high energies, where α_s is small, physical quantities can be expanded in powers of α_s: $\sigma = \sigma_0 + \alpha_s\sigma_1 + \alpha_s^2\sigma_2 + \mathcal{O}\left(\alpha_s^3\right)$. Here, σ represents a physical observable, such as a cross section, and σ_0, σ_1, σ_2 are the coefficients calculated using Feynman diagrams. The order of a Feynman diagram is determined by the number of vertices it contains: *Leading Order* (LO) diagrams (also known as *tree-level* diagrams) are the ones with the fewest vertices, *Next-to-Leading Order* (NLO) diagrams have one additional vertex, *Next-to-Next-to-Leading Order* (NNLO) diagrams have two vertices more than LO ones, and all higher order diagrams have even more vertices, corresponding to higher powers of α_s. However, perturbation theory is not valid at low energies, where α_s becomes large, leading to non-perturbative effects. In this regime, other techniques are used, such as lattice QCD [17, 18], a non-perturbative approach to solve QCD at low energies, at which analytic perturbative solutions are difficult or even impossible to obtain due to the nonlinear nature of the strong force and the large α_s value. It is a lattice gauge theory formulated on a grid of points in space-time. In the limit of an infinitely large size of the lattice, and of infinitesimally close sites, the continuum QCD is recovered. A momentum cut-off of the order of $1/a$ is introduced, where a is the lattice dimension, which regularizes the theory, and makes it mathematically well-defined, allowing to develop predictions also in low energy regions.

1.3 Deep Inelastic Scattering and Parton Distribution Functions

The features of QCD cross sections in processes with initial-state hadrons can be described within the *Deep Inelastic Scattering* (DIS) theory [19,20]. Strictly speaking, it is a class of processes used to probe the insides of hadrons using electrons, muons and neutrinos, but its jargon applies also in the case of hadronic collisions, like the *pp* ones studied in this thesis work. *Inelastic* here means that the structure of the initial hadron is destroyed in the scattering, and its internal structure can be deduced from the remnants of the process. To illustrate the theory, it is convenient to take as an example the scattering $l(k) + N(P) \to l'(k') + X(P')$, where typical initial and final states are $l, l' = e^{\pm}$, $\mu^{\pm}$, $\nu_\mu, \bar{\nu}_\mu$, $\nu_e, \bar{\nu}_e$, N is a nucleon, and X symbolizes the recoiling hadronic system. The transferred four-momentum from l to p carried by the vector boson γ, Z^0 or $W^{\pm}$ is usually named q. Given these, Q^2 is related to the exchanged four-momentum, $Q^2 = -q^2 = -(k - k')^2$, and is also called *virtuality*. Several invariant kinematic variables can be written, in particular the so called *Bjorken scaling variable* x and the *inelasticity parameter* y, which in the nucleon's rest frame read:

$$x = \frac{Q^2}{2q \cdot P} \qquad y = \frac{q \cdot P}{k \cdot P} \tag{1.6}$$

x represents the fraction of the initial nucleon's four-momentum carried by the *parton* (the collective name for a quark, antiquark or gluon) involved in the scattering, which is then xP, and y is the fraction

of the lepton's energy lost. The differential cross section for a DIS process of a lepton colliding with a proton is expressed in terms of two *structure functions* relative to the proton, $F_1\left(x,Q^2\right)$ and $F_2\left(x,Q^2\right)$. In the case of a photon exchange, the cross section is approximated by:

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{xQ^2}\left[\left(1-y-\frac{M^2x^2y^2}{Q^2}\right)F_2\left(x,Q^2\right)+\left(1-\frac{y}{2}\right)^2 2xF_1\left(x,Q^2\right)\right] \tag{1.7}$$

where α is the fine-structure constant - the fundamental constant characterizing the strength of the electromagnetic interaction between elementary charged particles - and M is the mass of the proton. Particularly, $F_1\left(x,Q^2\right)$ is linked to the transverse polarization of the photon, and represents the probability distribution of finding a parton carrying a fraction x of the proton's momentum when the photon is transversely polarized. Instead, $F_2\left(x,Q^2\right)$ includes contributions from both the longitudinal and transverse polarizations of the photon, and is more directly related to the total probability distribution of partons within the proton. Consequently, $F_1\left(x,Q^2\right)$ is proportional to the distribution of the transverse momentum of the partons, while $F_2\left(x,Q^2\right)$ provides information on the overall parton distribution, summing contributions from both transverse and longitudinal components.

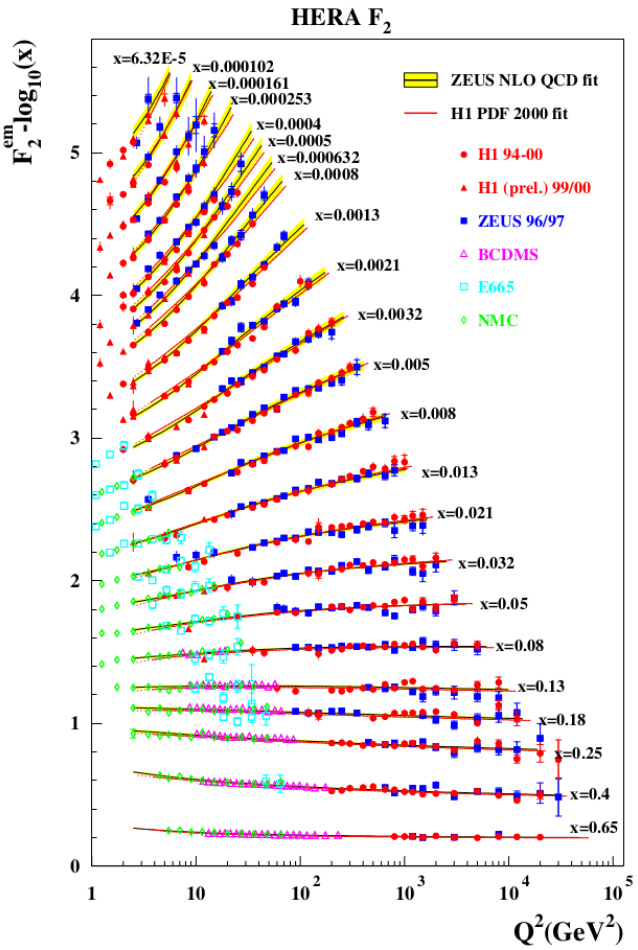

Figure 1.4: The proton structure function, $F_2\left(x,Q^2\right)$, as a function of the virtuality of the exchanged photon at different values of the Bjorken scaling variable x in $e^{\pm}p$ collisions, measured at HERA and fixed target experiments [21].

It is interesting to look explicitly at the dependence of $F_2\left(x,Q^2\right)$ on Q^2 and x (figure 1.4). The plot can be read in two ways, with fixed Q^2 or x. At fixed virtuality, the lower x, the higher $F_2\left(x,Q^2\right)$ gets. The naive interpretation is that at small x the number of visible partons is large, as a small x means that there are more available partons among which to split the proton's momentum. On the other hand, at fixed high x, the number of partons is flat with respect to Q^2, whereas at fixed low x, $F_2\left(x,Q^2\right)$ decreases with decreasing Q^2, i.e. the higher the virtuality, the more visible partons there are. Without going into much detail, the latter fact occurs because of *gluon splitting* and *quark splitting* processes, which happen inside the proton at small x values.

This whole discussion can be generalized to *p-p* scatterings. How can the momentum distributions of the partons inside the protons be known? The parameters describing the internal structure of the proton must be measured experimentally, and the framework to model it is given by the *Parton Distribution Functions* (PDFs) $f_i(x)$, which quantify the probability that the i-th parton in the proton carries a momentum between x and $x + dx$. Subsequently, at Leading Order, the structure functions of the proton can be rewritten as:

$$F_1\left(x, Q^2\right) = \frac{1}{2}\sum_i e_i^2 f_i(x) \qquad F_2\left(x, Q^2\right) = \sum_i x e_i^2 f_i(x) \tag{1.8}$$

where e_i is the electric charge of the parton i in units of the proton charge. Also the cross section for the photon exchange in equation 1.7 can be reformulated considering that the lepton interacts not with the proton as a whole, but with one if its constituting partons:

$$\frac{d^2\sigma}{dxdQ^2} = \sum_i \int_0^1 dx f_i(x) \frac{d^2\sigma}{dxdQ^2}\bigg|_{lq\to lq} \tag{1.9}$$

The number N_i of partons i inside the proton is then:

$$N_i = \int_0^1 fi(x)dx \tag{1.10}$$

and momentum conservation imposes:

$$\sum_i \int_0^1 x f_i(x)dx = 1 \tag{1.11}$$

A compelling consequence of DIS theory is that the partons present inside the protons are not only its valence quarks, two u and one d quarks, but there are also other quarks, antiquarks and gluons, which all participate to the momentum distribution. *Valence quarks* are the building blocks that determine the quantum numbers (such as charge and flavor) and identity of a hadron. Then, the *sea quarks* are the virtual quark-antiquark pairs that are created and annihilated inside the hadrons due to the strong force, carried by the gluons. They contribute to the hadron internal structure and interactions, but do not change its defining quantum properties. As a matter of fact, experimental results show that in a proton:

$$\sum_i \int_0^1 x f(x)dx \approx 0.5 \tag{1.12}$$

which means that the neutral partons, the gluons, carry approximately 50% of the proton's momentum. In the following plots 1.5, the distributions of the different partons inside the proton are shown as a function of x for a specific type of PDF, CT10NNLO [22]. As expected, the sea quarks increase in number as x decreases: more partons are visible, and so the fraction of momentum carried by each parton is small. There exist many PDF sets, distinguished by the Collaboration producing them, which applies its own methodologies and parametrization methods, e.g. traditional polynomial-like functional forms for CT10NNLO PDF, chosen based on theoretical expectations, and fitted to experimental data. Uncertainty estimation, data samples, fit procedures, and order of QCD calculations are also peculiar of each PDF set. An important parameter for a PDF is the α_s value used for the computations.

The evolution of the PDFs with respect to Q^2 is governed by the *Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations* [23–26]. They are integro-differential equations that account for the probability of a parton splitting into other partons as the energy scale increases, which affects the distribution of partons inside a hadron. DGLAP equations can be written as:

$$\frac{\partial\, q_i\left(x, Q^2\right)}{\partial \ln Q^2} = \frac{\alpha_s\left(Q^2\right)}{2\pi}\sum_j \int_x^1 \frac{dz}{z} P_{ji}(z)\, q_j\left(\frac{x}{z}, Q^2\right) \tag{1.13}$$

$$\frac{\partial\, g\left(x, Q^2\right)}{\partial \ln Q^2} = \frac{\alpha_s\left(Q^2\right)}{2\pi}\left[\sum_j \int_x^1 \frac{dz}{z} P_{jg}(z)\, q_j\left(\frac{x}{z}, Q^2\right) + \int_x^1 \frac{dz}{z} P_{gg}(z)\, g\left(\frac{x}{z}, Q^2\right)\right] \tag{1.14}$$

where $q_i\left(\frac{x}{z}, Q^2\right)$ is the distribution function of a quark of flavor i, $g\left(\frac{x}{z}, Q^2\right)$ is the distribution function of a gluon, and the $P_{ji}(z)$ are the *splitting functions*, which represent the probability of a parton of type i splitting into a parton of type j with momentum fraction z. They are specific to the kind of parton splitting, and are calculated in perturbative QCD: $P_{qq}(z)$ is the probability of a quark emitting a quark and a gluon ($q \to qg$); $P_{qg}(z)$ is the probability of a gluon splitting into a quark-antiquark pair ($g \to q\bar{q}$); $P_{gg}(z)$ is the probability of a gluon splitting into two gluons ($g \to gg$); $P_{gq}(z)$ is the probability of a quark emitting a quark and a gluon, and is different from $P_{qq}(z)$, as $P_{gq}(z)$ considers that the quark firstly evolves into a gluon, which in turn evolves back into a quark. To solve numerically the equations, a reference scale Q_0^2 must be set, and then the evolution of the PDFs is determined from Q_0^2 to higher scales. This evolution considers that the number of partons increases as the energy scale increases, a direct consequence of the already discussed asymptotic freedom property of QCD. The DGLAP equations are essential for making predictions, and for testing those predictions by comparison with experimental measurements at different energies.

Therefore, at a pp collider, the cross section of a collision producing the final state X reads:

$$\sigma_{pp\to X} = \sum_{a,b} \int_0^1 dx_a\, dx_b\, f_a^A\left(x_a, Q^2\right) f_b^B\left(x_b, Q^2\right) \hat{\sigma}_{ab\to X} \tag{1.15}$$

where A and B denote the two colliding protons, a and b their respective partons involved in the process, and $\hat{\sigma}_{ab\to X}$ is the partonic cross section that can be represented with a Feynman diagram. Equation 1.15 proves that PDFs enter in many theoretical predictions, since they are a required input for the computation of cross sections. Hence, it is fundamental to measure them with high precision.

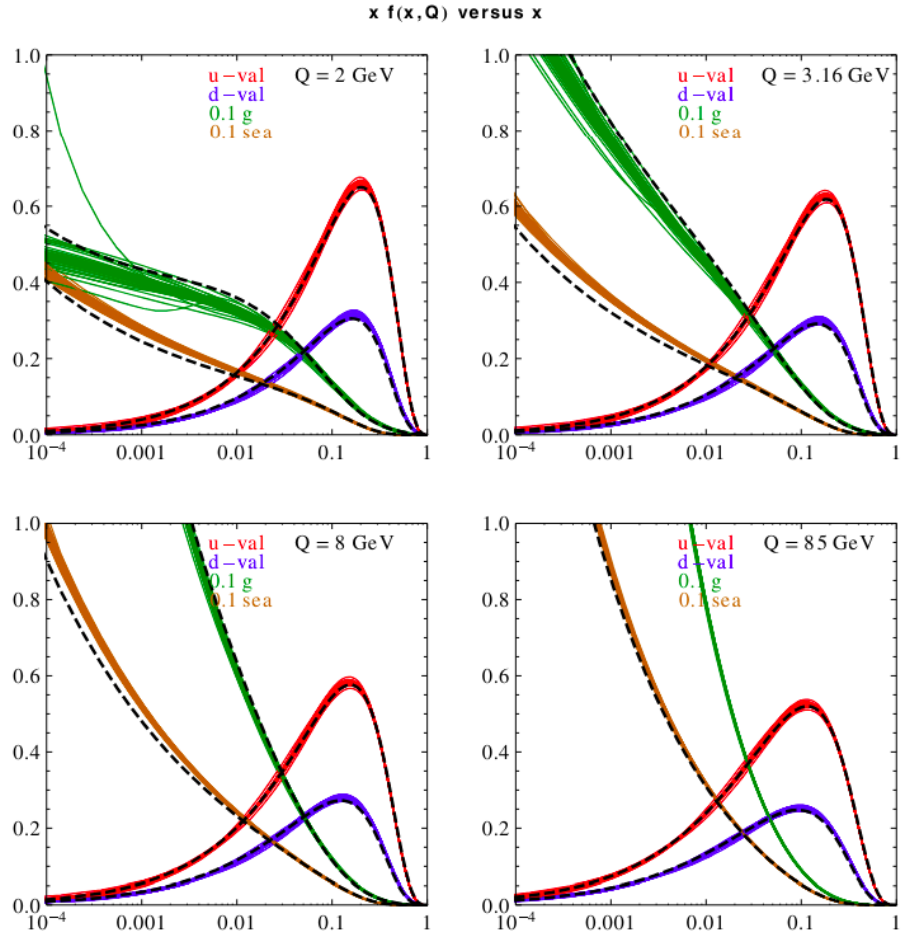

Figure 1.5: CT10NNLO Parton Distribution Functions [22]. Each graph shows $x\,u_{valence} = x\,(u - \bar{u})$, $x\,d_{valence} = x\,(d - \bar{d})$, $0.1\,x\,g$ and $0.1\,x\,q_{sea}$ as functions of x for four different fixed values of Q. The quark sea contribution is $q_{sea} = 2\,(\bar{d} + \bar{u} + \bar{s})$. Here u, d, $\bar{u}$, $\bar{d}$, $\bar{s}$ represent the number of up, down, anti-up, anti-down and anti-strange quarks, respectively.

1.4 α_s measurement techniques

At present, α_s is the least precisely measured fundamental constant. It is crucial to improve its measurement precision: theoretically, since α_s enters every calculation about strongly generated particles, it limits the prediction ability of the models for production and decay processes; experimentally, it prevents effects of new physics to be visible, as it is a non-negligible contribution to the total uncertainty of a measurement of a strongly-induced process. In other words, lowering α_s uncertainty can answer some cardinal open questions: could unknown interactions beyond the SM be modifying α_s in certain processes or at certain energies? Could all the forces be of equal strength at very high energy, suggesting a common origin?

The strong coupling constant is not a physical observable in itself, as it is a variable defined in the context of perturbation theory, which only enters the predictions for experimentally measured observables. Therefore, it can not be extracted directly, but it must be inferred measuring such quantities, and is affected by experimental and theoretical uncertainties. The incomplete knowledge of α_s propagates into uncertainties in numerous precision tests of the SM.

The most precise single determination of α_s to date, compatible with the world average, was obtained by the ATLAS Collaboration in 2023 [27]. In their work, ATLAS physicists studied the production and subsequent decay into a $\mu^+\mu^-$ or e^+e^- pair of the Z^0 boson in pp collisions data. Namely, they analyzed the emission of gluons from the initial quarks, that causes the Z^0 boson to recoil in the plane transverse to the beam axis; the size of this "kick" is measured as transverse momentum and depends on α_s. Since the Z^0 boson decays instantly after being produced, its transverse momentum is measured via its decay products. Other values of α_s have been extracted from a plethora of processes at different Q^2 values. The sub-fields considered to calculate the world average value of $\alpha_s\left(m^2_{Z^0}\right)$, summarized in figure 1.6, are [13]: hadronic τ decays and low Q^2 continuum, heavy quarkonia decays, PDF fits, hadronic final states of e^+e^- annihilations, observables from hadron-induced collisions, electroweak precision fits, lattice QCD.

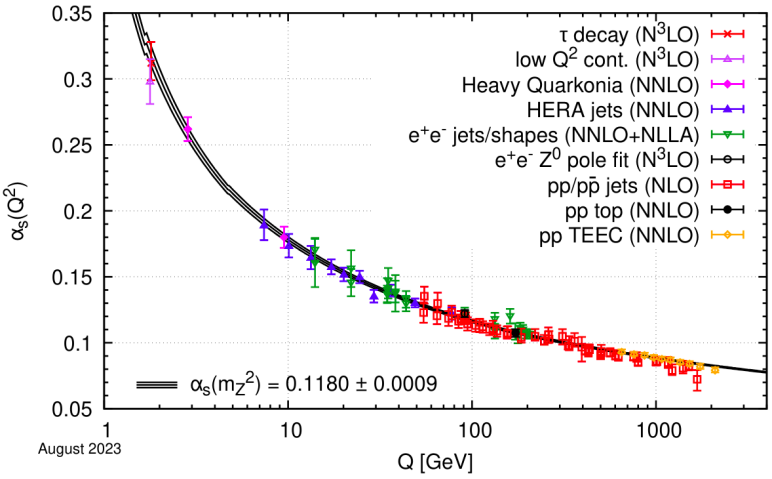

Figure 1.6: Contributions to the calculation of the world average value of $\alpha_s\left(m^2_{Z^0}\right)$ [13].

1.5 LHCb forward measurements and impact on α_s

As reviewed in chapter 2, the peculiarity of the LHCb experiment is being a spectrometer built in the forward region of a pp interaction, spanning a phase space of $2 < \eta < 5$. This has a deep consequence: the collisions producing massive vector bosons and jets at LHCb involve one parton with high Bjorken-x and one parton with low x. This happens because the collisions must be boosted into the forward region to be in the LHCb acceptance, and the rapidity y of a directly produced particle of mass M is related to the x values of the colliding partons (here labelled 1 and 2) by $x_{1,2} \simeq \frac{M}{\sqrt{s}} e^{\pm y}$, where s is the center of mass energy of the pp collision. The resulting regions probed by

LHCb are visible in figure 1.7: at high x values, close to unity, LHCb produces measurements which are comparable with other experiments, while at low x and high Q^2 values, it covers an unexplored kinematic range, complementary to other experiments.

With its unique features, LHCb is able to study the PDFs discussed in section 1.3, that are a fundamental input for all collider experiments measurements and for theoretical calculations, and that must be determined experimentally. LHCb tests them not only at high x, where they are already constrained by results from HERA and other experiments, but most importantly at low x and high Q^2. The behavior of the DGLAP equations that govern the evolution of PDFs is also investigated in this regime, and consequently the internal proton structure is probed. Figure 1.8 shows the impact of the LHCb experiment in testing PDFs: the inputs to the NNPDF4.0 PDF set for its development are reported, where the LHCb contribution is included in the *Collider gauge boson production* pink data points, which at low x and high Q^2 are the only available measurements.

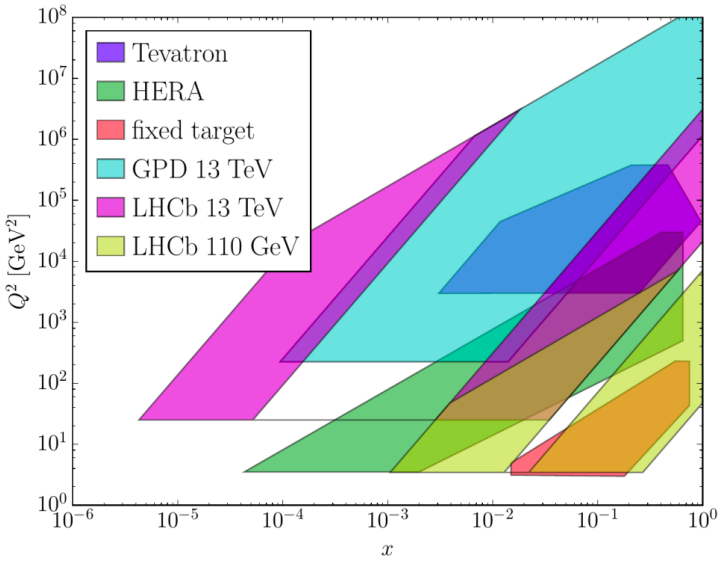

Figure 1.7: Experimental kinematic regions covered by the LHCb detector at 13 TeV (in bright pink), complementary to other relevant experiments [28].

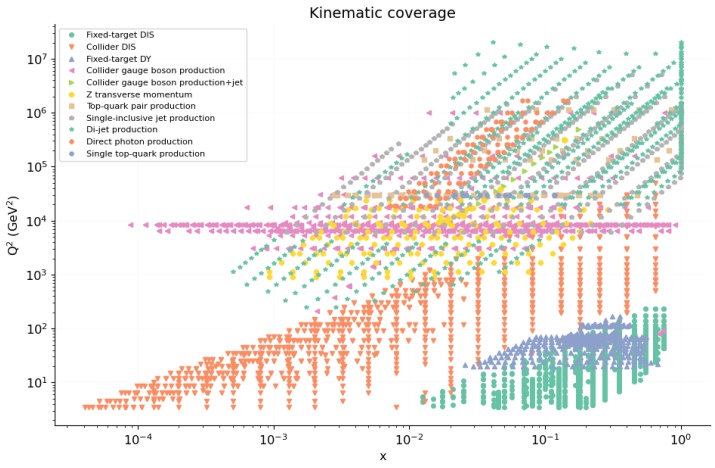

Figure 1.8: NNPDF4.0 PDF set kinematic coverage. The dataset is described at length in [29] and includes several processes and experiments. LHCb contributes in the pink *Collider gauge boson production* data points.

For these reasons, the kinematic complementarity of LHCb is a significant opportunity for electroweak measurements, and also for new QCD studies like the work presented in this thesis. Not only measuring α_s in a new x-Q^2 regime could test the PDFs in that region, but also combining an LHCb α_s estimate with other complementary measurements could lower its global uncertainty with respect to the uncertainty of single measurements.

Chapter 2

The LHCb experiment

The Large Hadron Collider beauty (LHCb) experiment is one of the four largest detectors at the LHC accelerator at CERN, Geneva. It was primarily designed to search for indirect evidence of new CP-violation sources beyond the SM and rare decays of beauty and charm hadrons, such as B_d, B_s and D mesons. A CP-violating element is already present in the SM, in the form of a single complex phase in the weak sector appearing in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which describes the quark mixing in flavor-changing weak interactions. However, this phase is not sufficient to explain the amount of matter in the universe. With much more precision, the effect of an extra source of CP violation might be seen in heavy-flavor physics. LHCb looks for even the slightest deviation from the SM predictions, which could address the presence of new phenomena. Over the first years of operation, LHCb exhibited great capability not only in performing measurements in heavy-flavor physics, but also in many other domains, including electroweak, heavy-ion, fixed target, and hadronic jets physics.

The LHCb detector has been designed to meet specific requirements, particularly those related to the tracking system. Given that pp collisions in the TeV center of mass energy levels generate a lot of tracks per collision, leading to a high track density, the detector must have high granularity systems to accurately reconstruct tracks and photons, thereby reducing the rate of fake signals. Moreover, since the momentum distribution of forward tracks is relatively soft, it is essential to minimize the amount of material in the active tracking volume to reduce multiple scattering and secondary interactions. Lastly, the trigger system must efficiently reconstruct tracks within the allowed system latency.

LHCb started taking data in 2010, and worked during the two first LHC runs, Run 1 (2010-2012) and Run 2 (2015-2018). In Run 1 the center of mass energy of the p-p collisions was initially $\sqrt{s} = 7\,\text{TeV}$, and then it was increased to $\sqrt{s} = 8\,\text{TeV}$, allowing LHCb to collect $3\,\text{fb}^{-1}$ of integrated luminosity. In Run 2, $\sqrt{s} = 13\,\text{TeV}$ was reached, and the integrated luminosity almost doubled, getting to $5.7\,\text{fb}^{-1}$. Also, in Run 1 and Run 2 about $30\,\text{nb}^{-1}$ of lead-lead and p-lead collisions, and about $200\,\text{nb}^{-1}$ of fixed target data were collected. LHCb was originally designed to take data at a maximum instantaneous luminosity $\mathcal{L} = 2 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$ to keep the average number of visible primary pp interactions (the so called *pile-up*) close to 1. Nonetheless, LHCb has been successfully operated for most of Run 1 and Run 2 at $\mathcal{L} \sim 4 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$, proving to be capable to produce excellent physics results at higher luminosity and with a larger pile-up.

Despite the successes achieved by LHCb in its wide physics program, the precision of many key flavor physics observables in Runs 1 and 2 was statistically limited. Much larger data sets were demanded to probe the SM at the precision level achieved by theoretical calculations, and to obtain enough sensitivity to observe possible new physics effects. Hence, a first upgrade of the experiment was implemented during the LHC Long Shutdown 2 (2019-2021), which is expected to function in the ongoing Run 3 (2022-2025) and in Run 4 (2029-2032), the first High Luminosity (HL) LHC run. Then, the proposed second upgrade will perform during Run 5 (2035-2038) and Run 6 (2040-2041). The long term LHC schedule can be found here. The original LHCb configuration is described in section 2.1, while Upgrade Ia and future Upgrades Ib and II are outlined in sections 2.2 and 2.3.

2.1 The LHCb detector: Run 1 and Run 2 configuration

LHCb [30] is located at the Intersection Point 8 of the LHC ring, previously dedicated to the DELPHI experiment. Some adjustments permitted to adapt the cavern, using at maximum the existing facilities. LHCb is a single-arm forward spectrometer almost 21 m long, about 10 m high and 6 m wide. The chosen coordinate system has the origin at the nominal pp interaction point, the z axis along the beam pointing towards the muon system, the y axis pointing vertically upwards, and the x axis defining a right-handed system. Most of the subdetectors, with the exception of vertex and Cherenkov detectors, are split into two mechanically independent halves, which can be opened for maintenance and to guarantee access to the beam pipe. The access side or Side A is at $x > 0$ and the cryogenic side or Side C is at $x < 0$. The experimental angular coverage is of $\sim [10, 300]$ mrad in the horizontal xz plane, and $\sim [10, 250]$ mrad in the vertical xy plane. This results in a pseudorapidity coverage in the range $2 < \eta < 5$. The *pseudorapidity* η is a widespread quantity in particle physics used to describe the angle of a particle relative to the colliding beam axis, defined as:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.1}$$

where θ is the angle between the particle three-momentum and the positive direction of the beam axis z. The LHCb geometry is peculiar, and is justified by the fact that at high energies the main mechanism for heavy-quark production is gluon-gluon fusion, which generates $b\bar{b}$ and $c\bar{c}$ highly correlated pairs with an angular distribution peaked at small angles with respect to the beam line. Therefore, b- and c-hadrons are produced mainly in the same forward or backward cone, in a relatively small solid angle.

An additional peculiarity of LHCb is that most subdetectors, especially near the interaction point of the proton beams, are built in halves that are separately movable in the horizontal plane. This has a twofold advantage: it simplifies the assembly and maintenance procedures, and it allows to place the detectors very close to the beams only during the actual data-taking; pulling them away from the beam reduces the radiation damage and slows the aging of the detectors.

A side view of the initial LHCb configuration is schematized in figure 2.1. In the sections below, a description of the detector components will be given.

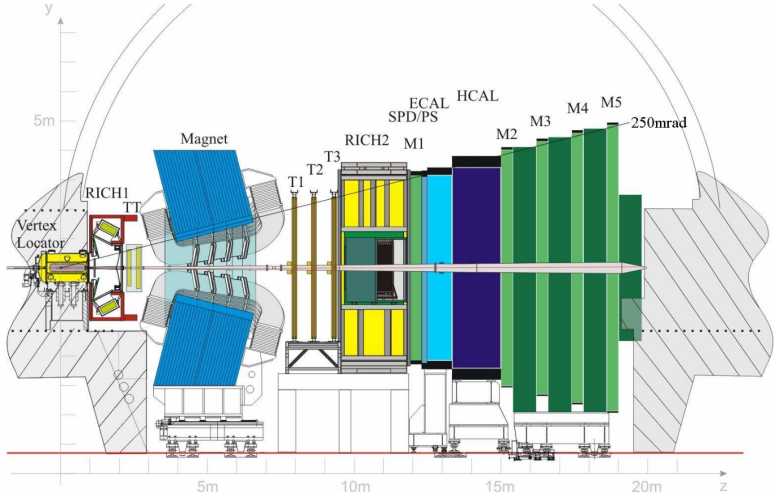

Figure 2.1: Side view of the LHCb Run 1 and Run 2 layout [30].

2.1.1 The magnet

In LHCb a warm dipole magnet [31] is used to measure the momentum of charged particles, covering the experiment forward angular acceptance of ± 300 mrad horizontally and ± 250 mrad vertically. As depicted in the left side of figure 2.2, the coils are saddle-shaped and placed in a window-frame yoke with sloping poles that match the detector acceptance. There are two identical coils made of pure Al-99.7 hollow conductor, and are assembled mirror-symmetrically. The conductor resistance is below

$28\,\Omega \cdot$ m at 20 °C. Overall, the magnet system measures 11 m horizontally, 8 m vertically, and is 5 m wide, with a mass of 1600 tons. The Magnet Control System monitors the power supply and some operational parameters, e.g. temperatures and mechanical movements, while the fully independent Magnet Safety System (MSS) guarantees the safe operation, enforcing autonomously a discharge of the magnet if its critical parameters are out of range. The nominal current in the conductor is 5.85 kA.

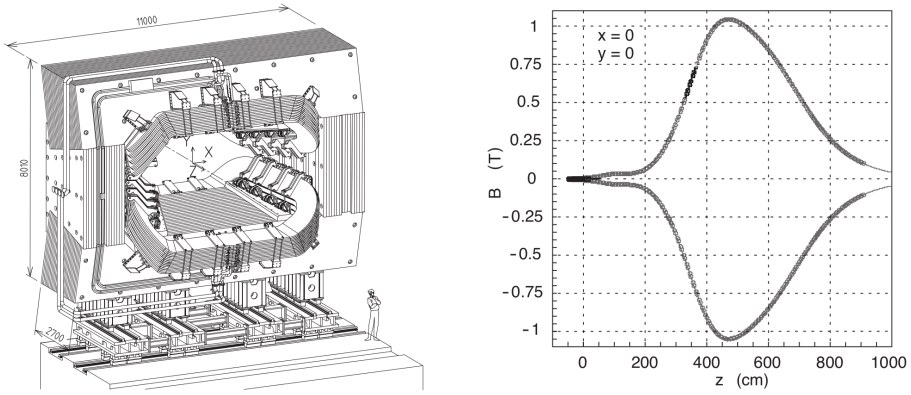

Figure 2.2: Schematic view of the magnet (left) and magnetic field B_y along the z axis (right) [30].

The right plot in figure 2.2 shows the behavior of both polarities of the main B-field component (B_y) along the z direction, resulting from the particular shape of the coils. They satisfy two contrasting requirements: the integrated magnetic field $\int B\,dl$ must be as high as possible in the region between the Vertex Locator (VELO) and the Tracker Turicensis (TT) stations, and at the same time it is needed to be low inside the RICHs envelope (refer to figure 2.1 for the position of these detectors in LHCb). For these reasons, $\int B\,dl = 0.1159\,\mathrm{T}\cdot\mathrm{m}$ in the region between 0 m and 2.5 m, and $\int B\,dl = 3.615\,\mathrm{T}\cdot\mathrm{m}$ in the region between 2.5 m and 7.95 m. The magnetic field peak was measured with a precision of a few millimeters, and the integrated field with a relative precision of about 4×10^{-4}. It is essential to keep the value of the latter very low, to achieve the necessary charged-particle momentum resolution.

2.1.2 The tracking system

The LHCb tracking system, described in more detail in the sections below, comprises the Vertex Locator system (VELO) and four planar tracking stations: the Tracker Turicensis (TT) upstream[1] of the dipole magnet and T1-T3 downstream of the magnet. Both VELO and TT are made of silicon microstrip detectors. In T1-T3, silicon microstrips are placed in the region close to the beam pipe (Inner Tracker, IT), whereas straw-tubes are employed in the outer region (Outer Tracker, OT). The TT and the IT were developed in a common project called the Silicon Tracker (ST).

2.1.2.1 The Vertex Locator

The first detector system which particles intercept is the VErtex LOcator (VELO) [32]. It measures track coordinates close to the interaction point with high precision, identifying the displaced secondary vertices, which are distinctive of b- and c-hadron decays. It detects particles in the angular range $1.6 < \eta < 4.9$, and emerging from primary interaction vertices in $|z| < 10.6$ cm. Figure 2.3 shows that VELO consists of silicon modules placed along the beam axis z, perpendicularly to it. They are built in two retractable and staggered semi-circular halves that can slide in the vertical plane xy, and overlap in the fully closed configuration, where they are at a radial distance from the beams which is smaller than the aperture required by LHC during injection. There are two detector stations located upstream of the VELO sensors, that form the so called pile-up veto system, which is part of the first level of the trigger, as discussed in section 2.1.4. The VELO components are enclosed in a secondary vacuum vessel, separated from the primary beam vacuum by an aluminum sheet, that is very thin to minimize the amount of material traversed by charged particles before reaching the sensors.

[1] Upstream and downstream are referred to the direction of increasing z.

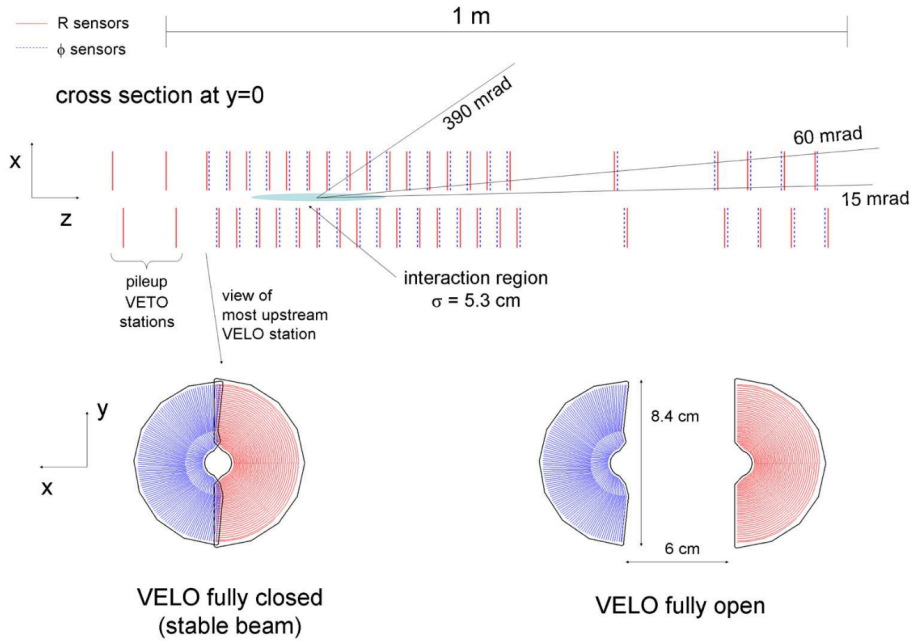

Figure 2.3: Cross section of the VELO silicon sensors in the xz plane, at $y = 0$, in the fully closed position. The front face of the modules is illustrated in both the closed (left) and open (right) positions. The two pile-up veto stations are located upstream of the VELO sensors [30].

The geometry of VELO sensors is cylindrical with $r\phi$ coordinates, where r is the radial direction from the beam pipe in the xy plane, and ϕ is the azimuthal angle in the same plane. This choice enables a faster reconstruction of tracks and vertices in the trigger than a rectilinear geometry. Therefore, two types of sensors are installed in each VELO module: a ϕ-sensor, providing information on the azimuthal coordinate, and a R-sensor, measuring the radial distance from the beam axis. To achieve a three-dimensional measurement, the position of each sensor plane along the z axis is used as the third coordinate.

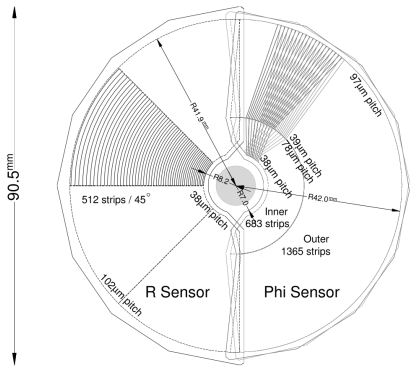

Figure 2.4: Sketch of the $r\phi$ geometry of the VELO ϕ- and R-sensors. Only a portion of the strips is illustrated. In the ϕ-sensor, the strips on two adjacent modules are indicated, to highlight the stereo angle. The radius of the R-sensor is slightly larger, but the sensitive area is identical [30].

The sensors are made by silicon microstrips with a thickness of 300 µm. To achieve high radiation tolerance, the n^+n strip technology was adopted. For both R- and ϕ-sensors, the minimum pitch is at the inner radius to optimize the vertex resolution. In figure 2.4 the layout of the strips in the two sensors is illustrated. In the R-sensor, the implanted strips are concentric semi-circles with center at the nominal LHC beam position. To minimize the occupancy and reduce the strip capacitance, each strip is subdivided into four 45° sectors. To ensure that measurements along the track contribute to the impact parameter precision with roughly equal weight, the pitch at the innermost radius of 8.2 mm is 38 µm, increasing linearly to 101.6 µm at the outer radius of 41.9 mm. Instead, the *phi*-sensor is

sectioned into an inner and an outer region. The outer region starts at a radius of 17.25 mm, and its pitch (39.3 µm) is set to be half that of the inner region (78.3 µm), which ends at the same radius. A skew of the strips is introduced to improve pattern recognition. At 8.2 mm from the beam the inner strips have an angle of $\sim 20°$ to the radial, while the outer strips make an angle of approximately $\sim 10°$ to the radial at 17 mm. The skew of inner and outer sections is reversed, giving the strips a distinctive *dog-leg* shape. Two adjacent modules have their ϕ-sensors with opposite skew, and through this stereo view the ability to distinguish ghost hits from true hits is ensured.

2.1.2.2 The Silicon Tracker

The Silicon Tracker (ST) includes the Tracker Turicensis and the three Inner Tracker stations. They all use single-sided p^+n silicon microstrip sensors with a pitch of about 200 µm. Each of the four ST stations has four layers, where the strips are arranged vertically in the first and the last layer, while in the second and third layer the strips are tilted by a stereo angle of $-5°$ and $5°$, respectively. A trade-off between maintaining the strip occupancies at a level of at maximum a few percent, and minimizing the number of readout channels, brought to different readout strip lengths for different regions of the devices. The ST detectors have a single-hit spatial resolution of 50 µm, and their readout electronics, power supply, and control and monitor systems are in common. The four layers of each ST station are mounted in a light-tight, thermally and electrically insulated detector volume, in which a temperature below 5 °C is kept. The volume is continuously flushed with nitrogen to avoid condensation on the cold surfaces.

Tracker Turicensis. The Tracker Turicensis (TT) [33], formerly known as Trigger Tracker, is located upstream of the magnet, and covers the full LHCb acceptance area. As illustrated in the left part of figure 2.5, the detection layers are organized into two pairs, (x, u) and (v, x), separated by approximately 27 cm along the beam axis. The building block of a layer is a half module, each covering half the height of the LHCb acceptance, with full modules made by joining two half modules end-to-end. A half module is composed by seven silicon sensors and a stack of readout hybrids. In the region close to the beam pipe, where the particle density is the highest, the sensors are organized into three readout sectors (4-2-1 type modules), and further from the beam they are divided into two readout sectors (4-3 type modules). Adjacent modules within a detection layer are staggered in z and overlapped in x to avoid acceptance gaps, and to facilitate their relative alignment and the track reconstruction algorithms. The advantage of this detector design is that all front-end hybrids and the infrastructure for cooling and module support are located above and below the active area of the detector, outside of the experimental acceptance.

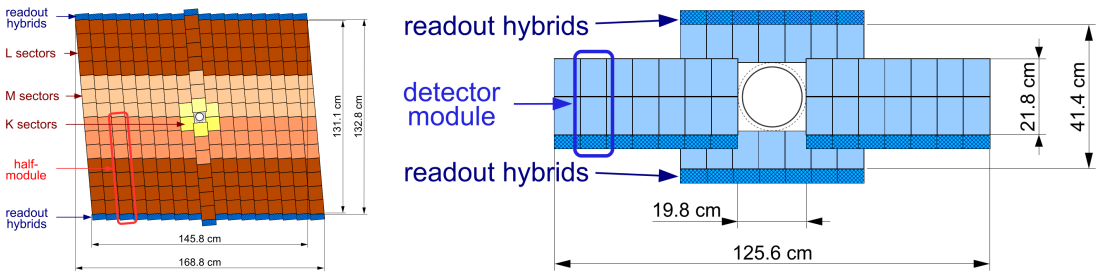

Figure 2.5: Pictorial representation of the third Tracker Turicensis detection layer (left) and the x detection layer in the second Inner Tracker station (right) [30].

Inner Tracker. The Inner Tracker (IT) [34] stations are cross-shaped, and each station consists of four detector boxes arranged around the beam pipe and located in the central part of the three T1-T3 trackers downstream of the LHCb magnet. Each detector box contains four detection layers, and each detection layer consists of seven detector modules. Adjacent modules in a detection layer are staggered by 4 mm in the z direction and overlap by 3 mm in the x direction. Detector modules in the top and bottom boxes placed above and below the beam pipe consist of a single silicon sensor and a readout hybrid, while detector modules in the side boxes placed to the left and right of the beam pipe consist of two silicon sensors and a readout hybrid. The resulting layout and dimensions of one

of the IT detection layers are illustrated in the right part of figure 2.5. Two types of silicon sensors of different thickness, but identical in design, are used in the IT. The thicknesses were chosen to ensure sufficiently high signal-to-noise ratios for each module type while minimizing the material budget of the detector.

2.1.2.3 The Outer Tracker

The Outer Tracker (OT) [35] is the last component of the LHCb tracking system, and it is in the farthest region of the T1-T3 stations away from the beam pipe. It is a drift-time detector designed to track charged particles and measure their momentum over a large acceptance area with excellent momentum resolution, which is crucial for determining the invariant mass of reconstructed b-hadrons with high precision. The OT guarantees high tracking efficiency and low rates of wrongly reconstructed tracks. The OT is an array of gas-tight straw-tube modules with two staggered monolayers of drift-tubes. The counting gas mixture, made of argon (70%) and CO_2 (30%), warrants fast drift times ($< 50\,\text{ns}$), and a drift-coordinate resolution of $200\,\mu\text{m}$. Like the ST, each station has four layers in an x-u-v-x geometry, with vertical x-layers, and u- and v-layers tilted by $\pm 5°$. The outer boundary of the OT covers the experimental acceptance, while the inner cross-shaped boundary is designed to keep occupancies below 10% at a luminosity of $2 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$. The stations are split into retractable halves, each consisting of two independently movable units of two half layers.

2.1.2.4 Track reconstruction

The reconstructed tracks at LHCb, listed below, are classified based on their trajectories in the spectrometer, as shown in figure 2.6. The reconstruction process takes as input initial track candidates, known as track *seeds*, in the VELO region and the T stations where the magnetic field is weak.

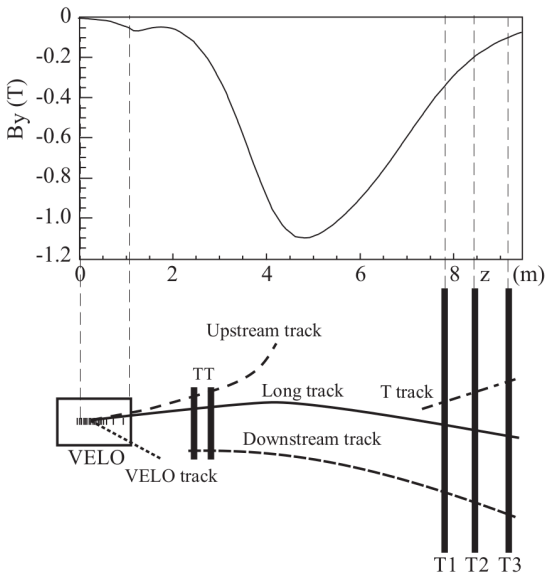

Figure 2.6: A schematic illustration of the track types reconstructed in LHCb: long, upstream, downstream, VELO and T tracks. The main B-field component (B_y) as a function of the z coordinate is also plotted [30].

Long tracks traverse the whole tracking system from the VELO to the T stations. Their momentum is the most precisely determined, therefore they are the most important set of tracks for b-hadron decay reconstruction.

Upstream tracks traverse only the VELO and TT stations. Their momentum is lower, thus they are bent out of the detector acceptance by the magnetic field. However, they manage to pass through the RICH 1, and may generate Cherenkov photons if they have velocities above threshold, so they are used to study backgrounds in the RICH PID algorithm. They are also used for b-hadron decay reconstruction or flavour tagging, even if their momentum resolution is rather poor.

Downstream tracks traverse only the TT and T stations. The most relevant cases are the decay products of K_s^0 and Λ that fall outside the VELO acceptance.

VELO tracks are measured in the VELO only, are generally large angle or backward tracks, and are useful for the Primary Vertex reconstruction.

T tracks are measured in the T stations only, are usually produced in secondary interactions, but are useful for the global pattern recognition in RICH 2.

2.1.3 The particle identification systems

Particle IDentification (PID) is a key requirement for LHCb, as it is essential to discriminate pions, kaons and protons in heavy-hadron decays. In the following, the LHCb detectors whose signals contribute to the PID are described. They are two Ring Imaging CHerenkov (RICH) detectors, a calorimetric system composed of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL), and a muon detection apparatus.

2.1.3.1 The Ring Imaging Cherenkov detectors

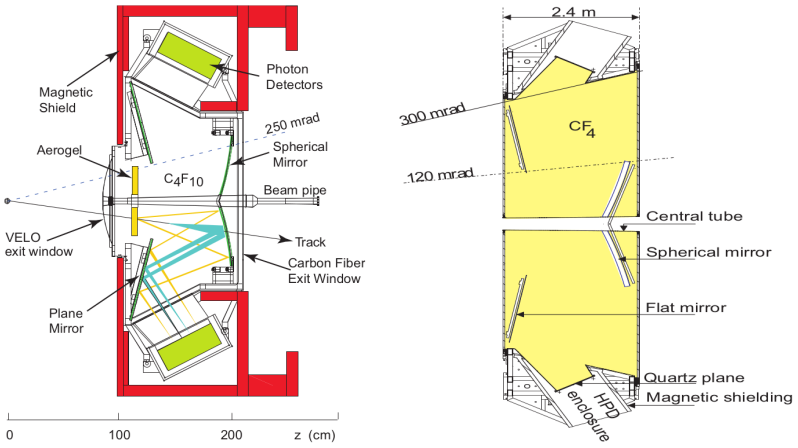

Figure 2.7: Side view schematic layout of the RICH 1 detector (left) and top view schematic layout of the RICH 2 detector (right) [30].

The two Ring Imaging CHerenkov (RICH) detectors [36] together cover the full momentum range of produced charged particles, which at large polar angles is softer, whereas at small polar angles is harder. The RICH 1 [37] is located upstream of the LHCb dipole magnet, between the VELO and the TT, and it covers the low momentum range, $\sim 160\,\text{GeV}$, using aerogel and fluorobutane C_4F_{10} radiators. It occupies the region $990\,\text{mm} < z < 2165\,\text{mm}$. The downstream detector, RICH 2 [38], covers the high momentum range, from 15 GeV up to and beyond 100 GeV, using a CF_4 radiator. Schematic views of the RICH 1 and RICH 2 are reported in figure 2.7. The RICH 1 covers the full LHCb acceptance from $\pm 25\,\text{mrad}$ to $\pm 300\,\text{mrad}$ horizontally and to $\pm 250\,\text{mrad}$ vertically. RICH 2 has a limited angular acceptance of about $\pm 15\,\text{mrad}$ to $\pm 120\,\text{mrad}$ horizontally and to $\pm 100\,\text{mrad}$ vertically. In both detectors the Cherenkov light is focused using a combination of spherical and flat mirrors to reflect the image out of the spectrometer acceptance. In the RICH 1 the optical layout is vertical whereas in RICH 2 is horizontal. Hybrid Photon Detectors (HPDs) are used to detect the Cherenkov photons in the wavelength range of 200-600 nm. The HPDs are protected by iron shields and are placed in MuMetal cylinders to allow operation in magnetic fields up to 50 mT. The RICH system allows to distinguish between final states of otherwise identical topologies, e.g. $B^0_{(s)} \to \pi^+\pi^-$, $K^+\pi^-$, K^+K^-; heavily reduce the combinatorial background in decay modes involving final-state hadrons, such as $B^0_s \to \phi\phi$, where $\phi \to K^+K^-$, that would be too large without PID requirements; perform the flavor tagging of a $B^0_{(s)}$ meson at the production vertex, basing on charged kaon identification from the $b \to c \to s$ decay chain. The information provided by the RICH system is also used to suppress the combinatorial background at the HLT2 level (see section 2.1.4).

2.1.3.2 The calorimeters

The calorimeter system [39] selects hadron, electron and photon candidates for the first trigger level, measuring their transverse energies and positions. The reconstruction with good accuracy of π^0's and prompt photons is essential for flavor tagging and for studying B-mesons decays. The latter require good background rejection and efficiency, which in turn define the general structure and characteristics of the calorimeters. In this sense, adequate resolution and shower separation are crucial.

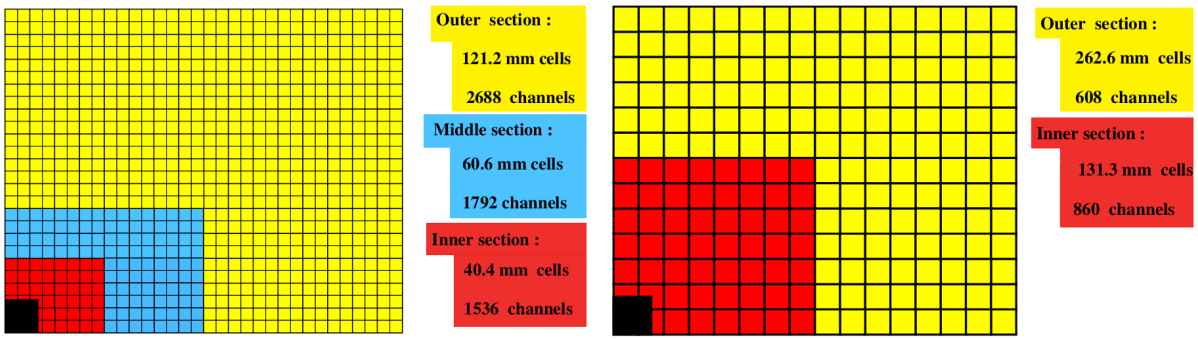

Figure 2.8: Lateral segmentation of one quarter of the detector front face of the SPD/PS and ECAL (left) and the HCAL (right) [30]. The cell dimensions in the left figure refer to the ECAL.

LHCb features a classical calorimeter structure, with an Electromagnetic CALorimeter (ECAL) followed by a Hadronic CALorimeter (HCAL). Identifying high transverse energy (E_T) electrons is crucial for distinguishing b-events from the abundant inelastic pp interactions. To reduce the high background of charged pions, a PreShower detector (PS) is installed, which reconstructs the longitudinal segmentation of the electromagnetic showers. Additionally, a Scintillator Pad Detector (SPD) in front of the PS helps the trigger discarding π^0's with high E_T by selecting charged particles. A thin lead converter is placed between the PS and SPD. Since photon conversions in the upstream spectrometer material dominate the electron trigger, an optimal energy resolution requires the full containment of high energy photon showers, achieved by building the ECAL 25 radiation lengths thick. Instead, due to space constraints, and considering that the trigger requirements on the HCAL resolution do not impose a stringent hadronic shower containment condition, the HCAL thickness is 5.6 interaction lengths. The lateral segmentation of the calorimeters, shown in figure 2.8, accommodates the hit density, which varies by two orders of magnitude over the calorimeter surface. When hit, all calorimeters produce scintillation light, which is then transmitted to PhotoMultiplier Tubes (PMTs) via Wave-Length Shifting (WLS) fibers. For the readout, the SPD/PS cells use MultiAnode PMTs (MAPMT), while the ECAL and HCAL modules use individual phototubes. To maintain a constant E_T scale, the phototubes gain in the ECAL and HCAL is adjusted based on their distance from the beam pipe.

The pad/preshower detector. The pad/preshower (SPD/PS) detector [40] utilizes scintillator pads read by 1.0 mm diameter WLS fibers coupled to MAPMT via clear plastic fibers. This detector comprises a 15 mm lead converter, $2.5X_0$ thick (where X_0 denotes a radiation length), situated between two almost identical high-granularity planes of rectangular scintillator pads. The centers of the PS and SPD scintillator planes are separated by 56 mm along the beam axis. Each detector plane is divided vertically into two halves, which can independently slide horizontally for service and maintenance. Due to projectivity requirements, the SPD plane dimensions are approximately 0.45% smaller than those of the PS, and each plane is divided into inner, middle, and outer sections with cell dimensions chosen to achieve a one-to-one projective correspondence with the ECAL segmentation. The light produced by ionizing particles in the scintillator is guided by the WLS fibers to the detector box exit, where optical connectors join the WLS fibers to long clear fibers, which transport the scintillation light over several meters to the MAPMT without significant attenuation. MAPMTs, initially considered robust in magnetic fields, were revealed to be significantly sensitive to fields as weak as 1 mT. Thus, dedicated magnetic shielding was designed using a MuMetal cylinder along the tube axis, and the MAPMTs are housed in a soft iron box. The PMT gain is set to 10^4, ensuring stable operation conditions.

The electromagnetic calorimeter. For the ECAL [41], the Shashlik technology has been employed, which is a sampling scintillator/lead structure read by plastic WLS fibers. The Shashlik technology boasts several advantages: modest energy resolution, quick time response, acceptable radiation hardness, reliability, and expertise from other experiments. The ECAL is positioned 12.5 m from the interaction point, and its outer dimensions match the tracking system projectively. As for the SPD/PS detector, the hit density varies steeply with the distance from the beam pipe, leading to the subdivision of ECAL into inner, middle, and outer sections with suitable cell sizes. Each ECAL module layer alternates 2 mm thick lead, 120 µm thick white reflecting paper, and 4 mm thick scintillator tiles, all wrapped in black paper to ensure light tightness. The tile edges are chemically treated to enhance diffusive reflection, light collection efficiency, and prevent cross-talk between tiles. The scintillation light from the tiles is absorbed, re-emitted, and transported by WLS fibers that traverse the entire module. Then, the light is readout with phototubes. The effective energy resolution of the calorimeter is measured from collisions recorded at $\sqrt{s} = 8\,\text{TeV}$, and includes the intrinsic resolution of the cells plus the electron reconstruction effects, the interaction with the material in front of the calorimeter and the pile-up. This effective resolution (with deposited energy E in GeV) is [42]:

$$\frac{\sigma_E}{E} = \frac{(13.5 \pm 0.7)\,\%}{\sqrt{E}} \oplus (5.2 \pm 0.1)\,\% \oplus \frac{(320 \pm 30)\,\text{MeV}}{E} \tag{2.2}$$

The hadronic calorimeter. The HCAL [39] is a sampling device with iron as the absorber material, and scintillating tiles as active material. A unique feature of this sampling structure is the orientation of the scintillating tiles parallel to the beam axis. Laterally, tiles are separated by 1 cm of iron, while longitudinally, the length of the tiles and iron spacers matches the hadron interaction length in steel. Light is collected by WLS fibers running along the detector towards the backside where PMTs are housed. The transversely segmented square cells of the HCAL are divided into inner and outer sections. Readout cells of varying sizes are defined by grouping different sets of fibers onto a single PMT, fixed to the rear of the sampling structure. Constructed as a self-supporting wall, the HCAL is positioned at $z = 13.33\,\text{m}$ from the interaction point. It is uniformly instrumented without dead zones, and divided vertically into two symmetric parts on movable platforms for accessing it. The optics design includes the 3 mm thick scintillating tile, the WLS fiber, and a small square light mixer in front of the PMT entrance window. Scintillating light travels through the tile to its edges, and is collected by WLS fibers. Tiles are wrapped in a reflective envelope to prevent light cross-talk between adjacent tiles, to enhance light collection in the WLS fiber, and to protect the tiles optical reflective surface. Light propagates along the fiber by total reflection, but tiles far from the PMT yield less light due to attenuation. This effect is mitigated by progressively reducing the tile-fiber optical contact for different tile layers, ensuring uniform response across tiles. Each cell is readout by one PMT, shielded from stray magnetic fields by an iron tube and a MuMetal foil. The HCAL absolute calibration enables a cross-check of the calibration of upstream detectors, like the electromagnetic calorimeter, by comparing the average energy deposition of hadronic showers in corresponding cells of two detectors. The energy resolution measured in test beams with pions (with deposited energy E in GeV) is [43]:

$$\frac{\sigma_E}{E} = \frac{(67 \pm 5)\,\%}{\sqrt{E}} \oplus (9 \pm 2)\,\% \tag{2.3}$$

2.1.3.3 The muon system

Two additional requirements of LHCb are muon triggering and offline muon identification. In fact, muons are present in the final states of many CP-sensitive B decays. They play a major role in measuring CP asymmetry and oscillation, since muons from semileptonic b decays are used to tag the initial-state flavor of the neutral B mesons. In addition, the study of rare B decays, such as $B_s^0 \to \mu^+\mu^-$, may indicate the presence of new physics beyond the Standard Model. The LHCb muon system provides fast information for the high-p_T first level muon trigger and muon identification for the high level trigger and the offline analysis.

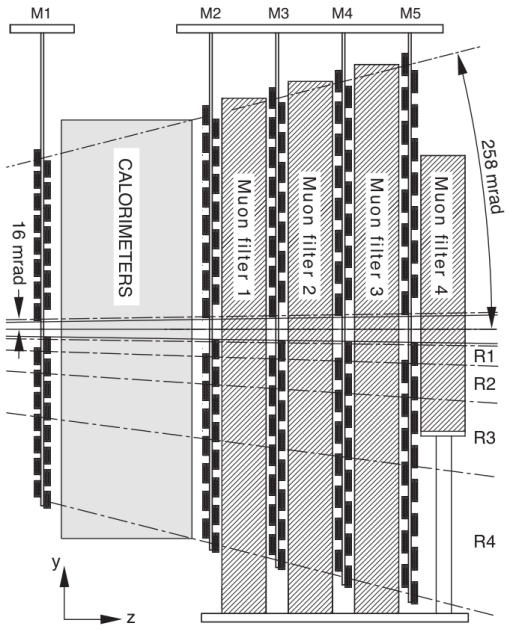

Figure 2.9: Side view of the layout of the muon system, showing the five stations M1-M5 and their segmentation regions R1-R4 [30].

The muon system [44–46], depicted in figure 2.9, consists of five rectangular stations M1-M5 positioned along the beam axis. The system includes 1380 chambers and spans a total area of 435 m^2. The inner and outer angular acceptances permit to identify about 20% of muons from inclusive b semileptonic decays. Stations M2 to M5 are located downstream of the calorimeters and interleaved with 80 cm thick iron absorbers to select penetrating muons, while station M1, placed in front of the calorimeters, enhances the p_T measurement in the trigger. The minimum momentum required to cross all five stations is approximately 6 GeV. The geometry of the five stations is projective, meaning their transverse dimensions scale with the distance from the interaction point. Stations M1-M3 offer high spatial resolution along the x coordinate (bending plane) to define the track direction, and calculate the p_T of the candidate muon with a 20% resolution. Stations M4 and M5 have limited spatial resolution, and primarily identify penetrating particles. Each muon station is divided into four regions, R1 to R4, with increasing y-distance from the beam axis, and linear dimensions and segmentations scaling in the ratio 1:2:4:8. This geometry ensures roughly equal particle flux and channel occupancy across the four regions of a given station. The (x, y) spatial resolution decreases moving away from the beam axis, where it is already limited by the increase of multiple scattering at large angles. The LHC bunch crossing rate of 40 MHz and the intense particle flux in the muon system impose stringent requirements on the detector efficiency, time resolution, rate capability, ageing characteristics, electronics speed, and radiation resistance. These necessities are satisfied by Multi-Wire Proportional Chambers (MWPC), which are used in all regions except the inner region of station M1, where the expected particle rate exceeds safety limits for ageing. In this inner region, triple-GEM (Gas Electron Multiplier) detectors are employed. The time resolution is achieved using a fast gas mixture and an optimized charge-collection geometry for both the MWPC and GEM detectors. The chambers consist of either four or two OR-ed gas gaps: in stations M2-M5, MWPCs have four gas gaps arranged in two sensitive layers with independent readout, while in station M1, the chambers have only two gas gaps to minimize material in front of the ECAL. In region M1R1, two superimposed OR-ed GEM chambers are used. Regarding the readout, different technical solutions are employed for the MWPC in various stations and regions to meet the varying requirements of spatial resolution and rate capability. All chambers are segmented into physical pads - anode wire pads or cathode pads in the MWPCs, and anode pads in the GEM chambers - with each physical pad readout by one Front-End (FE) electronics channel. The FE electronics is based on custom radiation-hard chips, with the input stage wired to handle either signal polarity: negative for anode pads and positive for cathode pads.

MWPC. The LHCb muon system includes 1368 MWPCs, which use a fast, non-flammable gas mixture of $Ar/CO_2/CF_4$ in a 40:55:5 ratio. The gas gaps are in 5 mm thick with 2 mm spaced wire planes, and the produced signals from two adjacent wired gas gaps are OR-ed to form a double gap with over 95% efficiency within a 20 ns window at a gas gain of approximately 10^5, reached at a voltage of 2600-2700 V. The pad readout minimizes cross-talk between pads to less than 5%. Construction parameters are optimized to restrict the gas gain variation within ±20%.

GEM. In the innermost M1R1 region, the detector must handle a charged particle rate of up to 500 kHz/cm^2, so these chambers need to be radiation-hard to avoid ageing effects. This is challenging to achieve with wire chambers unless the gain is lowered, which worsens the signal-to-noise ratio. Therefore, 12 triple-GEM chambers were selected for M1R1. Each chamber consists of two triple-GEM detectors superimposed to form two logically OR-ed sensitive layers. The triple-GEM detector is made of three GEM foils between anode and cathode planes, and it is effective as a tracking detector with good time and position resolution. The ionization electrons produced in the drift gap between the cathode and the first GEM foil are drawn through the three GEM foils by electric fields, where they are multiplied. After crossing the last GEM foil, they reach the anode in the induction gap, generating a current signal on the pads. The chosen fast gas mixture of $Ar/CO_2/CF_4$ (45:15:40) allows to perform at a time resolution better than 3 ns. Prototype measurements with two OR-ed detectors showed over 96% efficiency within a 20 ns window at a gain of 6×10^3.

2.1.4 The trigger system

The planned LHCb average luminosity of $2 \times 10^{32}\,\mathrm{cm^{-2}s^{-1}}$ is much lower than the LHC maximum design luminosity, minimizing the radiation damage to detectors and electronics. For this reason, most bunch crossings involve single interactions, leading to low channel occupancy, which in turn simplifies triggering and reconstruction. The selection and reduction of the interaction rate to store for offline analysis is managed through a two-level trigger system [47], outlined below and pictorially shown in figure 2.10: Level-0 (L0) and High Level Trigger (HLT). The L0 trigger uses custom electronics synchronized with the 40 MHz bunch crossing frequency, while the HLT operates asynchronously on a commercial processor farm. The offline analysis employs stringent selection cuts based on B mesons masses and lifetimes to enhance signal over background. The trigger system is optimized for maximum efficiency in offline-selected events, with the L0 trigger reducing the 40 MHz beam crossing rate to 1 MHz to match the detector readout capabilities. Particularly, the L0 trigger makes use of high transverse momentum (p_T) and energy (E_T) particles, typically produced in B mesons decays.

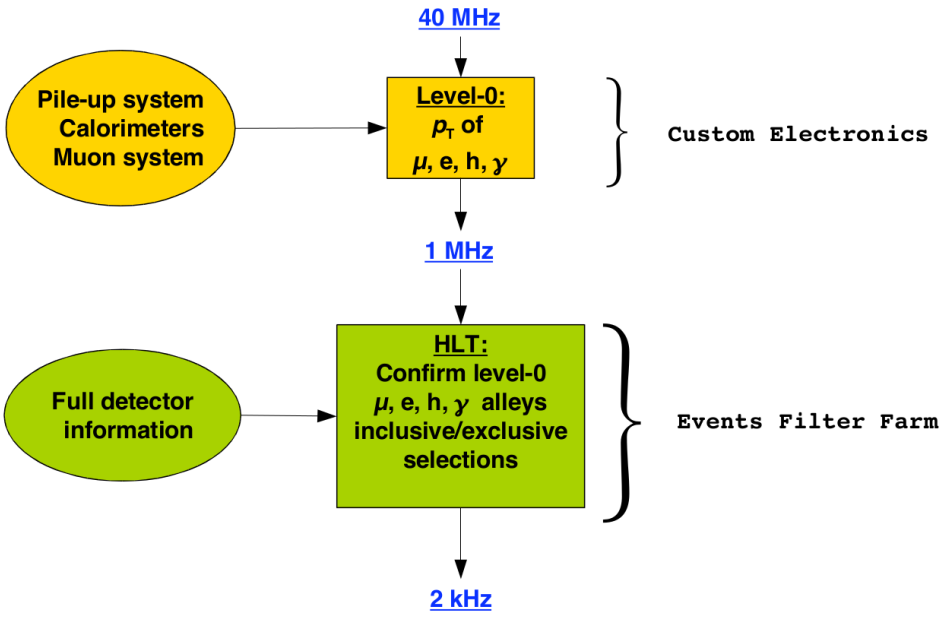

Figure 2.10: Description of the LHCb trigger system, displaying the two trigger levels, L0 and HLT, with their primary features [30].

A pile-up system in the VELO estimates the number of primary pp interactions in each bunch crossing. The calorimeters calculate the total observed energy and assess the number of tracks based on the number of hits in the SPD. These global quantities help rejecting events that would otherwise trigger due to large combinatorics, occupying an unreasonable fraction of the data-flow bandwidth or available HLT processing power. A L0 Decision Unit (DU) collects all this information and makes the final L0 trigger decision for each bunch crossing, pre-scaling and overlapping several trigger conditions. To reduce the event rate from 1 MHz to 2 kHz, the HLT uses the full event data. The generic HLT algorithms refine candidates found by the L0 trigger and sort them into independent alleys based on the L0 decision. Requiring candidate tracks with high p_T and/or large impact parameters reduces the rate to about 30 kHz, allowing the selection of interesting final states using inclusive and exclusive criteria.

Level-0 trigger. The L0 trigger is subdivided into the pile-up trigger, the calorimeter trigger, and the muon trigger. Each component is connected to one detector and to the L0 DU which collects all information to evaluate the final decision. The pile-up system uses four silicon sensors of the same type as the VELO ones to measure the radial position of tracks and of the Primary Vertices candidates along the beam line, and it estimates the total backward charged track multiplicity. It distinguishes between crossings with single and multiple visible interactions. The calorimeter trigger system looks for high E_T electrons, γ's, π^0's, or hadrons. It forms clusters by adding the E_T of 2×2 cells and selecting the clusters with the largest E_T, which are then identified as electron, γ, or hadron based on the information from the SPD, PS, ECAL, and HCAL. The muon chambers allow stand-alone muon reconstruction with a p_T resolution of $\sim 20\%$. Track finding is performed combining the strip and pad data from the five muon stations to form towers pointing towards the interaction region. The muon trigger selects the two muons with the highest p_T for each quadrant of the muon detector.

High Level Trigger. The all-software High Level Trigger consists of a `C++` application which runs on every CPU of the Event Filter Farm (EFF). Each HLT application has access to all data in one event, and thus, in principle, could execute the offline selection algorithms. However, given the 1 MHz output rate of the L0 trigger and CPU power limitations, the HLT aims at rejecting the majority of uninteresting events by using only part of the full event data. The typical L0 thresholds depend on luminosity and on the relative bandwidth division between the different L0 triggers. All L0 calorimeter clusters and muon tracks above threshold are passed to the HLT as part of the L0 trigger information, and are referred to as L0 objects. The HLT is subdivided in two stages, HLT1 and HLT2. The purpose of HLT1 is to reconstruct particles in the VELO and T-stations corresponding to the L0 objects, or, in the case of L0 γ and π^0 candidates, to confirm the absence of a charged particle which could be associated to these objects. HLT1 should reduce the rate to a sufficiently low level to allow the full pattern recognition on the remaining events, which corresponds to a rate of about 30 kHz. At this rate, HLT2 performs a combination of inclusive trigger algorithms, where the B decay is reconstructed only partially, and exclusive trigger algorithms, which fully reconstruct B-hadron final states.

Triggered events categories. In LHCb, the events accepted by the trigger are divided in three categories in order to estimate the trigger efficiency from data. The event categories are [48]:

1. Triggered On Signal (`TOS`): the signal is sufficient to produce a positive trigger decision;

2. Triggered Independent of Signal (`TIS`): the rest of the event - defined removing the signal and all corresponding detector hits - is sufficient to produce a positive trigger decision;

3. Triggered On Both (`TOB`): events that are neither `TIS` nor `TOS`, for which the simultaneous presence of the signal and the rest of the event are required to produce a positive trigger decision.

A single event can be at the same time `TIS` and `TOS` (`TISTOS`) if both the presence of the signal alone and the rest of the event alone are sufficient to produce a positive trigger decision. In other words, an event is either `TIS`, `TOS`, or `TOB` always with respect to a specified selection of trigger decisions, to be applied to the considered physical process.

2.1.4.1 Run 2 trigger and stripping lines involved in jets selection

In this thesis, two kinds of data sets from LHCb Run 2 are employed to measure the jet energy correction factors: a sample containing two reconstructed jets, thus called *dijet* sample, and a sample containing a Z^0 boson produced in association with a jet, and decaying into a $\mu^+\mu^-$ pair, named *Z+jet*. They will be discussed later in chapter 5, but here a short introduction on how the trigger selects candidates to build them is given.

Dijet samples. In LHCb Run 2, the selection of events with two jets is performed with the OR of all L0 lines, the OR of some HLT1 lines, and the HLT2 trigger line `HltQEEJetsDiJet`. The LHCb software has the ability to reconstruct jets with $p_T > 5\,\text{GeV}$, and the `HltQEEJetsDiJet` line selects events with two reconstructed jets with $p_T > 17\,\text{GeV}$, without any requirement on the Secondary Vertex. This line introduces a pre-scale factor due to the high rate of dijet events. The stripping line is `StrippingFullDiJetsLine`, that further constrains the jets transverse momentum, $p_T > 20\,\text{GeV}$, and adds another pre-scale factor. The L0 lines also consider the hit multiplicity measured in the SPD detector (see section 2.1.3.2), which has some problems in reproducing efficiencies, so another correction factor, called Global Event Cut (GEC), is applied in the form of an efficiency contribution. All the trigger lines and correction factors used for the dijet calibration sample in this work are specified in section 6.3.1.

Z+jet samples. Events containing a jet and a Z^0 boson decaying into a $\mu^+\mu^-$ pair are selected triggering only on muons: all events passing muon lines are kept by the trigger (online), then jets are reconstructed (offline) with the LHCb algorithm described in section 3.1. Data sets built in this way are good jet calibration samples, because the reconstructed jet variables (p_T, η, etc.) are not biased by the trigger selections. In fact, trigger cuts inevitably cause the deviation of the jet variables spectra from the truth-level ones. These biases affect the dijet samples, but can be eliminated analyzing asymmetry variables, in which they cancel out between numerator and denominator. The cuts applied to muons in this thesis are reported in section 6.2.1.

2.1.5 The data acquisition and control systems

The data acquisition and control systems [49, 50], formed by the three components briefly described below, ensure the transfer of data from the front-end electronics to permanent storage, and the configuration and monitoring of all operational and environmental parameters, such as temperatures and pressures. It also guarantees the proper synchronization of all detector channels with the LHC clock.

The Data AcQuisition (DAQ) system. It transports the data belonging to a given bunch crossing, and identified by the trigger, from the detector front-end electronics to permanent storage.

The Timing and Fast Control (TFC) system. It drives all stages of the data readout between the front-end electronics and the online processing farm, distributing the beam-synchronous clock, the L0 trigger, synchronous resets and fast control commands. The TFC system is a combination of electronic components common to all LHC experiments, and LHCb custom electronics.

The Experiment Control System (ECS). It controls and monitors the operational state of the whole LHCb detector, including not only the traditional detector control domains, such as high and low voltages, temperatures, gas flows, or pressures, but also the trigger, TFC, and DAQ systems. The hardware components of the ECS are heterogeneous, mainly because of the variety of equipments to be controlled, ranging from standard crates and power supplies to individual electronics boards.

2.1.6 The LHCb software

As discussed in section 2.1.4, 40 million *pp* collisions occur every second at LHCb: if all data were to be stored, $\sim 1\,\text{TB}$ would be recorded every second, and that is too large a quantity to retain, considering that the price of storage is too high. Moreover, data arrives very quickly, so it must be processed accordingly, and it is complex, so it can be handled in several ways. Thereby, the events must be filtered to keep only the ones containing interesting information, and in LHCb this is done

by a data flow designed to maximize the data-taking efficiency and data quality [51, 52]. It consists of several steps, which in Run 2 were the ones visible in figure 2.11, each controlled by an *application* that processes event-by-event data.

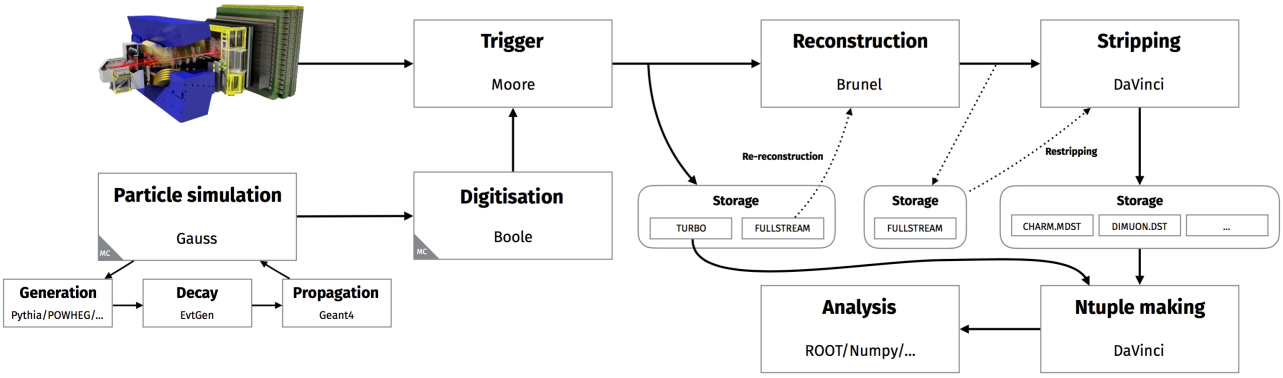

Figure 2.11: The LHCb Run 2 data flow with the software applications for real and simulated data [52].

The steps are the following.

1. Data pass from the detector through the trigger. The application in charge of the software trigger (HLT) is `Moore` [53].
2. The `Brunel` application [53] reconstructs triggered, raw detector hits into objects, such as tracks and clusters, which are stored into an output file in a DST format[2].
3. The reconstructed DST files, not yet accessible to users due to computing restrictions, are further filtered through a set of selections called the *stripping*, controlled by the `DaVinci` application [53], which outputs either a DST or µDST file. To save disk space and speed up access, the output files are grouped into streams containing similar selections, so that analysts interested in a specific physics channel do not need to deal with all the outputs.
4. Variables for analyses are extracted with `DaVinci`. The processing is different between DST and µDST, since some calculations need not only the signal, but also the other tracks in the event, which are not available in the µDST format.

Simulated Monte Carlo events are also produced, and they are managed similarly to real data. This analogy is beneficial, as in this way the simulated data is subject to the same deficiencies as real data. Two simulation steps mimic the *pp* collisions and the detector response.

1. The *pp* collisions, and the hadronization and decay of the resulting particles, are simulated by the `Gauss` application [55]. `Gauss` is responsible for implementing the supported Monte Carlo generators (the default one is `Pythia8` [56]), which simulate the *pp* collisions and hadronization. `Gauss` also controls `EvtGen` [57], which describes the decay of the simulated particles. `Geant4` [58] approximates the detector geometry and material content, taking care of the propagation and interaction of particles in the detector.
2. The virtual detector hits are digitized into signals that emulate the real detector ones by the `Boole` application [53]. It models the response of the subdetectors to passing particles and the subdetectors readout, as well as the L0 trigger. Background effects, such as noise and spillover[3], are simulated at this point. The `Boole` output closely matches the output of the real detector, so that the simulated data can be passed through the usual data processing chain described above, starting from the trigger stage. `Moore` is responsible for the emulation of L0 for Monte Carlo data.

[2] A DST file is a `ROOT` [54] file with the full event information, that is reconstructed objects and raw data. The µDST (micro-DST) format was created to save space by discarding the raw event and storing only the information about the objects selected to construct particle decays, like tracks.

[3] Spillover is a term that describes the situation where information from previous bunch crossings remain within the detector electronics also during the subsequent bunch crossing.

With few exceptions, only `DaVinci` is run by users, while every other step is operated centrally. The reconstruction with `Brunel` is computationally intensive, so it is only performed when data are taken, and when there is a new reconstruction configuration. The stripping runs on reconstructed data, so it can be executed more often.

2.1.7 Luminosity measurement

One of the main parameters to study and control at an accelerator is the luminosity [59], which has already been mentioned several times in the previous pages. In particle physics experiments, in order to produce new effects, not only the available center of mass energy of the colliding beams must be preserved without losses, but also the number of useful interactions is significant. This is even more meaningful when dealing with rare events, which have small production cross sections σ. The quantity that measures the ability of a particle accelerator to produce the required number of interactions is called *luminosity*, and represents the proportionality factor between the number of events per second dN/dt and the production cross section σ:

$$\frac{dN}{dt} = \sigma \mathcal{L} \tag{2.4}$$

The conventional unit of measurement of the luminosity is $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. After some calculations [60], the luminosity for one colliding bunch pair can be expressed as:

$$\mathcal{L} = \nu_{rev} N_1 N_2 \mathcal{O} \tag{2.5}$$

Here ν_{rev} is the collider revolution frequency, and N_i, with $i = 1, 2$, is the number of particles in a bunch in beam i. $\mathcal{O}$ defines the overlap integral:

$$\mathcal{O} = 2c \int \rho_1 (x, y, z, t)\, \rho_2 (x, y, z, t)\, dx\, dy\, dz\, dt \tag{2.6}$$

where ρ_i, with $i = 1, 2$, is the unit-normalized particle density function of the beam i. With the inclusion of the kinematic factor $2c$, $\mathcal{O}$ has the unit of a surface^{-1}, or cm^{-2}. The overlap integral accounts for the fact that the colliding beams are not stationary, but move towards and through each other. Hence, $\mathcal{O}$ depends on the longitudinal position of the bunches and on the time.

In LHCb Run 1 and Run 2 configuration, two methods to measure $\mathcal{L}$ were exploited, which are summarized in the following: the Van der Meer scan method, and the Beam-Gas Imaging (BGI) method.

2.1.7.1 The Van der Meer scan method

The luminosity defined in equation 2.5 requires the evaluation of the bunch population product $N_1 N_2$, and of the overlap integral $\mathcal{O}$. One method to measure the overlap is the Van der Meer (VDM) scan method [60,61], whose key principle is to express $\mathcal{O}$ in terms of rates that are experimental observables, without needing the bunch density functions. Experimentally, the VDM method consists in moving the beams across each other in the vertical and horizontal directions. Then, the overlap integral can be inferred from the rates measured at different beam positions. The starting assumptions are that the beam densities are independent in the x and y variables, and that the crossing angle between the beams is negligible. As just stated, the interaction rate $R = dN/dt$, which is linear with the luminosity (see equation 2.4), is used as observable. Accordingly, the absolute luminosity at a fixed beam displacement $(\Delta x_0, \Delta y_0)$ is given by [60]:

$$\mathcal{L} (\Delta x_0, \Delta y_0) = \nu_{rev} N_1 N_2 \frac{R (\Delta x_0, \Delta y_0) \cdot R (\Delta x_0, \Delta y_0)}{\int R (\Delta x, \Delta y_0)\, d (\Delta x) \cdot \int R (\Delta x_0, \Delta y)\, d (\Delta y)} \tag{2.7}$$

The pair of displacements $(\Delta x_0, \Delta y_0)$ is called *working point*, and is typically chosen to be as close to $(0, 0)$ as possible, to consider head-on beams. The VDM formula still remains valid for any Δx_0 and Δy_0 positions. It can be shown that equation 2.7 holds even in the presence of non-zero crossing angles.

The VDM method has two main advantages: it uses the detector rate as its only observable, which is experimentally simple, and it does not assume a specific beam shape. The experimental difficulties of the VDM method arise mostly from the fact that the beams must be moved to perform the measurement. The exact displacements steered by the LHC magnets must be calibrated, and, while the resulting corrections are typically of the order of 1%, some larger non-reproducibilities have been observed between two consecutive scans without being able to identify the cause. Another difficulty originates from beam-beam effects. When the beams are displaced, a change in the transverse size of the beams and a beam deflection may be produced, which influences the observed rate. In addition, the VDM method as performed in only one vertical and one horizontal scan, is valid only under the assumption that the transverse directions x and y are independent, i.e., that the x (y) shape measured at a working point Δy_0 (Δx_0) does not depend on the working point position. Unfortunately, the beams can only be considered factorizable in x and y for the measurements performed in November 2012, and in all other measurements a time dependence was observed in the cross section measurements traced back to the non-factorizable shape of the beams. All these effects are taken into account into correction factors applied by LHCb.

2.1.7.2 The Beam-Gas Imaging method and SMOG

An alternative direct luminosity measurement is the Beam-Gas Imaging (BGI) method [60, 61], which at the time of Run 1 and 2 was a unique capability of the LHCb experiment. The principle of this method is to evaluate the overlap integral of equation 2.6 by measuring all required observables using the interaction vertices of the beam with a gas. The employment of stationary beams has the advantage of avoiding effects due to changes of beam-beam conditions and of beam steering, which may be non reproducible. Furthermore, the measurements are made parasitically while other experiments perform VDM scans, and so they can be made in every dedicated luminosity calibration fill. At the same time, while the β^*[4] and crossing angles used at the LHC did not influence the VDM method, the BGI measurement demands the information on the vertex to determine the bunch shape. Therefore, not to be limited by the detector resolution, an increased β^* is more suitable. The knowledge of the crossing angle is important because it can reduce the luminosity up to 20%. The crossing angle is also necessary to avoid interactions between the main bunch and out-of-time charges. Such displaced collisions, centered at $z = \pm 37.5\,\text{cm}$, can not be cleanly disentangled from beam-gas interactions, and they can be avoided precisely by introducing a crossing angle. The VDM measurement can exclude interactions happening far from the interaction point, and is not affected by these so called *satellite* collisions. A precision luminosity measurement requires the *double gaussian model*, a shape description with a superposition of two gaussian functions, as a single gaussian model does not describe the bunch shape well enough. An additional factorizability parameter is required with double gaussian beams to account for the relation between the x and y description. The overlap integral with double gaussian beams can be written as the sum of partial single gaussian overlap integrals resulting from the combination of all partial double gaussian widths, and each term has an individual weight depending on the factorizability parameter. The non-gaussian shape and factorizability have to be taken into account also with the VDM method. Consequently, The VDM and BGI methods are complementary in the experimental difficulties. The systematic uncertainties on the overlap integral coming from the two methods are highly uncorrelated, and a luminosity calibration performed with both methods in the same fill allows to further constrain the systematic uncertainties. This could only be done by the LHCb experiment.

[4]The parameter β^* describes the focusing properties of a beam at the interaction point. It is related to the *beta function*, a characteristic of the beam optics that describes the transverse size of the beam along its nominal trajectory. Specifically, β^* indicates the level of beam focusing at the interaction point. The evolution of the beta function around the interaction point is given by $\beta(z) = \beta^* + z^2/\beta^*$, where z is the nominal beam direction.

As anticipated, the BGI method needs the interaction vertices between the circulating beam and the residual gas at the interaction point. This method led to the development of SMOG (System to Measure the Overlap integral with Gas) [60–62], which started operation in Run 2. SMOG injects noble gases (helium, neon, argon) into the beam pipe section within the VELO region. During SMOG operation, the two VELO's ion pumps are switched off, and gas flows towards vacuum pumps located approximately $\pm 20\,\mathrm{m}$ from the nominal LHCb collision point. The gas pressure in the VELO region is kept at about $2 \times 10^{-7}\,\mathrm{mbar}$, minimally affecting LHC operation. This setup achieves a luminosity for collisions in 1 m (the length of the vacuum vessel) of gas along the beam direction of $6 \times 10^{29}\,\mathrm{cm^{-2}s^{-1}}$ for an LHC beam of 10^{14} protons, roughly matching the LHCb vertex detector acceptance. SMOG also gave LHCb the opportunity to work as a fixed target experiment, opening up a new physics program, including: the study of cold nuclear matter effects through prompt heavy-flavor production in proton-nucleus collisions; Quark Gluon Plasma (QGP) formation in lead-nucleus collisions; searches for intrinsic charm in the proton at high Bjorken-x; measurements of prompt antiproton production in proton-helium collisions, relevant for dark matter searches.

2.2 LHCb Upgrade Ia

The LHCb Upgrade Ia experiment [63], now in operation during LHC Run 3, has been designed to face the wider physics program that LHCb has proven to be able to address as a general purpose experiment covering the forward region. Notwithstanding the considerable data set successfully collected during Runs 1 and 2, the precision on many key flavor physics observables remains statistically limited. Thus, significantly larger data sets are needed to probe the SM at the theoretical level of precision and to obtain the prescribed sensitivity to observe new physics effects. The instantaneous luminosity at which LHCb Upgrade Ia is working is substantially larger than in Runs 1 and 2. However, the original experimental design did not allow to significantly increase statistics, especially for fully hadronic final-state decays, the main limitation being the maximum allowed output rate of the L0 trigger. The simple p_T-based inclusive selection criteria implemented in L0 results in an efficiency loss with increasing luminosity, and in the saturation of the event yield. In addition, since inclusive flavor physics signals have relatively large cross sections, simple cuts based on displaced vertices or on p_T would be either not effective in rejecting background or would downscale the signal. Therefore, a solution was found by removing the hardware L0 trigger stage and introducing more discriminating selection criteria. A full-software trigger separating signal channels based on the full event reconstruction has been adopted. The first stage of the upgraded trigger is implemented on a GPU farm, while the output of the trigger produces a combination of totally reconstructed physics objects ready for final analysis, such as tracks and vertices, and of entire events, which need extra offline reprocessing.

Based on these considerations the LHCb Upgrade Ia has been designed to run at a nominal instantaneous luminosity $\mathcal{L} \sim 2 \times 10^{33}\,\mathrm{cm^{-2}s^{-1}}$ and to collect events at the LHC bunch crossing rate of 40 MHz. The all-software trigger reconstructs all events in real time at the visible interaction rate of $\sim 30\,\mathrm{MHz}$. By the end of Run 4, the total integrated luminosity (including Run 1 and Run 2) that is expected to be reached is $\sim 50\,\mathrm{fb^{-1}}$. The new trigger strategy, the higher luminosity and the subsequent higher pile-up required an overhaul of all the LHCb detectors and readout electronics, that are now able to read events at the LHC bunch crossing rate and cope with the larger event multiplicity thanks to a higher granularity. A revision of the experiment's software and of the data processing and computing procedures was also mandatory to deal with the large increase in data volume.

Figure 2.12 shows the layout of the upgraded detector. The tracking system comprises the VELO, the new Upstream Tracker (UT), and three new Scintillating Fibre Tracker (SciFi Tracker) stations. The VELO was upgraded replacing the silicon microstrips with arrays of hybrid silicon pixel detectors surrounding the interaction region. The UT is placed between the RICH1 and the magnet, and is made of four planes of silicon microstrip detectors organized in two stations. The SciFi Tracker stations are located downstream of the magnet, and substitute both the straw-tube Outer Tracker and silicon-strip Inner Tracker systems used in the original setup. The SciFi Tracker consists of twelve detection planes arranged in three stations, containing plastic scintillating fibres placed in multilayered fibre mats.

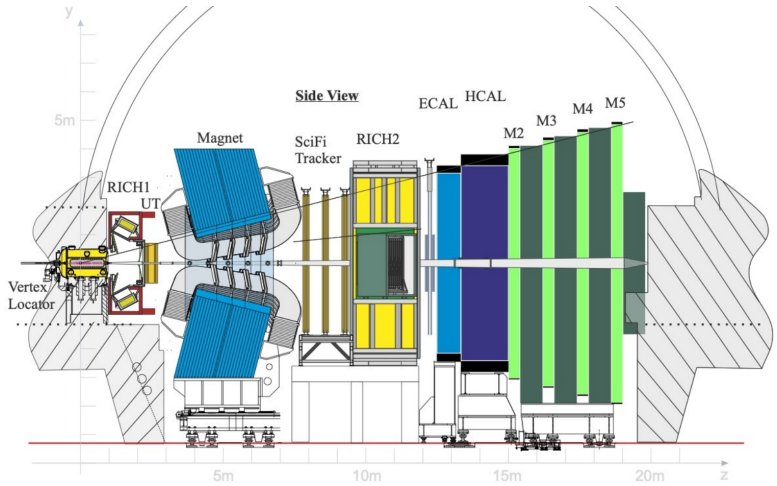

Figure 2.12: Layout of LHCb Upgrade Ia [63].

The Particle IDentification (PID) is provided by two Ring Imaging CHerenkov detectors (RICH1 and RICH2), a Shashlik-type Electromagnetic CALorimeter (ECAL), an iron-scintillator tile sampling Hadronic CALorimeter (HCAL), and four stations of muon chambers (M2–5). The overall layout of the RICH system remains unchanged with respect to Run 1-2, but the single photon resolution has been improved to keep the excellent PID performance provided in Run 1 and Run 2 in the more challenging conditions of Run 3. The layout of both the ECAL and HCAL remains unchanged, but, to deal with the new LHCb readout system, the electronics of the calorimeters has been entirely replaced. Moreover, the Scintillating Pad Detector (SPD) and the PreShower (PS) have been removed, given their reduced role in the new LHCb all-software trigger. The muon stations, interleaved with iron shielding, are the same as the original LHCb setup. The most upstream M1 station present in the initial layout has been removed, since it was utilized in the hardware L0 trigger, and thus is no longer needed in the upgraded system. A dedicated luminosity subdetector, named Probe for LUminosity MEasurement (PLUME) [64], has been installed upstream of the *pp* interaction region (between $z = -1900\,\text{mm}$ and $z = -1680\,\text{mm}$. It measures the Cherenkov light produced by charged particles crossing a quartz radiator.

Also, an improved design for the gas target, SMOG2 [65], has been integrated in the design of the upgraded VELO detector. In this new design, the gas is contained in a storage cell made of two retractable halves, with the target placed upstream of the nominal LHCb collision point. The use of more gas species, notably H_2 and D_2, and heavier noble gases (Kr, Xe) is now possible. Outside the open ends of the cell, the gas density is suppressed by the VELO vacuum pumps, so that the beam-gas collisions occur mostly inside the cell, resulting in an increased effective areal density (and thus luminosity) in the target by a factor of about 8 (with H_2) to 35 (with Ar) with respect to SMOG for the same gas flow. This, combining the possibility to measure both the gas flow and the cell temperature precisely, allows for a significantly more precise determination of the target density (and luminosity), reaching the level of few percent accuracy. It is also possible to exploit all circulating bunches for fixed target physics, i.e. to run simultaneously in fixed target and in collider mode, thanks to the limited extension of the SMOG2 target region, displaced and not overlapping the *pp* interaction point. In particular, the use of H_2 and D_2 targets, available for the first time at the LHC, opens new physics frontiers that include the study of the proton structure in terms of quark and gluon distributions at unique kinematic conditions, and a significantly better defined interaction region (determined by the limited cell length of 20 cm).

2.3 LHCb Upgrades Ib and II

The LHCb Upgrade II program [66] aims at making full use of the potentialities of LHCb as a forward acceptance detector during the High Luminosity LHC (HL-LHC) operational period, which will start after LS3. LHCb Upgrade II will be installed during LHC Long Shutdown 4, with operations beginning in Run 5. This experiment will operate at instantaneous luminosities of up to $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, an order of magnitude above Upgrade Ia. New attributes, installed in LS3 and LS4, will enhance the detector capabilities to a wider range of physics signatures. Consolidation of the LHCb Upgrade Ia detector is required during LS3, an extended shutdown period, in which preparatory work for Upgrade II will also be performed. These changes are known as Upgrade Ib. In the meantime, LHCb will continue data-taking at an instantaneous luminosity of $2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ until LS4. At the planned luminosity of $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, compared to Upgrade Ia, there will be a further order of magnitude increase in data output rates, accompanied by a corresponding increases in radiation levels and occupancies. Efficient real-time reconstruction of charged particles and interaction vertices within this environment represents a significant challenge. To meet this challenge it is foreseen to modify the existing spectrometer components to further increase the granularity, and reduce the amount of material in the detector. A key addition will be to exploit the use of precision timing. It is expected that the performance of the Upgrade Ia detector will be maintained or improved in terms of track-finding efficiency and momentum resolution. Summarizing, the scheduled improvements of LHCb regard the subdetectors granularity and radiation hardness, and the addition of timing information.

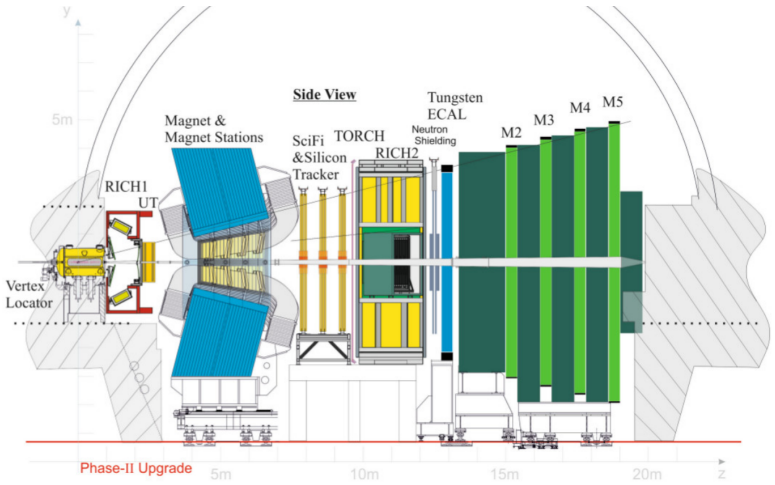

Figure 2.13: Layout of LHCb Upgrade II [66].

In figure 2.13, the planned layout for Upgrade II is depicted. The upgraded VELO will handle the increased pile-up from LHC bunch crossings by enabling real-time track reconstruction in the software trigger. This involves developing a 4D hybrid pixel detector with enhanced rate and timing capabilities, improving the mechanical design for periodic module replacement, and minimizing material before the first track measurement. These changes will improve impact parameter and momentum resolution, reducing the impact of secondary interactions on downstream detectors. Timing information will increase the efficiency in associating decay products to Primary Vertices, and reduce combinatorics at an early stage in track reconstruction, saving CPU space. The downstream tracking system will also be upgraded to cope with higher occupancies, increasing the detector granularity, and minimizing the rate of incorrect matching of upstream and downstream track segments. This will be achieved optimizing the UT, minimizing detector material, and incorporating timing information. The inner part of the T-stations will be replaced with a high-granularity silicon detector. As an intermediate step, the upgrade of the inner part of the SciFi during Upgrade Ib is proposed for the LS3. The design of this new silicon detector is known as the MIghTy tracker, featuring inner and middle tracker elements. The magnet internal surfaces will be instrumented with scintillating fibers to extend the spectrometer coverage for low momentum tracks, achieving the required momentum resolution with a spatial resolution of $\sim 1\,\mathrm{mm}$ in the bending plane, with a stereo layer arrangement for y segmentation.

The LHCb Upgrade II focuses on enhancing the already high quality PID systems with improved granularity and fast timing (of the order of a few tens of picoseconds) capabilities to manage the increased track multiplicity. The RICH system will evolve with new high-granularity photodetectors, replacing the MAPMTs of Upgrade Ia with SiPMs, and will enhance the Cherenkov angle resolution through redesigning the optics and using smaller pixel sizes. The potential addition of a TORCH detector upstream of RICH 2 could improve the low-momentum hadron identification by measuring the time-of-flight through detecting internally reflected Cherenkov light produced in a thin quartz plane with MicroChannel Plate (MCP) photodetectors. The electromagnetic calorimeter will undergo a significant overhaul to address several issues: radiation damage; high luminosity, with consequently overlapping showers, and degradation in energy resolution and shower finding efficiency; large combinatoric background. The modifications include a radiation-hard design, reduced cell size for better energy resolution, and fast timing information. Possible solutions include a homogeneous crystal calorimeter with longitudinal segmentation, or a sampling calorimeter, either Shashlik or SpaCal, with a tungsten-alloy converter. The HCAL is replaced by up to 1.7 m of iron, providing an additional shielding compared to the current situation. The primary role of the HCAL was to contribute to the hardware trigger, which is no longer required, as discussed in section 2.2. The muon system will be upgraded to handle the higher background flux by increasing granularity, replacing some detectors with new pad chambers, and adding shielding around the beam pipe. For Upgrade II, extra shielding and new high granularity, high rate detectors, such as the micro-resistive WELL detector (μ-RWELL), will be installed. These enhancements aim at improving the system's ability to operate under severe luminosity conditions.

As far as the luminosity determination is concerned, also SMOG2 has a planned upgrade. The proposals for Runs 4 and 5 are: installing a polarized gas target upstream of the VELO; an off-axis solid target coupled to a bent crystal, hit by a secondary proton beam, to measure the magnetic and electric dipole moments of heavy baryons and, possibly, τ leptons; a metal microstrip target, with a large variety of nuclear targets.

At an instantaneous luminosity of up to $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the LHCb detector will produce up to 400-500 Tb of data per second, which must be processed in real time and reduced by at least 4-5 orders of magnitude before permanent storage. Advances in radiation-hard optical links and commercial networking technology, coupled with LHCb's geometry, will facilitate moving this data into a processor farm by Upgrade II. The primary challenge is managing up to ~ 55 interactions per bunch crossing, containing multiple heavy-flavor hadrons. Traditional trigger strategies will not suffice, necessitating early pile-up suppression and precise timing information to discard irrelevant detector hits. Early full reconstruction of charged and neutral particles is crucial for optimal matching of heavy flavor signatures to pp interactions. LHCb experience in real-time data processing during Run 2 will aid in scaling up to the more demanding conditions of Upgrade II, leveraging future commercial hybrid processor architectures, likely involving a mix of CPU, GPU, and FPGA accelerators, in collaboration with computing institutes.

Chapter 3

Jets at LHCb Run 2

LHCb data of pp interactions taken in 2016 during LHC Run 2 and containing hadronic jets are analyzed in this thesis work. More specifically, from a sample of events with the production of at least two jets, the dijet cross section will be extracted, and it will be used as an observable to extract the value of α_s and to evaluate its main systematic uncertainties.

In a typical event, two proton beams, each with an energy of 6.5 GeV, collide in the established LHCb interaction point, resulting in a center of mass energy $\sqrt{s} = 13\,\text{TeV}$. The protons in the beams travel in *bunches* carrying several protons each. When the bunches intersect at the interaction point, the protons cross each other, leading to many, almost simultaneous collisions, and this phenomenon is called *pile-up*. A pp collision takes place in the so called *Primary Vertex* (PV). The interactions to be identified, stored, and studied are the ones producing physically relevant final states, while the other pile-up interactions must be discarded. In Run 2, due to luminosity considerations, it was planned to have only one pp interaction per bunch crossing, but LHCb proved to be capable to reach higher pile-up conditions.

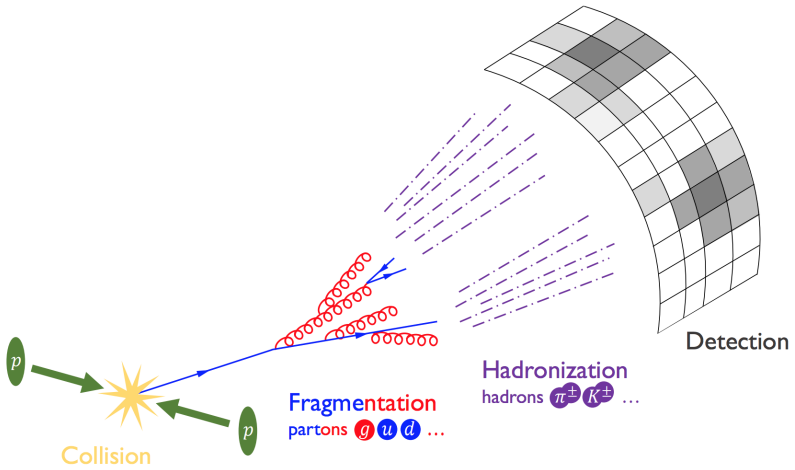

Figure 3.1: Diagram of a standard pp interaction: in the hard scattering (*Collision*) between the partons of the incoming protons, new partons are produced, that undergo fragmentation, and then hadronize into hadrons, which hit the detectors producing signals [67].

As described in section 1.4, a proton is made of partons. At LHC energies, pp collisions involve interactions between partons, referred to as *hard scattering* events, governed by the strong interaction. Partons produced in the initial collision undergo showering and decay, leading to the creation of new particles that mirror the momenta of the initial partons. Detectors identify the final particles, that are then analyzed to reconstruct the initial state and the whole collision event. The final-state particles are the long-lived stable[1] ones resulting from the decay of the unstable particles produced in the shower.

[1] In LHCb, a particle is considered stable if its mean lifetime is $\tau > 10^{-8}$ s.

The showering process consists of two crucial phases that determine the evolution of the collision event and the transformation of high-energy partons into final particles (see figure 3.1). Firstly, *fragmentation* occurs immediately after the hard scattering event, where a high-energy parton converts into multiple lower-energy partons through gluon emissions and quark-antiquark pairs creation, leading to a cascade of partons. In the second phase, *hadronization*, these partons are combined into stable hadrons that can be detected. Hadronization happens at a larger scale compared to fragmentation, generally at distances of the order of 1 fm.

At this point, the concept of particle jets must be introduced [68, 69]. Individual particles emerging from hadronization carry little to no information about the underlying hard scattering. However, a sufficiently hard parton can produce a collimated spray of high-energy particles - a *jet* - whose collimation is directly proportional to the momentum of the parton. Jets leave clear signatures in an event display, and their measured energy and direction can be used to infer the nature of the original parton. Since a jet represents a collection of particles rather than a single one, jet reconstruction methods are needed to define how to group particles into a jet, and to assign a momentum to it. The four steps of a standard LHCb jet reconstruction algorithm will be specified in section 3.1.

Jets are widely used in various physics analyses, and can be classified according to the origins of the partons that give rise to them. At hadron colliders like the LHC, the inclusive jet spectrum is a commonly used observable, associated with high-momentum-transfer partonic $2 \to 2$ scattering. In this case, the jet energy closely relates to the original parton energy, providing insight into partons distribution and their interaction strength. Partons forming jets can also originate from the hadronic decay of heavy particles, such as top quarks, Higgs bosons, or new unknown resonances, which produce multiple jets whose combined momenta correspond to the invariant mass of the original particle, supporting PID techniques. Additionally, jets can form through radiative processes, e.g. gluon emissions, which inform on the strong coupling constant and the events color structure, distinguishing between different production mechanisms. QCD radiative corrections also give substructures to jets, and managing these emissions helps avoid background interference in multi-jet signals, ensuring accurate matching between tree-level matrix-element calculations and Monte Carlo parton showers.

The previous paragraphs highlight the non-triviality of identifying a jet as coming from a single parton. Actually, the association is rather ambiguous, also because jets are not physical objects, but are defined by specific algorithms. To be addressed as good, a jet definition must be applied to experimental measurements, to the output of parton-showering simulations, and to partonic calculations, giving as a result jets with a common and consistent representation of all these kinds of events.

3.1 Jets reconstruction

Several components of the LHCb detector are involved in jet reconstruction: VELO provides precise measurements of the primary and secondary vertices, essential for identifying the origin of particles; the Tracker Turicensis and the Inner and Outer Trackers measure the momentum of charged particles; ECAL and HCAL measure the energy of particles and play a crucial role in identifying jets; the muon system assists in identifying muons, which are important for distinguishing between different types of jets, such as those containing b quarks. In this section, the four stages of the jet reconstruction in LHCb are presented: the selection of the input particles with the Particle Flow method, the clustering of the particles into jets with a jet reconstruction algorithm, the assignment of a four-momentum to each jet in a recombination scheme, and the correction of the resulting jet four-momentum.

3.1.1 Particle Flow algorithm

The first step of the LHCb jet reconstruction procedure is the *Particle Flow* method [70], which gathers information from throughout the detector to provide a list of input particles to be reconstructed into jets by a jet algorithm. Considering that the LHCb calorimeter cells are coarse and saturate at high energies, giving the jet clustering algorithms calorimeter-based inputs only is not possible at LHCb. The tracking systems (see section 2.1.2) allow an adequate energy resolution of charged particles,

and therefore in LHCb the Particle Flow approach uses different subdetectors to measure different particles, optimizing for the jet energy resolution. It uses information from tracks where possible, and from calorimeters when no track is available, e.g. for neutrons and photons. As previously discussed, a peculiar advantage of LHCb is that collisions take place in a low pile-up environment, lessening the effects of particles produced in pp interactions other than the relevant ones. The effect of pile-up is also reduced thanks to the excellent position resolution of the VELO, so tracks originating in the VELO can have a Primary Vertex requirement placed on them. The calorimetric clusters reconstruction is performed for each Primary Vertex, and the difficulty is that there is no way to uniquely associate calorimeter clusters with a single vertex, so all calorimeter clusters are considered for each PV.

An experimentally observed object considered by the Particle Flow method implemented in LHCb Run 2 is categorized as:

1. *charged particle*, if it is associated to a track;
2. *isolated neutral particle*, if it produces a calorimetric signal, but can not be associated to any track;
3. *non-isolated neutral particle*, if it produces a calorimetric signal, and can also be associated to a track.

Below, the criteria applied to selected these particles are exposed. These particles are the inputs of the LHCb jet algorithm, which clusters them into jets.

3.1.1.1 Tracks

The quality of a track is determined by means of:

- the transverse momentum p_T;
- the Kalman fit χ^2 [71];
- the probability, P_{ghost}, that the track is a ghost track, i.e. an artifact of the reconstruction process rather than originating from a real particle;
- the momentum resolution $\frac{\sigma(q/p)}{q/p}$, where q is the charge of the considered track.

Among all the track types[2] that LHCb identifies, long, upstream, downstream, and VELO tracks are taken into account. The cuts on the values that have just been listed depend on the track type [72,73], and are reported in table 3.1.

Selection variable	Track type			
	Long	Upstream	Downstream	VELO
p_T [MeV]	-	> 100	-	-
χ^2	< 5	< 1.5	< 1.5	< 10
P_{ghost}	< 0.5	-	-	-
$\frac{\sigma(q/p)}{q/p}$	> 10	> 2	> 10	-

Table 3.1: Quality selection criteria required for tracks [72, 73].

3.1.1.2 Isolated neutral particles

Isolated neutral particles detected by ECAL are mainly photons and π^0 decayed into photons. π^0 mesons are named *merged* if they are reconstructed in a single ECAL cluster, and *resolved* if they are not. Three quantities are considered in the selection:

[2]The terminology for track types in LHCb can be found in section 2.1.2.4.

- the likelihood of the photon hypothesis, `PhotonID`, computed from the cluster shape;
- the transverse energy of the cluster, E_T;
- the matching chi square, χ^2_m, for each cluster-track combination, which expresses the quality of the matching itself.

The selection thresholds depending on the particle type [74] are shown in table 3.2. As far as HCAL is concerned, instead, the selection depends only on χ^2_m: it must be greater than 25 if the cluster energy is below 10 GeV, greater than 15 otherwise [74].

Selection variable	Isolated neutral particle type		
	Merged π^0	Resolved π^0	Photon
E_T [MeV]	-	-	> 200
`PhotonID`	-	> -4	> -1
χ^2_m	> 25	> 25	> 25

Table 3.2: Selection criteria required for isolated neutral particles [74].

3.1.1.3 Non-isolated neutral particles

A non-isolated cluster is typically the result of a neutral particle hitting the calorimeter near a charged particle, where both deposit energy in the same cluster. Track-cluster matching is crucial for discriminating between non-isolated and isolated clusters, and the matching chi square, χ^2_m, is a key variable used as selection rule. To reconstruct the energy of the neutral particle, an analytic parameterization of E/p is employed, where E represents the cluster energy and p the momentum of the matched track. This allows to foresee the energy released in the calorimeter by the charged particle based on the track measured momentum. The process of determining this parametric expression is known as E/p calibration [74], which has been performed on data collected with a minimum bias trigger configuration and specific criteria to discard backgrounds.

Now that χ^2_m and E/p are defined, it is possible to provide the selection criteria of non-isolated neutral particles [72, 73]:

- χ^2_m of ECAL clusters:
  - $\chi^2_m < 25$ for HCAL clusters below 10 GeV;
  - $\chi^2_m < 16$ for HCAL clusters above 10 GeV;
- ECAL and HCAL clusters are grouped not to share any matched track;
- the expected energy released in the calorimeters by the tracks pointing to each group, E_{exp}, is evaluated from the analytic expression for E/p;
- if the measured cluster energy, E_{meas}, is smaller than $1.8 \times E_{exp}$, the cluster group is discarded;
- if $E_{meas} > 1.8 \times E_{exp}$, E_{exp} is subtracted from E_{meas}.

The remaining energy is a candidate non-isolated neutral particle. If it has $E_T < 2$ GeV it is discarded, else it is added to the list of selected particles.

3.1.2 Jet clustering: the anti-k$_t$ algorithm

Given the variety of applications of jets, there is no single best method for defining them. Over the 30 years passed since their initial measurement proposal [75], numerous jet definitions have been developed. In any case, jet reconstruction algorithms cluster the final-state particles of an event into jets, with a set of parameters that quantify how close two particles must be for them to be grouped into the same jet. Jets are experimental observables containing significant and useful information about the partons emitted in the hard process, with the ultimate target of reducing the complexity

of the final-state objects while preserving the interesting physics information. A jet algorithm should also describe the QCD processes that partons undergo in the hadronization phase.

There are two fundamental features for well-behaved jet algorithms, leading to results that are not significantly altered by the presence of low-energy particles (infrared safety) or by particles that are very close in direction (collinear safety). The importance of infrared and collinear safety lies in ensuring that the theoretical pQCD predictions of jet properties are reliable and can be directly compared to experimental data. Algorithms that lack these properties can produce results that are highly sensitive to small changes in the particle configuration, making it difficult to achieve comparisons. In detail, *infrared safety* means that the results of the algorithm are not sensitive to the emission of soft gluons, i.e. the addition or removal of a very low-energy particle does not significantly change the reconstructed jets. The emission of soft gluons is a common occurrence in the QCD framework, so it is crucial to have infrared safe algorithms. On the other hand, *collinear safety* ensures that the algorithm results are not sensitive to collinear splittings, where a hard parton splits into two or more partons with similar momenta. This is important because partons often undergo collinear splitting, leading to a cascade of particles moving in nearly the same direction.

In LHCb, the clustering of the list of input particles is handled using the `FastJet` [76] interface, which enables a fast execution of the infrared and collinear safe anti-k_t algorithm [77]. It belongs to the k_t family of sequential recombination algorithms, which iteratively cluster particles into jets based on their relative distance in momentum space. The most important sequential recombination algorithms, besides anti-k_t, are k_t [78, 79] and Cambridge/Aachen [80, 81].

The steps of the k_t algorithms are:

1. introducing distances d_{ij} between any two particles i and j, and d_{iB} between each particle i and the beam B:

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{3.1}$$

$$d_{iB} = k_{ti}^{2p} \tag{3.2}$$

   where k_{ti} is the transverse momentum of the i-th particle, η_i its rapidity, ϕ_i its azimuth, and $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the separation of particles i and j in the η-ϕ metric; R is the value of the cluster cone radius, and the number p governs the relative power of the energy with respect to the geometrical scale (ΔR_{ij});

2. taking the smallest distance; if it is a d_{ij}, particles i and j are combined into an overall particle, while if it is d_{iB}, i is identified as a jet and removed from the list of particles;

3. repeating step 1. and 2. until all particles in the event are clustered into jets.

The free parameters which determine the algorithm strategy are ΔR_{ij} and R, whereas the kind of algorithm of the family is defined depending on p:

- $p = 1$ in the case of the k_t algorithm:

$$d_{ij} = \min\left(k_{ti}^{2}, k_{tj}^{2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{3.3}$$

$$d_{iB} = k_{ti}^{2} \tag{3.4}$$

  so particles with similar transverse momenta are combined;

- $p = 0$ in the case of the Cambridge/Aachen algorithm:

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \tag{3.5}$$

$$d_{iB} = 1 \tag{3.6}$$

  so only the geometric separation ΔR_{ij} in the η-ϕ plane is considered, without the k_t weighting;

- $p = -1$ in the case of the anti-k_t algorithm:

$$d_{ij} = \min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{3.7}$$

$$d_{iB} = \frac{1}{k_{ti}^2} \tag{3.8}$$

In general, when it comes to soft radiation, the behavior of any $p > 0$ algorithm is similar to the k_t one: the ordering between particles must be preserved, and indeed it is for all positive values of p. The behavior of all $p < 0$ algorithms is also similar.

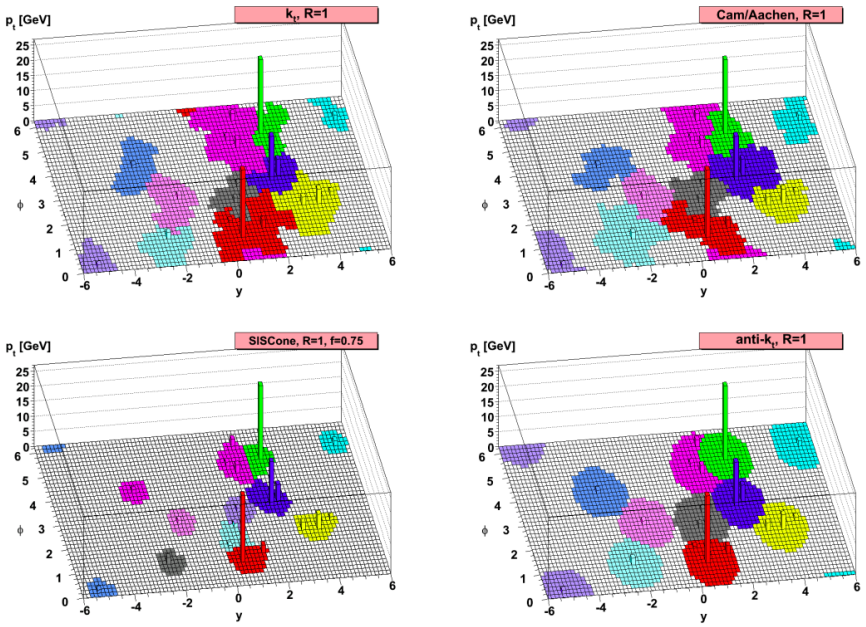

Figure 3.2: A simulated parton-level event with many random soft *ghosts*, clustered with four different jets algorithms. The active catchment areas of the resulting hard jets are illustrated [77]. In the bottom left SISCone panel, f stands for the overlap parameter, which manages the decision of splitting or merging two nearby jets.

The properties of the anti-k_t algorithm can be understood by means of the following simple cases.

- Two particles, 1 and 2, in the range $\Delta R_{12} < R$ are clustered together in a jet.
- The distance d_{ij} between a hard particle and a soft particle is dominated by the momentum of the hard particle. If there are no other hard particles within a distance $2R$ from the first hard particle, then all soft particles within a radius R from the first hard particle are clustered with it to form a jet with radius R.
- If there is another hard particle in the range $R < \Delta R_{12} < 2R$ from the first hard particle, then two jets are formed. The hardest particle of the couple forms a conical jet, and the softest one forms a non-conical jet. If, instead, the two hard particles have similar transverse momenta, then neither jet is conical.
- If there is a collinear splitting of a parton, k, into particles i and j, then the ΔR_{ij} value after the splitting is 0, and the particle formed by recombining i and j is the same as the particle present before the collinear splitting, k. The collinear splitting of partons, therefore, has no effect on the clustered jets, that is anti-k_t is collinear safe.
- The emission of a soft particle from a hard parton with non-zero angular separation does not modify the kinematic properties of the jet. This is precisely the concept of an infrared safe algorithm, that can be well defined at all pQCD orders.

Jet shapes are established by adding many soft particles to the event across the η-ϕ metric. The addition of soft partons does not alter the shape or kinematics of the jets in the event, as the algorithm is infrared safe. The shape of jets reconstructed using different algorithms is compared in figure 3.2. One of the nice features of the anti-k_t algorithm is that it tends to produce more conical jets, with a nearly circular cross section, compared to other k_t-algorithms and cone algorithms, like SISCone (the `C++` implementation of the Seedless Infrared-Safe Cone algorithm [82]), as shown in the bottom left plot. This is why anti-k_t has become one of the mostly used clustering methods nowadays.

It is worth discussing that the reconstructed jets depend on the stage of a pp collision being considered. In particular, jets reconstructed at the parton level, originated from final-state partons produced in a matrix-element-level calculation and any subsequent parton shower, differ from jets reconstructed after hadronization and fragmentation, corresponding to those that would be reconstructed in an ideal detector. Multiple effects give rise to these differences. The underlying event leads to particle production throughout the whole η-ϕ plane, so including it in the clustering increases the jets momenta. This outcome is smaller for smaller R values. During hadronization, the final-state partons cluster into particles that subsequently decay. This can cause particles formed from partons within a jet at the parton level to fall outside the jet at the hadron level, reducing the momentum of jets at the hadron level compared to that at the parton level. This reduction is less significant for larger cone sizes R.

The previously described processes also generate mass for jets, even for jets which at the parton level consist of one massless parton (light quarks are commonly deemed massless, as their mass is much lower than the probed energy scales). These effects are stronger for gluon-initiated jets than quark-initiated jets. Basically, as they carry more color charge, the amount of radiation in gluon-initiated jets tends to be greater than in quark-initiated jets. There is also a dependence on the momentum and type of emitted parton, with a broad minimum in the region between $R = 0.5$ and $R = 0.7$.

3.1.3 Jet four-momentum

To attribute a momentum to a reconstructed jet, a recombination scheme [83] must be adopted. The jets identified by the LHCb jet clustering algorithm are assigned a four-momentum in the energy recombination scheme, according to which the kinematic variables of a cluster of particles are derived by direct addition of the four-momenta of the individual particles:

$$p^{\mu}_{jet} = \sum_i p^{\mu}_i \tag{3.9}$$

where i is the index running on the particles clustered in the jet. Therefore, since $p^{\mu}_{jet} = (E_{jet}, \vec{p}_{jet})$, one gets an expression for the energy and the three-momentum as well:

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

The resulting jets have a clearly defined mass, hence this recombination scheme allows to distinguish transverse energy $E_{T,jet}$ from transverse momentum $p_{T,jet}$, and pseudorapidity η_{jet} from rapidity y_{jet}. Auxiliary quantities can be computed starting from the components of the four-momentum: the angle between the jet three-momentum and the positive direction of the beam line θ_{jet}, the transverse energy $E_{T,jet}$, and the pseudorapidity η_{jet}. All these variables are listed here:

$$p_{T,jet} = \sqrt{p^2_{x,jet} + p^2_{y,jet}}, \qquad \phi_{jet} = \arctan\left(\frac{p_{x,jet}}{p_{y,jet}}\right), \qquad y_{jet} = \frac{1}{2}\ln\left(\frac{E_{jet} + p_{z,jet}}{E_{jet} - p_{z,jet}}\right), \tag{3.11}$$

$$\theta_{jet} = \arccos\left(\frac{p_{z,jet}}{\sqrt{p^2_{x,jet} + p^2_{y,jet} + p^2_{z,jet}}}\right), \quad E_{T,jet} = E_{jet}\sin\left(\theta_{jet}\right), \quad \eta_{jet} = -\ln\left[\tan\left(\frac{\theta_{jet}}{2}\right)\right] \tag{3.12}$$

3.1.4 Jet selection efficiency

Since jets typically contain a large number of particles, the probability that a MC jet is reconstructed is $\sim$ 99%. However, a small percentage of the reconstructed jets are *fake jets*, built from neutral particles produced in pile-up interactions, or from badly measured particles. To reject fake jets, and optimize the fake/good jet separation, the following criteria are imposed to select candidate jets [69].

- $n_{pointing} \geq 2$: at least two tracks must point to the Primary Vertex associated to the jet;
- mtf < 0.75: there must not be a charged particle carrying more than 75% of the total jet p_T;
- mpt > 1.8: the jet must contain a track with $p_T > 1.8\,\text{GeV}$;
- cpf > 0.1: the fraction of charged particles inside the jet must exceed 0.1.

The first requirement alone allows to reject about 40% of fake jets, removing only 3% of good jets. The second criterion discards fake jets made of one single particle, usually a poorly reconstructed one. The third condition removes fake jets made of one single calorimeter cluster, often having low p_T.

The jet selection efficiency ε_{jet} [73], that is the fraction of MC jets associated to a reconstructed jet passing the selection requirements, is shown in figure 3.3. It has been determined on a simulated $Z^0 \to (\mu^+\mu^-) + jets$ sample produced with Run 1 conditions at a center of mass energy $\sqrt{s} = 7\,\text{TeV}$. To obtain a ε_{jet} for data, the simulation results have been corrected by weighting MC events to match the MC jet multiplicity distributions to data. The weights are a function of jet p_T.

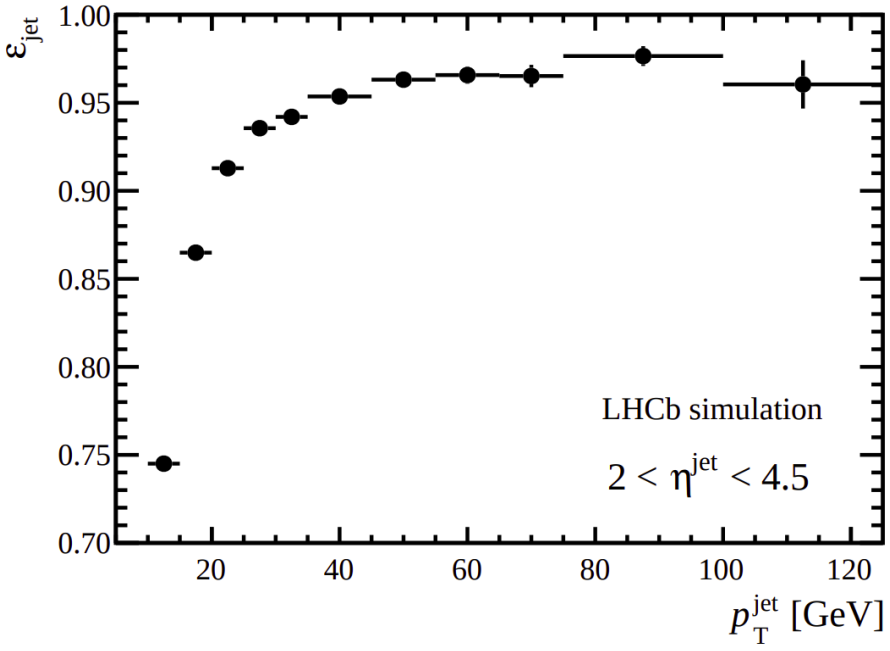

Figure 3.3: Jet selection efficiency as a function of the jet p_T [73].

3.1.5 Monte Carlo validation of the reconstructed jets

At this point, Monte Carlo (MC) simulated samples can be used to validate the quality of the applied jet reconstruction. MC jets are defined as jets clustered with the same anti-k_t algorithm, but with stable MC particles as inputs instead of the reconstructed particles. MC particles carry true kinematical values, as they are not affected by any detector effect. To avoid biases when reconstructing energy, invisible particles like neutrinos are removed from the list of inputs. A MC jet is associated to a reconstructed jet if it is the closest jet lying within a distance from the reconstructed jet in the η-ϕ plane below 0.4, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$.

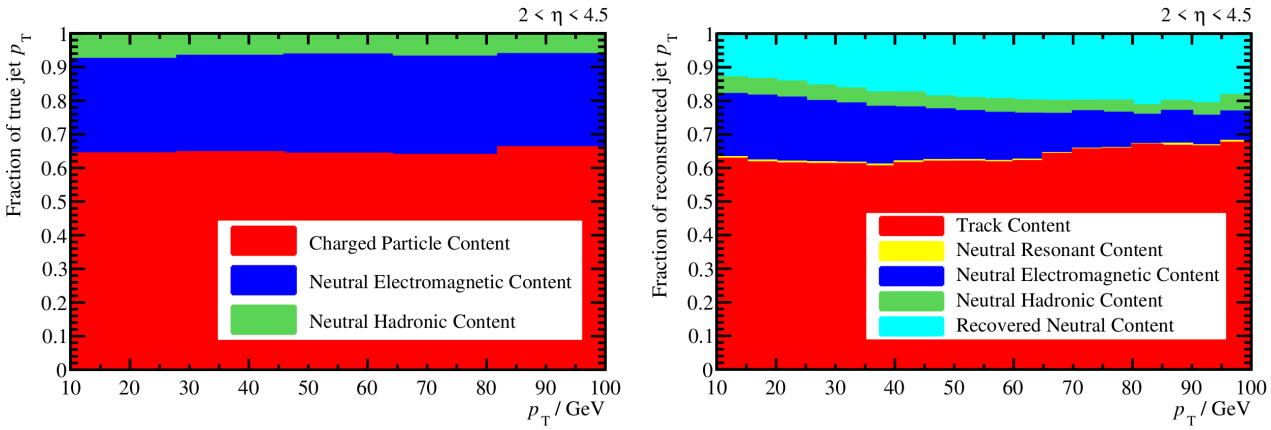

Figure 3.4: Mean fraction of the jet p_T carried by different types of particles within the jet, at MC or truth level (left); mean fraction of the jet p_T carried by different types of particles within the jet in simulation at reconstructed or particle level (right) [69].

The p_T fraction of light jets carried by different particle species at MC level as a function of the particle p_T is shown in the plot of the top of figure 3.4. About 60% of jet constituents are charged particles, about 30% are neutral electromagnetic particles (π^0 and photons), and about 10% are neutral hadrons. Comparing this behavior with the plot on the bottom of figure 3.4, it can be seen that the particle content at reconstructed level has a larger dependence on jet p_T. In this plot, *Neutral Resonant Content* is the contribution from neutral resonances decayed to tracks, while *Recovered Neutral Content* is the contribution from non-isolated neutral particles. In particular:

- the isolated neutral particle content decreases with increasing p_T, as at higher p_T jets become more collimated and contain more particles, so fewer neutral clusters are isolated; consequently, the content from neutrally recovered non-isolated particles increases;
- due to the jet selection criteria discussed in the following section 3.1.4, the charged particle content at the lowest jet p_T slightly decreases with increasing p_T;
- above 30 GeV the track content increases, because part of the jet neutral energy is lost - with no possibility of being recovered - due to the saturation of calorimeter clusters;
- at low p_T, the fraction of reconstructed jet p_T associated to tracks is about 62%, while at truth level it is about 65%; this disagreement is due to track detection inefficiencies.

3.1.6 Jet Energy Correction

The energy of reconstructed jets, E_{jet}^{reco}, differs from the energy of the associated MC jets, E_{jet}^{MC}. To correct the reconstructed jet energy to its true value, a multiplicative correction factor, k_{MC}, is obtained from simulation:

$$E_{jet}^{MC} = k_{MC} \times E_{jet}^{reco} \tag{3.13}$$

Given that the MC jet direction is close to the reconstructed jet one, the same correction factor is applied to each jet four-momentum component. k_{MC} accounts for pile-up effects, non-uniformity and inefficiencies of the detectors, and contributions from background processes. Also, errors coming from the built-in limitations of simulations, both in terms of interaction physics and detector models, are eliminated. k_{MC} is evaluated in simulated events of jets from b-, c-, light quarks and gluons with Run 2 conditions. It is parameterized as a function of the jets p_T in bins of η, ϕ, cpf (the fraction of charged particles in the jet), and number of Primary Vertices.

In the context of this thesis, two additional corrections are applied to take into account possible differences between the jet energy in MC and real data: they refer to Jet Energy Scale (JES) and Jet Energy Resolution (JER), and they are measured in chapter 6.

Chapter 4

Muons at LHCb Run 2

Besides jets, another relevant physics object for the purposes of this thesis is the muon. One of the data sets analyzed in this work, used as a calibration sample, contains events with a Z^0 boson decaying into a $\mu^+\mu^-$ pair. This sample will be used, as discussed in section 6.4.1, to extract the correction factors for the MC jets that account for differences between experimental and MC jets due to detector non-uniformities and inefficiencies. Specifically, these correction factors will contribute to the estimation of α_s in the form of an efficiency that propagates inside its systematic uncertainty.

This chapter offers an overview of the techniques adopted by LHCb in reconstructing muons. Indeed, muons are involved in a large fraction of physics processes happening at LHCb, e.g. studies of CP violation through semileptonic decays, and branching fraction measurements of rare decays. To improve the efficiency and purity of identified muons, during the Long Shutdown 1 between Run 1 and Run 2 the muon identification algorithm [84, 85] has been re-optimized. A unified code, that can be run both online in the HLT and in the offline reconstruction, has been implemented.

4.1 Muon reconstruction

4.1.1 Muon identification variables

It has just been specified that the identification of muons in LHCb is primarily based on the five detecting stations of the muon system (refer to section 2.1.3.3). Muons are highly penetrating particles: they are the only ones that survive all layers of the calorimeters and leave signals in the muon chambers. Exploiting the muon detector information, various identification variables were employed during LHCb Run 2 [86]. The first one is the boolean decision called `IsMuon`, which extrapolates a long or downstream track (for the definition of track types, refer to section 2.1.2.4) within the acceptance of the muon system, yielding the track coordinates in the M2-M5 stations. A search for hits is performed in an elliptic, momentum-dependent region around the extrapolated track in each station. The track satisfies the `IsMuon` hypothesis if a minimum number of stations, depending on the track momentum, have at least one hit within the elliptic region. Table 4.1 shows the required stations and their dependence on the track momentum. The minimum momentum of $p > 3\,\text{GeV}$ is necessary for a muon to reach the M3 station - included in all selection criteria - without being absorbed, thus serving as a threshold for classifying a track as a muon.

The other boolean variables are `IsMuonLoose`, which, as shown in table 4.1, requires a fewer amount of hits with respect to `IsMuon`, and `IsMuonTight`, which requires the same amount of hits, but using only crossed hits, that are signals resulting from the OR of horizontal and vertical readout pads.

p [GeV]	`IsMuon`	`IsMuonLoose`
$p < 3$	False	False
$p < 6$	M2 & M3	At least two of M2-M4
$6 < p < 10$	M2 & M3 & (M4 || M5)	At least three of M2-M5
$p > 10$	M2 & M3 & M4 & M5	At least three of M2-M5

Table 4.1: Stations required to have hits inside the elliptic region for the muon identification boolean variables `IsMuon` and `IsMuonLoose` and their dependence on the track momentum [86].

4.1.2 Muon reconstruction efficiency

The muon efficiency is decisive in computing the production cross section of processes such as $Z^0 \to \mu^+\mu^-$, as it is an important source of systematic uncertainties. The muon reconstruction efficiency in LHCb, ε_{rec}, is factorized into three components [87]:

$$\varepsilon_{rec} = \varepsilon_{trg} \times \varepsilon_{trk} \times \varepsilon_{id} \tag{4.1}$$

where ε_{trg} is the muon trigger efficiency, ε_{trk} is the track reconstruction efficiency, and ε_{id} is the muon identification efficiency. All these contributions are determined with the highly pure $Z^0 \to \mu^+\mu^-$ real events. The *tag-and-probe* method is used to estimate the three efficiencies, according to which one of the muons, the *tag*, is required to be triggered by dedicated muon trigger lines and positively identified as a muon, while the efficiency is measured with the second muon, the *probe*.

The tag-and-probe technique is also exploited to build the Z+jet calibration sample that will be used in this thesis work, detailed in section 6.3: the Z^0 boson is the tag, and the jet is the probe. LHCb identifies muons with very high efficiencies, as visible in figure 4.1, and this reflects in a clear Z^0 signature, which in turn allows to precisely reconstruct the accompanying jet.

4.1.3 Muon trigger

Both levels of the LHCb Run 2 trigger system (see section 2.1.4) are involved in recognizing muon signals. Also in the following description, the focus is put on muons used to make the Z+jet samples.

- The hardware-L0 muon trigger identifies high transverse momentum muons in the event using information from the muon stations. This trigger level is crucial for selecting events where a Z^0 boson decays into a pair of muons, as it quickly reduces the data rate while preserving events of interest.
- The software-HLT further refines the selection. The HLT uses full event information, including detailed tracking and calorimetry, to confirm the presence of high-p_T muons and ensure that they match the criteria for a $Z^0 \to \mu^+\mu^-$ decay. The HLT is responsible for selecting events with high precision, reducing the event rate to a manageable level for storage and analysis.

The high efficiency and precision of the muon trigger system at LHCb ensure that the $Z^0 \to \mu^+\mu^-$ decays are accurately selected. This makes the Z^0 boson an excellent tag in the tag-and-probe method, allowing to study the properties of the associated jet - the probe - as discussed in the previous section.

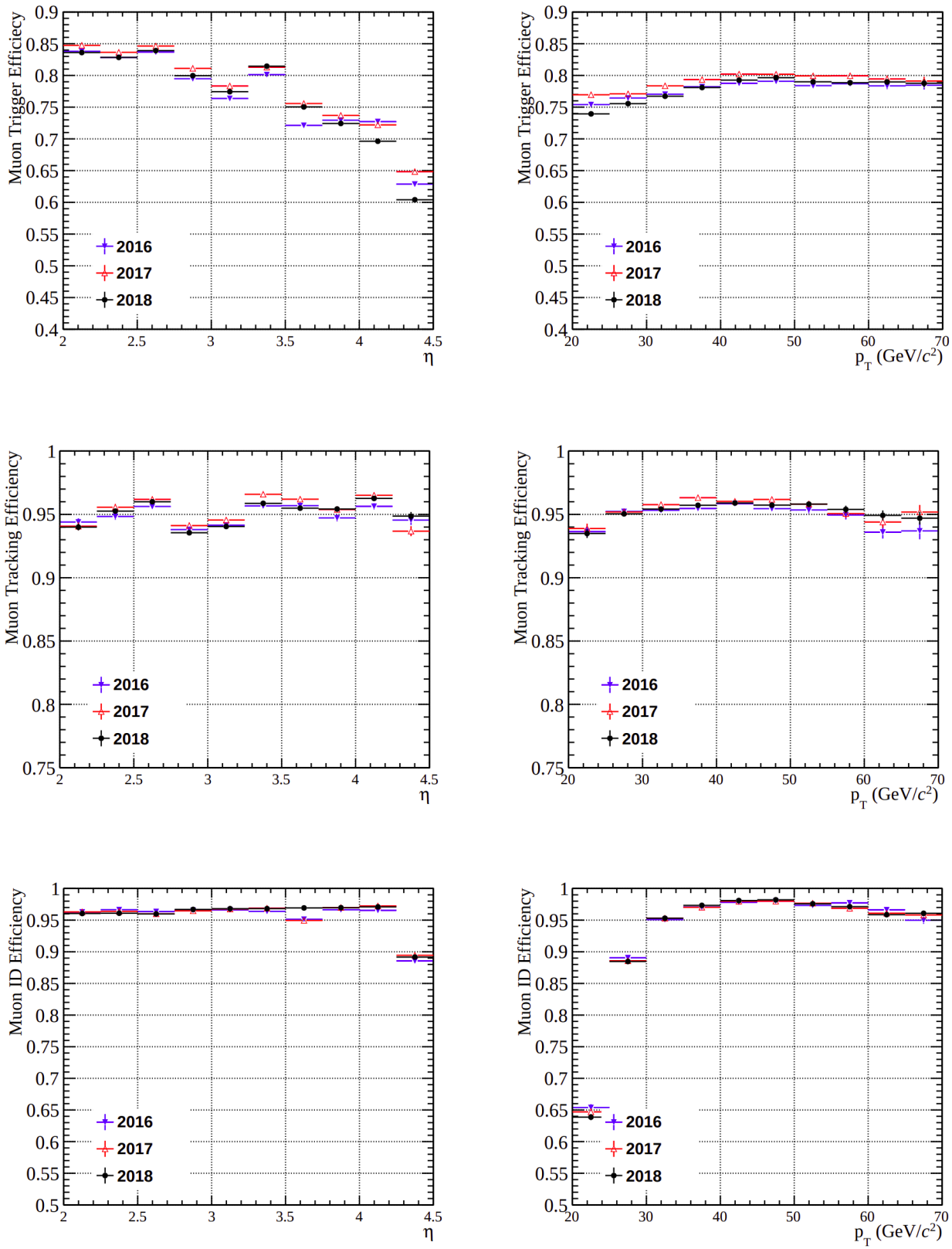

Figure 4.1: From top to bottom: muon trigger efficiency, muon tracking efficiency, and muon identification efficiency for events collected in 2016, 2017 and 2018 (LHCb Run 2) as a function of muon η and p_T [87].

Chapter 5

Simulation of dijet properties in the forward region

In this thesis work, a fast simulation has been developed in order to determine which observables related to dijet systems in the forward region are the most sensitive to the variation of α_s. Simulations of pp collision events at $\sqrt{s} = 13\,\text{TeV}$ - the energy condition of Run 2 - producing two jets have been implemented. The generation of events is handled with the program `MadGraph5_aMC@NLO` [88], and the hadronization with `Pythia8` [56]. Jets are reconstructed using the `FastJet` interface [76] with the canonical LHCb parameters, and selected according to the LHCb acceptance. Several variables of interest regarding each jet taken individually and possible multi-jet systems are extracted. In the following sections, the simulation process is outlined in full detail.

5.1 Events generation

The production mechanisms of two jets from a pp collision are governed by the strong interaction, so all the involved processes are mediated by the gluon. Specifically, they are:

- quark-quark scattering, $qq' \to qq'$, in which the quarks, one from each proton, exchange a gluon, resulting in two final-state quarks that form jets;
- quark-antiquark scattering, $q\bar{q} \to q'\bar{q}'$, in which a quark and an antiquark scatter off each other, producing another quark and an antiquark, that hadronize into jets;
- quark-antiquark annihilation into a gluon pair, $q\bar{q} \to gg$, in which the quark and antiquark annihilate via a gluon, which then splits into two gluons, leading to the formation of two jets;
- gluon-gluon scattering, $gg \to gg$, in which the two initial-state gluons scatter exchanging a gluon, generating two gluons in the final state, which become jets;
- gluon-gluon annihilation into a quark-antiquark pair, $gg \to q\bar{q}$, in which two gluons collide and yield a quark-antiquark pair, with subsequent hadronization into jets;
- quark-gluon scattering, $qg \to qg$, in which the quark and the gluon interact through the exchange of a gluon, which results in the final-state quark and gluon, giving origin to jets.

The corresponding Feynman diagrams are drawn in figure 5.1. Only one channel per process is shown.

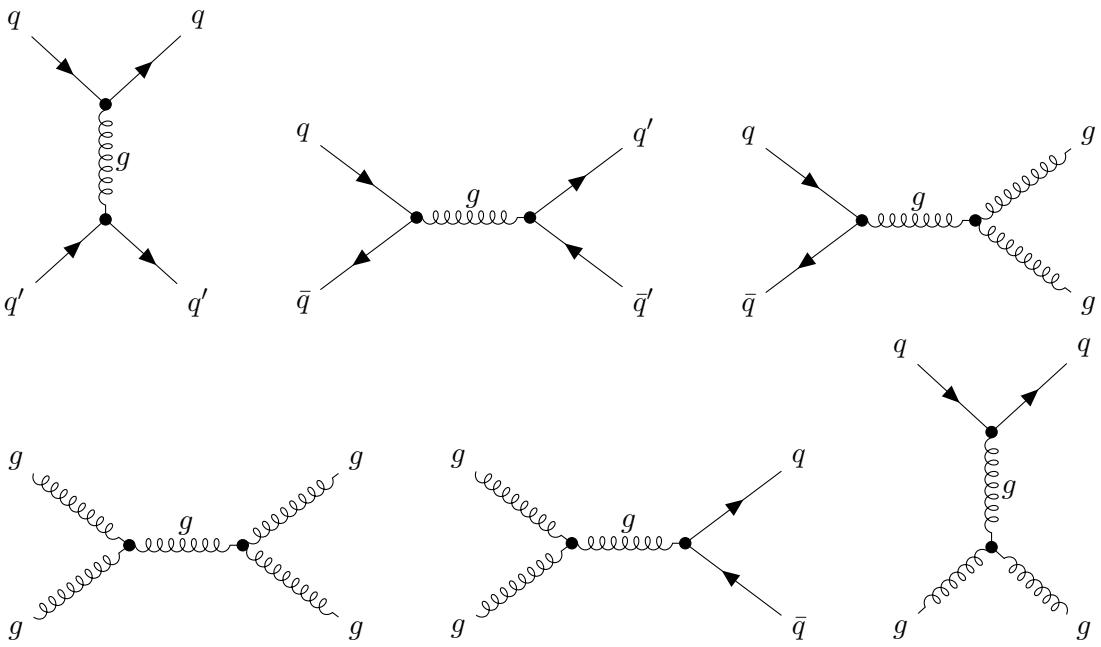

Figure 5.1: The LO Feynman diagrams for two jets production.

To simulate all these processes, events of the type

$$p\ p \to j\ j \tag{5.1}$$

are generated at LO in the `MadGraph5_aMC@NLO` [88] environment. Partons p inside the colliding protons and partons j that will originate jets are defined as, respectively:

$$p = g\ u\ d\ c\ s\ \bar{u}\ \bar{d}\ \bar{c}\ \bar{s} \tag{5.2}$$

$$j = g\ u\ d\ c\ s\ b\ \bar{u}\ \bar{d}\ \bar{c}\ \bar{s}\ \bar{b} \tag{5.3}$$

that is, the partons of the proton considered for the interactions are gluons, quarks and antiquarks, excluding b, $\bar{b}$, t and $\bar{t}$, and they are the objects to which `MadGraph5_aMC@NLO` applies the selected PDF that regulates how much each parton is active inside the proton, hence in the `MadGraph5_aMC@NLO` interface they are collectively called "proton". A jet, instead, is originated at parton level by a gluon, a quark or an antiquark, with the exception of t and $\bar{t}$.

Parameter	Value
Number of events	10^6
$p_{T,\,min}(j)$	10 GeV
$\eta_{max}(j)$	5
j-j separation ΔR	0.4

Table 5.1: Parameters for the events generation of $pp \to jj$ events. ΔR is the separation between two j's in the η-ϕ plane, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The parameters for generating the events are listed in table 5.1. The minimum transverse momentum for the parton j, $p_{T,\,min}(j)$, has been set to 10 GeV to allow a comparison between simulated and experimental jets, which in LHCb have a fiducial cut $p_T(jet) > 20$ GeV. This choice has been done because hadronization fluctuations and detector effects may lead to reconstructed jets with a transverse momentum different from the generated one. In other words, partons generated with a transverse momentum smaller than 20 GeV can be reconstructed as jets with transverse momentum larger than 20 GeV. Imposing a lower $p_{T,\,min}$ for the partons at the simulation level, then, allows to keep these jets that otherwise would be incorrectly discarded. The choice of the threshold for $p_{T,\,min}$ must also consider the computational time needed to achieve the desired statistics after the jets selection

requirements. Similar considerations can be formulated for the maximum pseudorapidity of the parton j, $\eta_{max}(j)$, remembering that the axis of a jet cone is identified with the direction of the jet three-momentum, and η is related to the angle of the jet axis with respect to the incoming beam axis, z. The value of $\eta_{max}(j)$ it is set to 5, that is the maximum pseudorapidity covered by LHCb, but then reconstructed jets will be selected inside the range $2.2 < \eta(jet) < 4.2$, to ensure that the cones of the reconstructed jets are entirely enclosed in the instrumented LHCb region.

To account for the effect of α_s in the evolution of these events, three sets of 10^6 events each have been generated, each with a specific value of α_s. To do so, in each generation process a PDF set has been loaded. Each PDF does not only carry information about the particular value of α_s, it also dictates the evolution of the simulated process, as it acts at the level of Feynman diagrams depending on the model it is built with. The chosen PDF family is CT10NNLO [22], whose fits are based on HERA combined data and the Tevatron jet data, and have a NNLO level of accuracy. The single PDFs have been downloaded from the LHAPDF6 catalogue [89].

The hadronization of generated partons is simulated by `Pythia8` [56] built inside `MadGraph5_aMC@NLO`. The resulting simulated cross sections of pp collisions producing two jets, σ_{sim}, are reported in table 5.2, where the errors are statistical. This step of the simulation gives an output file in the `HepMC` [90] format, an object oriented event record written in `C++` for High energy physics Monte Carlo event generators, such as `Pythia8`.

PDF set	α_s	σ_{sim} [$\times 10^9$ pb]
CT10nnlo_as_0110	0.1100	4.26194 ± 0.00099
CT10nnlo_as_0118	0.1180	4.6643 ± 0.0011
CT10nnlo_as_0130	0.1300	5.3889 ± 0.0014

Table 5.2: Simulated cross sections for the generated $pp \to jj$ events after the hadronization obtained with `Pythia8`, depending on the loaded PDF set, and so on the α_s value. The associated errors are purely statistical.

The previous table indicates that cross sections increase as the value of α_s enlarges. According to the discussion carried out in section 1.3, Parton Distribution Functions describe the probability of finding a particular parton with a specific fraction of the initial proton momentum, x, and at the same time they regulate the evolution of parton densities as a function of α_s. A larger value of α_s results in higher gluon densities at lower x, because stronger interactions increase the chance of having gluon emissions and gluon splittings. The cross section can be seen as a quantity measuring the probability of occurrence of a specific interaction, which depends on both the parton-level cross sections (which, in turn, depend on α_s) and on the parton densities given by the PDFs. Then, with a higher α_s, the partonic cross sections increase due to the stronger interactions between partons, and, additionally, PDFs with higher α_s predict higher densities of gluons and other partons in the relevant kinematic regions. All these factors mean that more partons are available to participate in the interactions that lead to the production of final states, like the jets, that are the objects analyzed in this thesis. The combined effect of stronger parton interactions and increased parton densities at relevant momentum fractions leads to an overall increase in the simulated cross sections. This is why, when using a PDF with a larger α_s value, an increase in the simulated cross sections is observed.

5.2 Jets reconstruction

The outputs of the hadronization produced by `Pythia8` and saved in the `HepMC` file are read as inputs to the jets reconstruction chain. Jets are reconstructed using the anti-k_T clustering algorithm embedded in the `FastJet C++` implementation, with cone size $R = 0.5$ and η-ϕ separation $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$. The chosen value of R gives reasonably small differences between the hadron level and the parton level stages, and is used also by the CMS experiment, allowing easier comparison with CMS results. The input particles to the clustering algorithm are stable ($\tau > 10^{-8}$ s, no children) hadrons, muons, electrons, and photons. Neutrinos are not included, as LHCb is not hermetic, and so it is not able to reconstruct invisible missing energy. The four-momentum of each jet is assigned in

the energy recombination scheme, by direct addition of the four-momenta of the individual particles. The jet axis is defined as the direction of its three-momentum, and the jet transverse momentum is computed as $p_T(jet) = \sqrt{p_x^2(jet) + p_y^2(jet)}$. In table 5.3 the important information is summarized.

Inputs	Anti-k_t parameters	Jet four-momentum
Hadrons Muons Electrons Photons	$R = 0.5$ $\Delta R = 0.4$ $p_{T,\,min}(jet) = 5\,\text{GeV}$	$p^{\mu}_{jet} = (E_{jet}, \vec{p}_{jet})$ $E_{jet} = \sum_i E_i$ $\vec{p}_{jet} = \sum_i \vec{p}_i$ (i: particle index)

Table 5.3: Variables and parameters used for reconstructing jets in this thesis.

In this step, among all the other objects of an event, only jets are selected, and then saved in a `ROOT` output file.

It is worth pointing out that, even if two partons are generated in the simulated $p\ p \to j\ j$ scattering, the number of final-state reconstructed jets in an event can be larger than two. Indeed, `Pythia8` simulates the initial- and final-state gluon emissions, as well as the gluon splittings and all the other parton-level mechanisms that then lead to jet formation. Hence, `Pythia8` is responsible for the development of the shower of the initial partons, in which many interactions occur, giving rise to the particles which are then clustered into jets. In the parton shower, fragmentation happens, and so many vertices are created, increasing the considered order of α_s, and to each α_s order the presence of a jet is associated.

5.3 Jets selection and jet variables extraction

Starting from the simulated jets four-momentum, the jets pseudorapidity is calculated according to the formula appearing in equation 3.12. All jets in an event are required to have a pseudorapidity in the range of the LHCb instrumented coverage, $2.2 < \eta(jet) < 4.2$.

Several variables regarding each jet taken individually and possible multi-jet systems are extracted, with particular attention to distributions with a potentially interesting behavior with respect to the varying α_s value. Multi-jet systems with two or three jets have been considered. From now on, jets will be labelled 1, 2, or 3. The jet 1 is the leading jet, i.e. the jet with the highest transverse momentum in an event, then jets 2 and 3 have lower transverse momenta: $p_T(jet_1) > p_T(jet_2) > p_T(jet_3)$. Only events with at least one reconstructed jet in the fully instrumented LHCb acceptance are selected, and the number of remaining events after these cuts for each generated sample, N_{sel}, is given in the second column of table 5.4. The third column shows the number of events surviving the additional cut imposed on the transverse momentum of the three leading jets, $p_T(jet) > 20\,\text{GeV}$. It results that about 0.01% of generated events contain at least one jet with $2.2 < \eta(jet) < 4.2$ and $p_T(jet) > 20\,\text{GeV}$.

Every spectrum is scaled by a factor F, which translates the simulated number of events into experimentally foreseen ones. In fact, neglecting detector efficiency contributions for the moment, the number of expected events in a collision, N_{exp}, is:

$$N_{exp} = \sigma \cdot \mathcal{L} \cdot A \tag{5.4}$$

where σ is the cross section of the process under study and $\mathcal{L}$ the integrated luminosity at the considered experiment, and A is the acceptance, a measure of the fraction of events that survive the applied cuts, N_{sel}. At simulation level, the cross section is σ_{sim}, and A is:

$$A = \frac{N_{sel}}{N_{gen}} \tag{5.5}$$

where $N_{gen} = 10^6$ is the number of generated events, while $\mathcal{L}$ is a link to experimental data, and in this case it is $\mathcal{L} = 1.6\,\text{fb}^{-1}$, the integrated luminosity of 2016 LHCb data-taking; this was chosen

because the trigger and stripping lines set used for it were fully available. N_{sel} is the integral of the distribution.

The scale factor F applied to simulated distributions, therefore, is:

$$F = \frac{\mathcal{L} \cdot \sigma_{sim}}{N_{gen}} \tag{5.6}$$

and its values are written in the fourth column of table 5.4. The fifth column contains the number of expected events in a collision, N_{exp}, resulting after scaling with the F factors.

α_s value	$N_{sel}^{\eta \text{ cut}}$	$N_{sel}^{\eta+p_T \text{ cut}}$	F $[\times 10^6]$	$N_{exp} = F \cdot N_{sel}^{\eta+p_T \text{ cut}}$ $[\times 10^8]$
0.1100	619972	111	6.8 ± 0.1	8.0 ± 0.7
0.1180	621619	152	7.5 ± 0.1	12 ± 1
0.1300	622422	156	8.6 ± 0.2	15 ± 1

Table 5.4: The number of events after η and $\eta + p_T$ cuts, for each generated set, is given in columns two and three, respectively. The fourth column contains the scale factors F applied to the three generation runs,and the fifth column shows the number of expected events after the application of the scale factor F.

Some of the most interesting jets distributions are reported below. The errors are statistical, considering the error on the number of entries of a bin N to be the square root of the number of entries itself, $e_N = \sqrt{N}$. The distributions are normalized to the same area in order to identify shape differences among them.

5.3.1 Jets transverse momentum

The distributions of the transverse momentum of the two leading jets, jet_1 and jet_2, with the three values of α_s are plotted in figures 5.2 and 5.3 for the range $20\,\text{GeV} < p_T(jet_{1,2}) < 100\,\text{GeV}$.

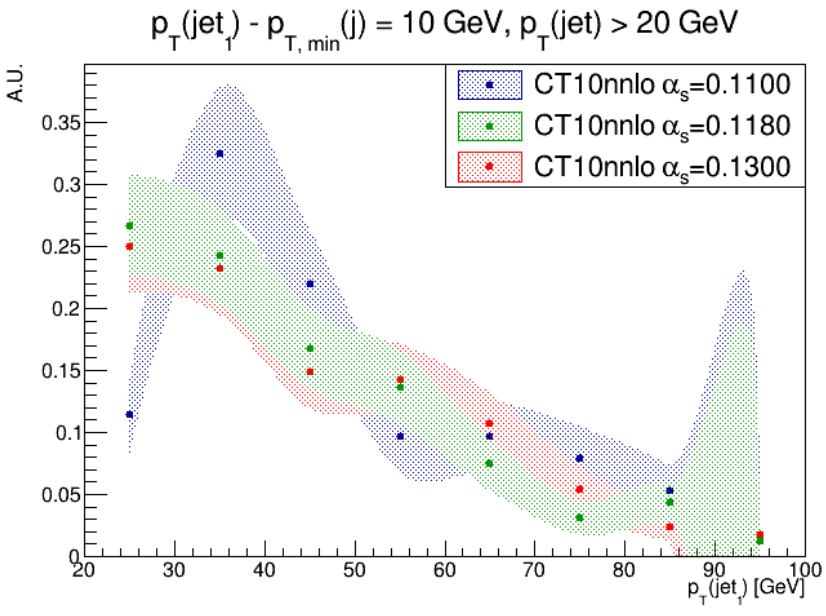

Figure 5.2: $p_T(jet_1)$ distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

As expected, the distribution relative to jet_1 shows larger p_T values than the jet_2 distribution. It is interesting to observe that for both jets the p_T peak shifts towards smaller values as α_s increases. This could be explained by the fact that as α_s gets larger, it accesses lower energy regimes.

The three distributions of $p_T(jet_1)$ and of $p_T(jet_2)$ are compatible within their errors.

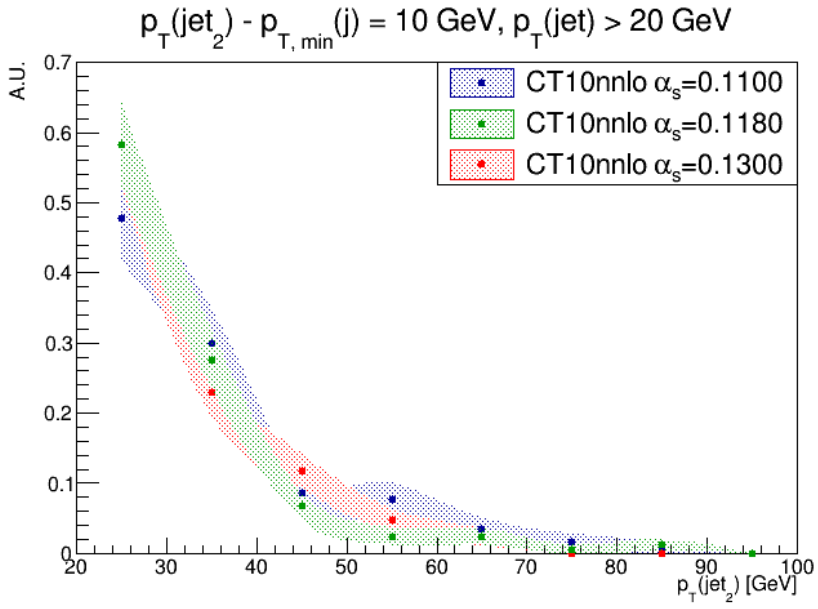

Figure 5.3: $p_T(jet_2)$ distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

5.3.2 Average transverse momentum of dijet systems

A relevant quantity belonging to this category is the average transverse momentum of the two leading jets in the event, Q, defined as:

$$Q = \langle p_T(jet_{1,2}) \rangle = \frac{p_T(jet_1) + p_T(jet_2)}{2} \tag{5.7}$$

The distributions to vary α_s, plotted in figure 5.4 for the range $20\,\text{GeV} < Q < 100\,\text{GeV}$, are a direct consequence of those of the two leading jets taken individually and described in the previous section. Specifically, the Q peak shifts towards smaller values as α_s increases.

Also in this case, the three Q distributions are compatible within their errors.

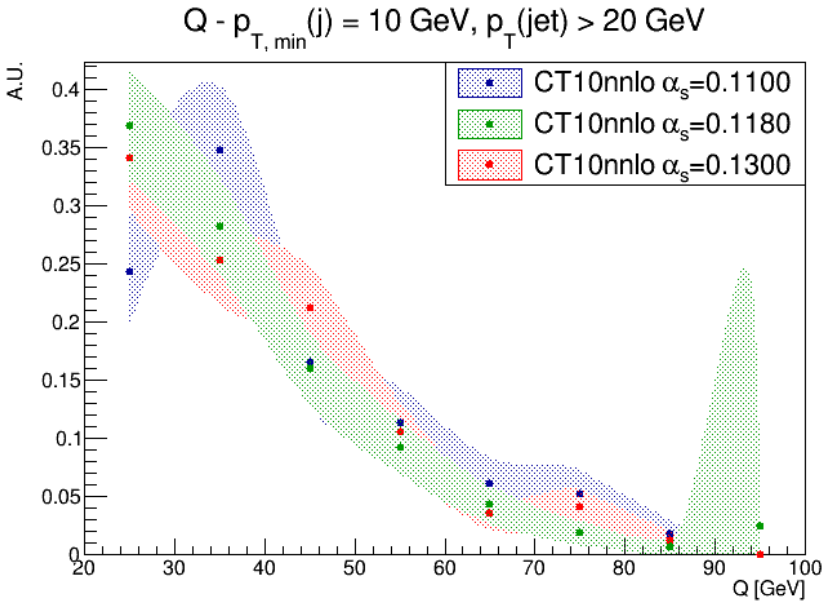

Figure 5.4: Q distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

5.3.3 Multi-jet systems mass

Here, the distributions worth mentioning are the mass of the system formed by the two leading jets, m_{12}, and the mass of the system of the three leading jets, m_{TOT}:

$$m_{12} = \sqrt{\left(E(jet_1) + E(jet_2)\right)^2 - \left(\vec{p}(jet_1) + \vec{p}(jet_2)\right)^2} \tag{5.8}$$

$$m_{TOT} = \sqrt{\left(E(jet_1) + E(jet_2) + E(jet_3)\right)^2 - \left(\vec{p}(jet_1) + \vec{p}(jet_2) + \vec{p}(jet_3)\right)^2} \tag{5.9}$$

The distributions are plotted in figures 5.5 and 5.6 for the ranges $20\,\text{GeV} < m_{12} < 200\,\text{GeV}$ and $60\,\text{GeV} < m_{TOT} < 240\,\text{GeV}$, respectively. Their shapes are broad, and this behavior is consistent with the fact that in QCD multi-jet productions no specific resonances - which would appear as peaks in the invariant mass distribution - are expecte7

Again, the three distributions of m_{12} and of m_{TOT} are compatible within their errors.

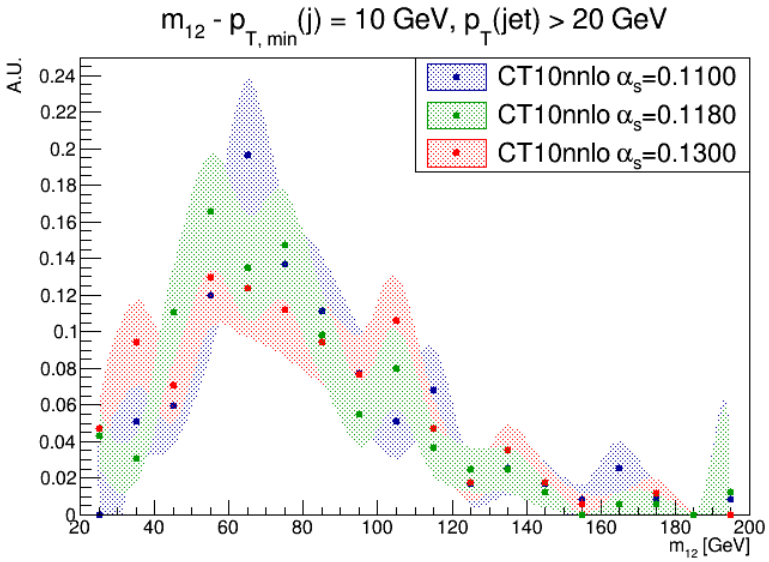

Figure 5.5: m_{12} distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

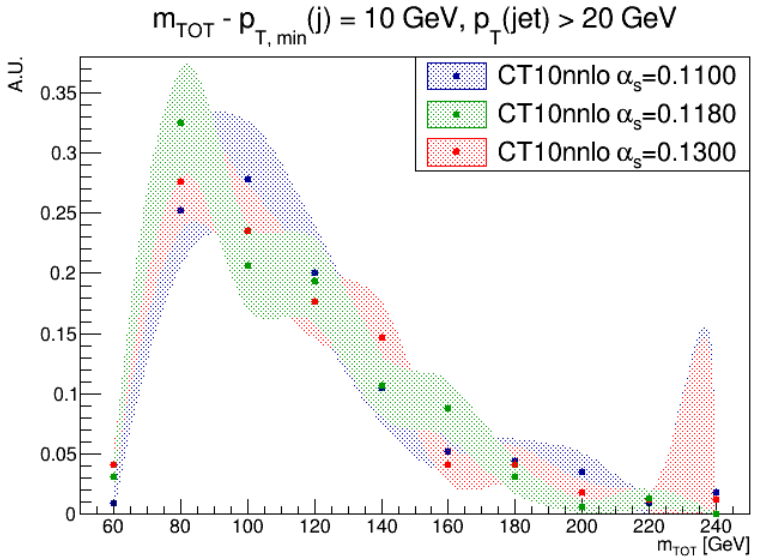

Figure 5.6: m_{TOT} distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

5.3.4 Pseudorapidity separation in dijet systems

The pseudorapidity separation between the two leading jets is considered:

$$\Delta\eta_{12} = \eta(jet_1) - \eta(jet_2) \tag{5.10}$$

where the pseudorapidity of a single jet is defined in equation 3.12. The plot in figure 5.7 reports the three distributions of $\Delta\eta_{12}$ in the range $[-2, 2]$. They are broad and almost symmetric with respect to the peak at $\Delta\eta_{12} = 0$, which means that in most of the events the two leading jets are produced close to each other in the η coordinate. As a consequence, since η is related to the angle relative to the beam axis, θ, a small difference in pseudorapidity indicates that the two jets are emitted at similar angles relative to the beam axis, implying they are close to each other in the detector angular coverage, and possibly produced from a common interaction or the same parent particle. The nearly identical pseudorapidity suggests that the two jets are back-to-back and correlated in other variables, such as the azimuthal angle - which they are (see figure 5.8). Additionally, a small $\Delta\eta_{12}$ suggests that the jets are moving in nearly the same direction in the laboratory frame of reference, providing insights into the event geometry and the kinematics of the underlying event. Pseudorapidity is also related to the detector coverage, so a small $\Delta\eta_{12}$ between two objects means that they are detected within a narrow region of the detector. This impacts the resolution and efficiency of the measurement, with particular relevance for the detector acceptance and performance.

Once more, the three distributions of $\Delta\eta_{12}$ are compatible within their errors.

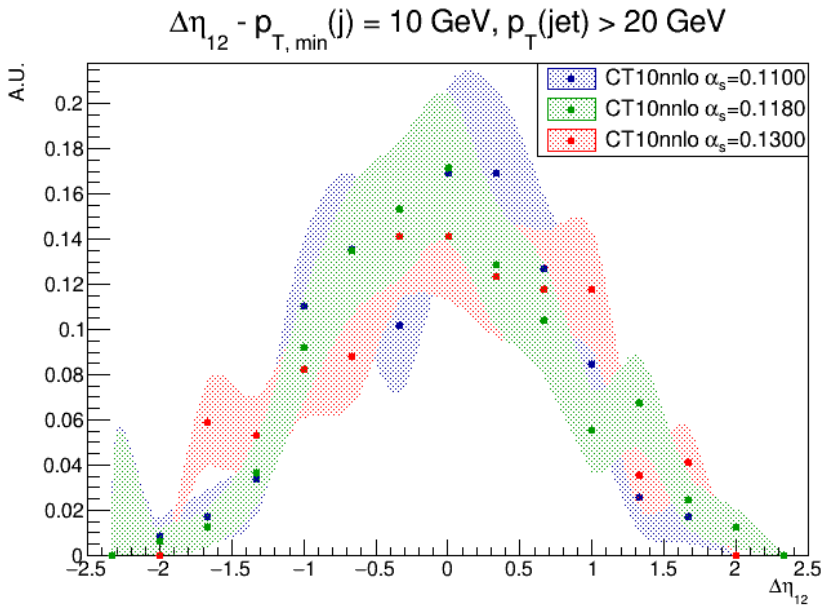

Figure 5.7: η_{12} distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

5.3.5 Azimuthal separation in dijet systems

The azimuthal separation between the two leading jets, $\Delta\phi_{12}$, is studied. In collider physics, the azimuth ϕ is the angle between the three-momentum of an object and the x axis in the xy plane, perpendicular to the direction of the incoming beams, z. In terms of the components of the jets four-momentum, $\Delta\phi_{12}$ is computed as:

$$\Delta\phi_{12} = \arccos \frac{p_x(jet_1) \cdot p_x(jet_2) + p_y(jet_1) \cdot p_y(jet_2)}{\sqrt{\left(p_x^2(jet_1) + p_y^2(jet_1)\right) \cdot \left(p_x^2(jet_2) + p_y^2(jet_2)\right)}} \tag{5.11}$$

In figure 5.8, the $\Delta\phi_{12}$ distributions for the three chosen values of α_s are shown. It is clear that the majority of events contains the two leading jets produced almost back-to-back, $\Delta\phi_{12} \simeq \pi$, confirming

the discussion done for the $\Delta\eta_{12}$ trend. In fact, the two leading jets are the ones carrying the largest amount of the available transverse momentum, thus kinematic considerations impose a π separation in the transverse plane.

Also these three distributions are compatible within their errors.

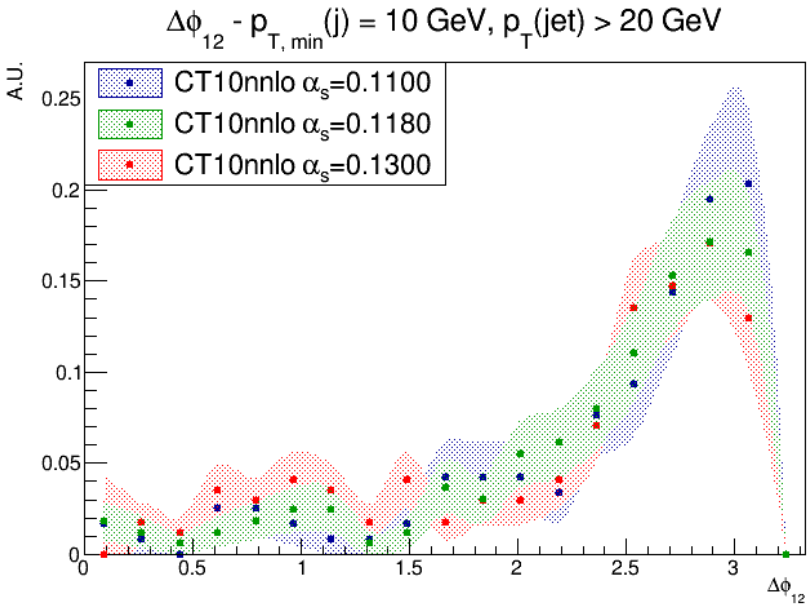

Figure 5.8: $\Delta\phi_{12}$ distributions normalized to the same area with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

5.3.6 R_{32} ratio

Following the lead of the analysis developed by the CMS Collaboration in [91], the first approach adopted for the analysis presented here was to determine α_s from the R_{32} distribution. R_{32} is the ratio of the number of selected inclusive three-jet events to the number of selected inclusive 2-jet events. Therefore, since the number of selected inclusive three- and two-jet events are proportional to α_s^3 and α_s^2, respectively, R_{32} is proportional to $\alpha_s^3/\alpha_s^2 = \alpha_s$. R_{32} is a robust observable for precision α_s measurements mainly because, compared to three- and two-jet cross sections taken individually, it has the advantage that numerous uncertainties cancel in the ratio, at least partially. In fact, ratios of event rates help reduce the sensitivity to experimental uncertainties related to luminosity measurements and detector effects, as well as many theoretical systematic uncertainties linked to the choice of the renormalization and factorization scales, μ_R and μ_F, or to non-perturbative effects.

The resulting R_{32} distributions in bins of Q and also m_{12} are shown in figures 5.9 and 5.10, where the errors of R_{32} are computed by propagation, considering the error on the number of entries of a bin N to be the square root of the number of entries itself, $e_N = \sqrt{N}$. No significant differences in shape can be recognized in the distributions with the three α_s values.

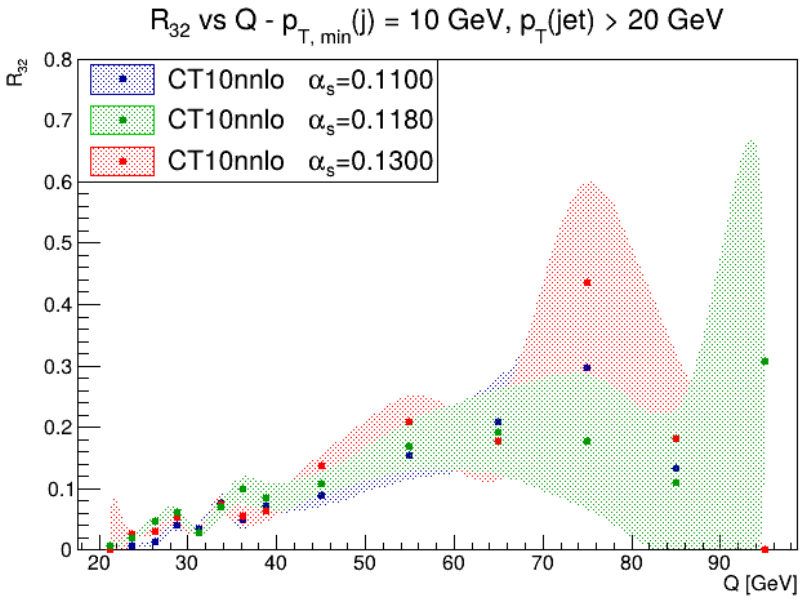

Figure 5.9: R_{32} distributions in bins of Q with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

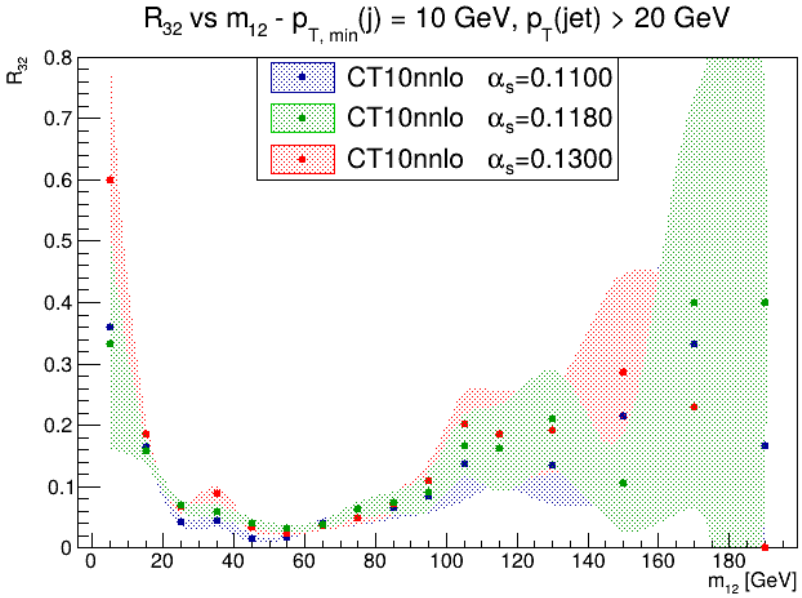

Figure 5.10: R_{32} distributions in bins of m_{12} with the three α_s values. The colored bands are delimited by the errors (statistical) of the distributions.

5.3.7 Inclusive dijet cross section

To extract a cross section for a specific process, its general relation to the number of events connected to that process is exploited:

$$N = \sigma \cdot \varepsilon \cdot \mathcal{L} \tag{5.12}$$

where ε is the efficiency, and $\mathcal{L}$ the integrated luminosity. Hence, the cross section can be extracted by counting the number of selected events. In this work, inclusive events are considered, where *inclusive* means events with a number of jets $\geq n$. To give an example that will be useful later, inclusive Z^0 production refers to events with ≥ 0 jets, and inclusive Z+jet production refers to events with ≥ 1 jet. In contrast, exclusive Z+jet production corresponds to events containing exactly 1 jet.

The inclusive number of jets per event has been extracted from the simulated samples, and they are available in table 5.5. The usual cuts $2.2 < \eta(jet) < 4.2$ and $p_T(jet) > 20\,\text{GeV}$ are applied to

Number of jets	Number of events		
	$\alpha_s = 0.1100$	$\alpha_s = 0.1180$	$\alpha_s = 0.1300$
≥ 0	960629	959387	958091
≥ 1	36396	37519	38597
≥ 2	2857	2933	3146
≥ 3	111	152	156
≥ 4	6	9	10

Table 5.5: Number of events with a minimum number of jets from 0 to 4 for each generation run. The cuts $2.2 < \eta(jet) < 4.2$ and $p_T(jet) > 20\,\text{GeV}$ are applied.

all jets in the event. The number of events for each sample is then scaled by the relative factor F computed as shown in equation 5.6. In figure 5.11 the results are plotted. It can be seen that the number of produced jets per event decreases as the value of α_s decreases. Actually, the strong coupling constant governs the strength of the interaction between quarks and gluons in QCD: a higher α_s value results in quarks and gluons interacting more strongly, leading to more frequent gluon emissions and splittings, which are the processes responsible for generating jets. Jets, in fact, are formed from the fragmentation of high-energy quarks or gluons, and the probability of these emissions increases together with the value of α_s. Conversely, when α_s is reduced, the likelihood of gluon emissions and parton splittings decreases: fewer partons are produced in the final state, and consequently the probability of forming additional jets is reduced, because there are fewer high-energy partons available to initiate jet formation. With fewer emissions, an event is less likely to produce complex, multi-jet topologies, resulting in simpler final states with fewer jets.

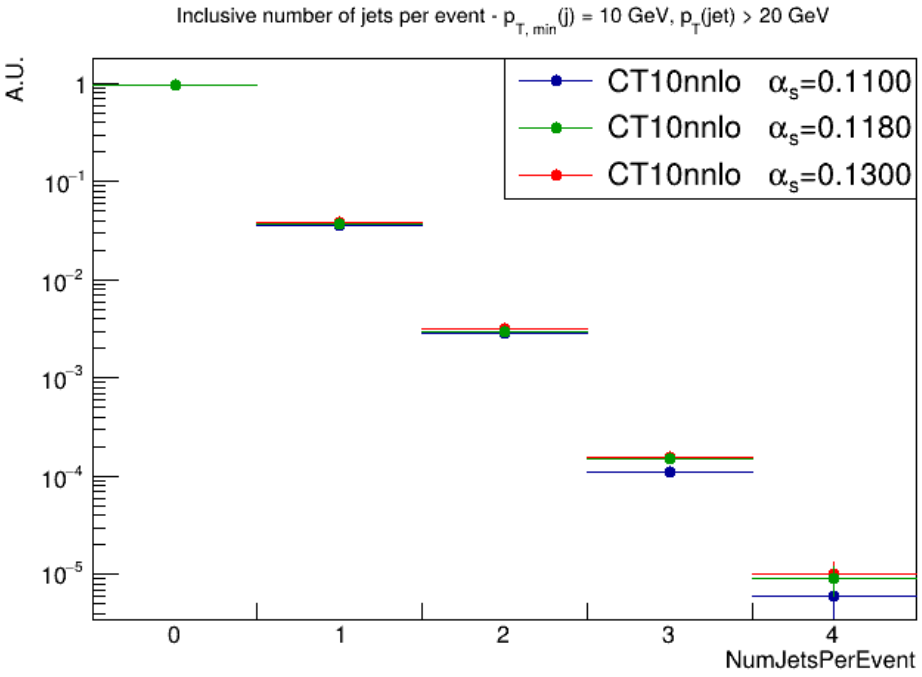

Figure 5.11: Number of events with a minimum number of jets for each generation run. The cuts $2.2 < \eta(jet) < 4.2$ and $p_T(jet) > 20\,\text{GeV}$ are applied, and the number of events is corrected with the relative F factor.

Then, for each bin of plot 5.11, the acceptance is computed as the ratio of the number of events with jets satisfying the $2.2 < \eta(jet) < 4.2$ and $p_T(jet) > 20\,\text{GeV}$ cuts, $N_{sel,nj}$, to the number of generated events, N_{gen}:

$$A_{nj} = \frac{N_{sel,nj}}{N_{gen}} \tag{5.13}$$

These acceptances are used to weight the simulated cross sections, obtaining the cross sections for processes including a number $\geq n$ jets, σ_{nj}:

$$\sigma_{nj} = A_{nj} \times \sigma_{sim} \tag{5.14}$$

These corrected cross sections are listed in table 5.6, where the errors are computed by propagation, and their behavior as a function of the inclusive number of jets per event is shown in figures 5.12 and 5.13. The quantities plotted in 5.11 and 5.12 are closely linked, being σ_{nj} linearly dependent on the number of events with at least n jets, $N_{sel,\,nj}$.

Number of jets	Weighted cross section σ_{nj} [pb]		
	$\alpha_s = 0.1100$	$\alpha_s = 0.1180$	$\alpha_s = 0.1300$
≥ 0	$(4.094 \pm 0.004) \times 10^9$	$(4.475 \pm 0.005) \times 10^9$	$(5.163 \pm 0.005) \times 10^9$
≥ 1	$(1.551 \pm 0.008) \times 10^8$	$(1.750 \pm 0.009) \times 10^8$	$(2.08 \pm 0.01) \times 10^8$
≥ 2	$(1.22 \pm 0.02) \times 10^7$	$(1.37 \pm 0.03) \times 10^7$	$(1.70 \pm 0.03) \times 10^7$
≥ 3	$(4.7 \pm 0.4) \times 10^5$	$(7.1 \pm 0.6) \times 10^5$	$(8.4 \pm 0.7) \times 10^5$
≥ 4	$(2 \pm 1) \times 10^4$	$(4 \pm 1) \times 10^4$	$(5 \pm 2) \times 10^4$

Table 5.6: Values of the cross sections of events with $\geq n$ jets weighted by acceptance A for the three generated samples.

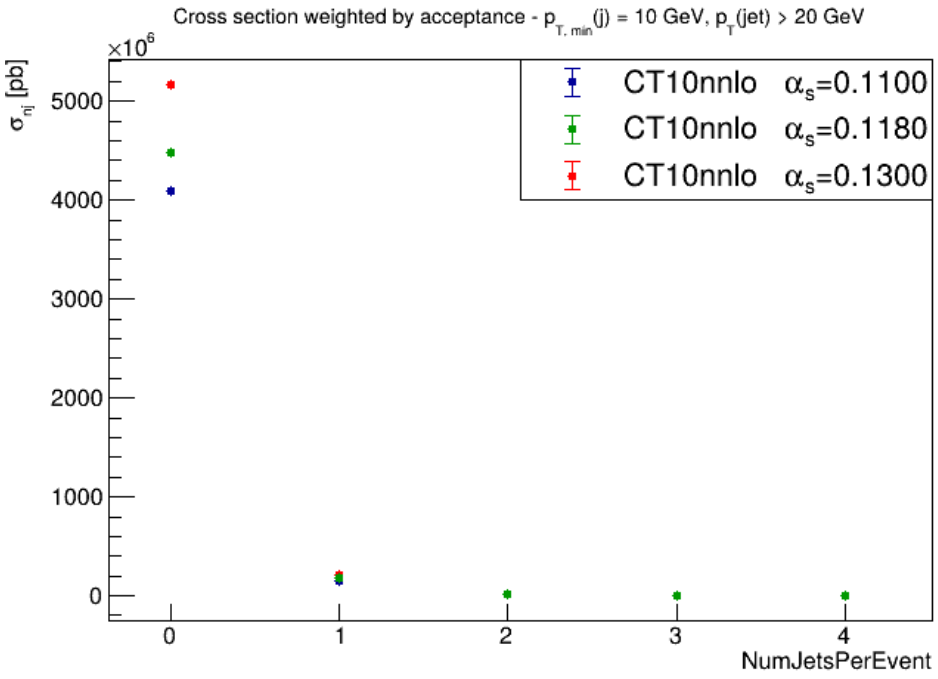

Figure 5.12: Cross sections of events with $\geq n$ jets weighted by acceptance A as a function of the inclusive number of jets per event.

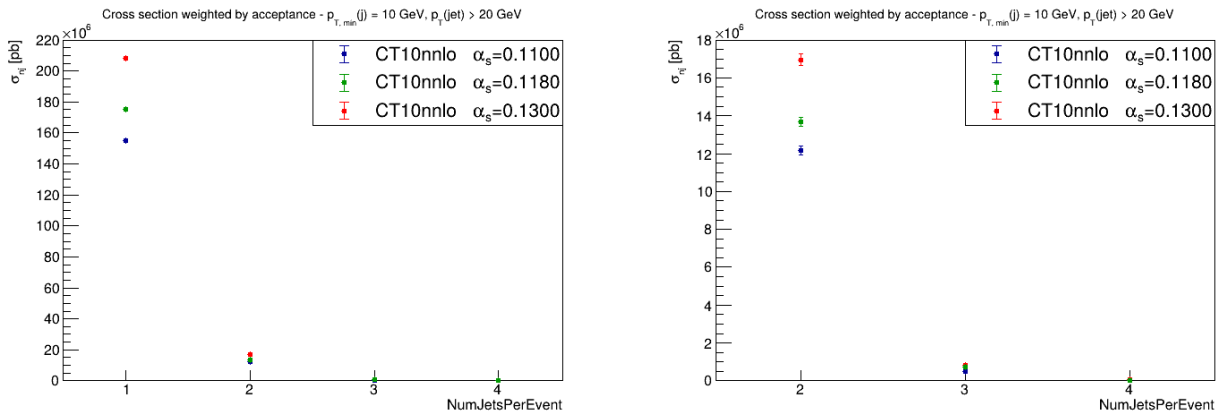

Figure 5.13: Zoom of the plot of cross sections of events with $\geq n$ jets weighted by acceptance A as a function of the inclusive number of jets per event.

The behavior of the weighted cross section σ_{nj} as a function of the inclusive number of jets per event displays sensitivity with respect to the value of α_s. In particular, the focus will be on the weighted cross section for $n \geq 2$ jets: $\sigma_{2j} \equiv \sigma_{jj}^{sim}$.

Therefore, by means of the simulation studies, the chosen variable to focus on to get an estimation of α_s is the dijet cross section, σ_{jj}. The simulated ones with the three values of α_s will be compared to the experimental one to extract α_s and the main contributions to its systematic uncertainty.

Except for the cross sections, the other studied observables do not reveal any sensitivity with respect to the variation of α_s. To further investigate if some dependence arises, the statistics should be increased. In this work, it was not possible to carry out this particular task, due to computational time constraints and the standard time frame available for a thesis project.

Chapter 6

Measurement of α_s from dijet cross section

This chapter outlines the analysis performed on real LHCb data: the value of the strong coupling constant is estimated from the comparison of the measured inclusive dijet cross section and the simulated ones obtained as described in chapter 5. Also, the main systematic uncertainties of α_s are determined. They stem from the integrated luminosity contribution, and from the correction factors required to correct the LHCb Monte Carlo samples for the detector non-uniformities and inefficiencies, that are experimental effects not included in the LHCb simulation.

6.1 Analysis strategy

Two data samples of events collected by LHCb during Run 2 in 2016 at a center of mass energy $\sqrt{s} = 13\,\text{TeV}$ and integrated luminosity $\mathcal{L} = 1.6\,\text{fb}^{-1}$ are analyzed:

1. a set of inclusive events containing two jets, in the following called *dijet* sample;
2. a set of inclusive events containing a Z^0 boson produced in association with a jet and decaying into a $\mu^+\mu^-$ pair, in the following called *Z+jet* sample.

In sections and 6.2 6.3, the two data sets are described in detail. They both contain experimental and simulated data.

The steps implemented to get to an estimate of α_s and its main systematic uncertainties, which are the topics of this chapter, are summarized here.

- From the dijet sample, the experimental inclusive dijet cross section is extracted according to the general formula:

$$N = \sigma \cdot \varepsilon \cdot \mathcal{L} \tag{6.1}$$

  where N is the number of events relative to the process under study, σ is the unknown cross section, ε incorporates all efficiency terms, and $\mathcal{L}$ is the integrated luminosity.

- Both dijet and Z+jet samples are used to compute two calibration factors, K^* and F^*, which are linked to the Jet Energy Scale (JES) and Jet Energy Resolution (JER), respectively. They are applied to the measured cross section, and fix the differences between Monte Carlo and experimental data, on account of detector non-uniformities and inefficiencies, that are not considered in the LHCb simulation. In section 6.4 the procedures developed to extract K^* and F^* are explained.

- From the comparison between the experimental cross section and the fit of the simulated cross sections derived in chapter 5, all as a function of α_s, the value of the strong coupling constant is drawn.

- The main systematic uncertainties involved in the estimation of α_s are propagated from the uncertainties of the dijet cross section measurement, and they come from (the order is arbitrary, not by relevance):
  - Jet Energy Scale;
  - Jet Energy Resolution;
  - integrated luminosity, that in 2016, 2017 and 2018 data-takings of LHC Run 2, is measured with a relative uncertainty of 2%; thus, it introduces a constant uncertainty on the cross section itself, and is independent of the specific analysis under study;
  - trigger efficiencies, due to differences between experimental and simulated data;
  - Initial State Radiation and Final State Radiation, related to gluon emissions that are poorly simulated by `Pythia8`;
  - the PDFs chosen for the simulated cross sections (refer to chapter 5); since PDFs are extracted from experimental data and theoretical calculations, they come with associated uncertainties due to limited data, assumptions in the fitting procedure, and the choice of theoretical models, which translate directly into the uncertainty of the cross section, because the latter is obtained by integrating over the possible parton momentum fractions, weighted by the PDFs.

  The relevant contributions examined in this work are the first three ones. The JES and JER ones are computed in section 6.5. These uncertainties are accounted for as corrections to the selection efficiency with the aim of matching the experimental and MC jet transverse momentum. The corrected efficiency appears in the formula that allows to calculate the cross section from the number of selected events. Section 6.5 is dedicated to explaining the procedures conducted to estimate the corrected efficiency.

In section 6.6 the value of the measured inclusive dijet cross section is reported, with the corrections that have just been discussed. Ultimately, in section 6.7 the preliminary values estimated for the strong coupling constant and its systematic uncertainty are provided.

6.2 Dijet sample

The processes that lead to the production of two jets have already been examined in section 5.1. Retrieving the information disclosed in section 2.1.6 about the LHCb software responsible for managing experimental data and also MC simulations, here the dijet experimental and MC samples are briefly specified.

This experimental dijet sample is based on pp collision events that are collected using specific jet triggers designed to select high-p_T jets (see section 6.2.1). The selected events are then reconstructed, with jets identified and their energies calibrated to account for detector effects. The data undergoes quality checks and selection criteria to ensure only well-reconstructed dijet events are saved. Background suppression techniques are applied to minimize contamination, resulting in a clean dijet sample.

The process of generating the MC dijet sample, instead, begins with simulating pp collisions using `Pythia8` to model the production of high-energy partons, which then hadronize into jets. Any unstable particles produced are decayed using `EvtGen`. The resulting particles are then passed through a detailed simulation of the LHCb detector using `Geant4`, which models how they interact with the detector components. After this, the signals are digitized to simulate the electronic response of the detector. The events are then subjected to the same experimental LHCb trigger conditions, ensuring that only those matching the real data selection criteria are kept. The events are reconstructed with the same software used for the actual data, and finally, corrections are applied to account for differences between the simulated and real detector performance, ensuring the MC sample closely mirrors the observed dijet events.

The total number of generated events considered to build the MC dijet sample, N_{tot}, is the sum of four simulated sets of 2×10^6 events, each with a different value of the transverse momentum exchanged by quarks and gluons at parton level, $\hat{p}_T$, in the pp interaction. The weights, reported in table 6.1 for each $\hat{p}_T$ range, allow to merge the four sets into a single sample, and each weight is proportional to the cross section of the corresponding individual MC set. This, in turn, ensures a good statistics in each of the four intervals.

$\hat{p}_T$ [GeV]	Weight	Eff. events number
$[10, 15]$	1	2×10^6
$[15, 20]$	0.22	4.4×10^5
$[20, 50]$	0.12	2.4×10^5
> 50	0.0035	7×10^3

Table 6.1: Weights applied to the four generated dijet samples depending on the transverse momentum exchanged by quarks and gluons at parton level, $\hat{p}_T$, in the pp interaction. The third column shows the effective number of events considered for each generated set.

Therefore, N_{tot}, written in the following equation, should be properly defined as an effective number of generated events, as the actual number of generated events is simply $2 \times 10^6 \times 4 = 8 \times 10^6$.

$$N_{tot} = (2 \times 10^6 \times 1) + (2 \times 10^6 \times 0.22) + (2 \times 10^6 \times 0.12) + (2 \times 10^6 \times 0.0035) = 2.687 \times 10^6 \quad (6.2)$$

Since the efficiency is a quantity that will be important for this analysis, and it is mathematically defined as the ratio of the number of selected events relative to a process to the number of total events of a sample, every time the MC events enter an efficiency calculation, they are adjusted with the weights reported in table 6.1.

6.2.1 Events selection

The involved trigger lines with the corresponding cuts for the dijet sample are reported in table 6.2, and the fiducial region parameters are listed in table 6.3. The range of acceptance for the two jets, $2.2 < \eta(jet_{1,2}) < 4.2$, does not correspond to the full LHCb coverage, $2 < \eta < 5$, and this is chosen to ensure that the jet cones are entirely inside the fully instrumented LHCb acceptance. Also, the transverse momentum of the jets must be greater than 20 GeV, which is a safe trade-off threshold chosen to collect high quality reconstructed jets, at the same time reducing the rate of fake jets, which becomes negligible with this cut. Two lower limits on the azimuthal separation between the two jets of this sample are used in two different steps of the analysis: $\Delta\phi > 2.8$ is chosen for determining the jets transverse momentum correction factors, K^* and F^* (see section 6.4), while $\Delta\phi > 1$ is set for the extraction of the efficiency adjusted with the contribution of K^* and F^* (see section 6.5). The $\Delta\phi$ threshold was relaxed from 2.8 to 1 not to waste too many events in the efficiency estimation, while assuring to avoid the characteristic anti-k$_t$ artifact at low ϕ-separation values.

The values of the applied pre-scales are displayed in table 6.4. For their definition, see section 2.1.4.1. Notice that the pre-scale factor introduced in the HLT means that 1 dijet event over 1000, taken randomly, is saved, and the others are discarded. Both pre-scale values are significantly large, and this is justified by considering that the employed trigger lines were not initially designed to perform analysis like the one proposed in this work.

The number of events before and after the cuts reported in table 6.3 are listed in the following table 6.5. Notice that the selected number of events after the cuts with $\Delta\phi_{jet_{1,2}} > 2.8$ is not an integer, as it is scaled by the weights appearing in equation 6.2. The selected number of events after the cuts with $\Delta\phi_{jet_{1,2}} > 1$ is weighted as well.

A last consideration must be done: the dijet cross section simulated by the LHCb software is $\sigma_{LHCb} = 8.6 \times 10^8$ pb. It is obtained with $p_T(jet) > 10$ GeV and $\theta < 400$ mrad. To account for the mismatch of the LHCb simulated phase space and the one of this work, the measured dijet cross section will be multiplied by a correction factor, named PS, that will be discussed in section 6.6.

Trigger stage	Trigger lines: OR of
L0	L0HadronDecision_TOS L0MuonDecision_TOS L0PhotonDecision_TOS L0DiMuonDecision_TOS L0ElectronDecision_TOS L0MuonEWDecision_TOS L0JetPhotonDecision
HLT1	Hlt1TrackMVADecision_TOS Hlt1TwoTrackMVADecision_TOS Hlt1TrackMuonDecision_TOS Hlt1TrackMVATightDecision_TOS Hlt1TwoTrackMVATightDecision_TOS Hlt1DiMuonHighMassDecision_TOS Hlt1DiMuonLowMassDecision_TOS Hlt1SingleMuonHighPTDecision_TOS Hlt1DiMuonNoL0Decision_TOS
HLT2	HltQEEJetsDiJet

Table 6.2: Run 2 trigger lines for dijet samples.

Quantity	Cuts
Jets transverse momentum	$20\,\text{GeV} < p_T(jet_{1,2}) < 100\,\text{GeV}$
Jets pseudorapidity	$2.2 < \eta(jet_{1,2}) < 4.2$
Jets azimuthal separation for K^*, F^*	$\Delta\phi_{jet_{1,2}} > 2.8$
Jets azimuthal separation for ε_{corr}	$\Delta\phi_{jet_{1,2}} > 1$

Table 6.3: Fiducial region for the dijet data sample. Different thresholds on the azimuthal separation between the two jets are used in two different steps of the analysis: $\Delta\phi > 2.8$ for determining the jets transverse momentum correction factors, K^* and F^*, and $\Delta\phi > 1$ for the extraction of the efficiency adjusted with the contribution of K^* and F^*.

Correction factor	Symbol	Value
HLT pre-scale	pre_{HLT}	0.001
Stripping pre-scale	pre_{strip}	0.013
GEC efficiency	ε_{GEC}	60%

Table 6.4: Correction factors applied to build the dijet sample.

Sample	Before cuts	After cuts ($\Delta\phi_{jet_{1,2}} > 2.8$)	After cuts ($\Delta\phi_{jet_{1,2}} > 1$)
Data	24970	12661	23281
MC	513657	5004.07	468311

Table 6.5: Number of events contained in dijet sample before and after the fiducial cuts.

6.3 Z+jet sample

The Z+jet data set contains events of the type:

$$Z^0 \left(\rightarrow \mu^+\mu^-\right) + jet \tag{6.3}$$

where a Z^0 boson is produced in association with a jet, and decays into a $\mu^+\mu^-$ pair. Such a production mechanism at hadron colliders occurs according to multiple Feynman diagrams, with the most important ones at LO shown in figure 6.1. Processes like $Z^0 + n$ jets are useful to test pQCD, where the fact that α_s becomes smaller at high energy is exploited to use a perturbative approach. Including a higher number of jets allows to check pQCD at higher orders, as each jet introduces an α_s factor.

For comparison, the LO Feynman diagram for inclusive Z^0 production is shown in figure 6.2.

In this production mechanism, and also in other similar cases, it is important to neglect the tracks associated with the decay particles of the Z^0 boson in the jet reconstruction. This is handled easily by the jet finding algorithm, and the candidate tracks corresponding to the Z^0 boson decay are not included in the list of particles to be clustered into jets.

Z+jet is used as a calibration sample by virtue of the transverse momentum balance between the Z^0 and the jet. The $Z^0 \to \mu^+\mu^-$ decay is indeed reconstructed with high precision (as detailed in the chapter 4 on muon reconstruction), leading to a strong correlation between the transverse momentum of the Z^0 boson and that of the jet. Another important quality of this sample is its very low background (whose main sources are listed below), owing to the precise identification of the Z^0 boson. Additionally, the trigger lines used for this sample, reported in the section 6.3.1 below, are designed to select the Z^0 boson, ensuring that the jet selection is free from any trigger bias.

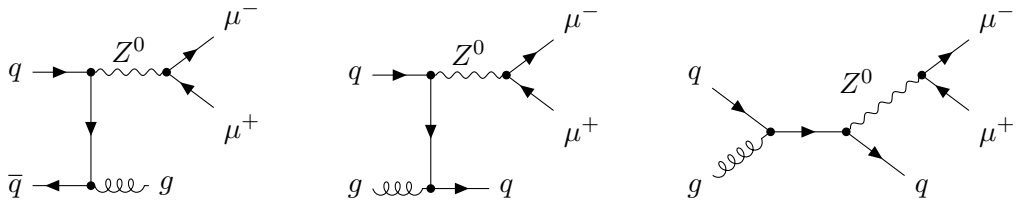

Figure 6.1: The LO Feynman diagrams for Z+jet production at a hadron collider, with the boson subsequently decaying to a $\mu^+\mu^-$ pair. All particles in each diagram can be swapped for their corresponding anti-particles.

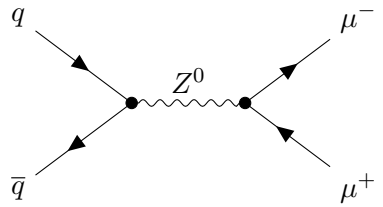

Figure 6.2: The LO Feynman diagram for Z^0 boson production, with the subsequent decay into a $\mu^+\mu^-$ pair. All particles in each diagram can be swapped for their corresponding anti-particles.

To build the experimental Z+jet sample, the process begins with the collection of pp collision events, recorded using the LHCb detector. The events are selected by triggers designed to identify high-p_T muon pairs (signature of $Z^0 \to \mu^+\mu^-$ decays) and high-p_T jets (see section 6.3.1). The collected events are reconstructed to identify the Z^0 boson decays into muons and the associated jets. This involves tracking muons to measure their momentum and reconstructing the invariant mass of the muon pair to identify the $Z^0 \to \mu^+\mu^-$ decays. Jets are formed from hadrons in the event and their properties are reconstructed. The events undergo quality checks to ensure accurate reconstruction. Selection criteria are applied to filter events based on the muons invariant mass (near the Z^0 boson mass) and the jets kinematic properties, such as transverse momentum and pseudorapidity. Techniques are used to suppress background events that could mimic the Z+jet signature. These expected backgrounds have different origins, indexed here.

- $Z^0 \to \tau^+\tau^-$: in this channel, the Z^0 boson decays to taus, and each tau decays to a muon. Around 7 ± 3 $\tau^+\tau^-$ events can be attributed to this background for the 20 GeV cut on the jet transverse momentum [69].
- Muon mis-identification: this combinatorial background arises from muons in an event that are randomly produced with the correct invariant mass. These muons can come from semileptonic B decays, $W^\pm$ boson decays, or misidentified hadrons, which may decay in flight or pass through the calorimeters and reach the muon stations, incorrectly passing the `IsMuon` flag. Jets produced from unrelated QCD processes or multiple parton interactions may also contribute. The probability of a hadron being misidentified as a muon is charge-independent. Since LHCb has a high

muon identification efficiency, this background is low, with 5 ± 2 events foreseen for a minimum jet transverse momentum of 20 GeV [69].

- $t\bar{t}$ production: each top quark decays to a b quark and a $W^{\pm}$ boson. The b quark is associated to a jet, while if the $W^{\pm}$ boson decays to a muon and a neutrino, the whole final state could mimic the Z+jet final state. 5 ± 2 events of this type are expected for the 20 GeV jet transverse momentum threshold [69].
- Electroweak di-boson production (Z^0Z^0, $W^{\pm}W^{\mp}$, $Z^0W^{\pm}$): the process belonging to this class with the largest cross section is $W^{\pm}$ boson pair production, that is around 20% of the $t\bar{t}$ production cross section [92]. The requirement of a jet production associated with the $W^{\pm}$ boson pair further reduces this cross section.
- Heavy flavor Z^0 boson semileptonic decays: the decay products decay to muons, providing a very small contribution, as the muons generated in this way almost never have enough energy to pass the invariant mass and transverse momentum requirements.
- Fake jets: events which contain a correctly reconstructed dimuon decay of a Z^0 boson, but where a jet is incorrectly reconstructed.

Overall, the relevant sources of background taken altogether result in 17 ± 4 events for the 20 GeV cut on the jet transverse momentum [69]. Such value is negligible for the purposes of the present analysis. Moreover, previous studies [92, 93] have shown that the background level does not vary considering different kinematic distributions of the Z^0 boson. To check if the background contributions are actually negligible, in figure 6.3 the invariant mass of the $Z^0 \to \mu^+\mu^-$ process is plotted. It displays a clear and narrow peak centered at the nominal value of the Z^0 boson mass, $m_{Z^0} = 91.1880$ GeV, confirming that the backgrounds for its identification can indeed be neglected.

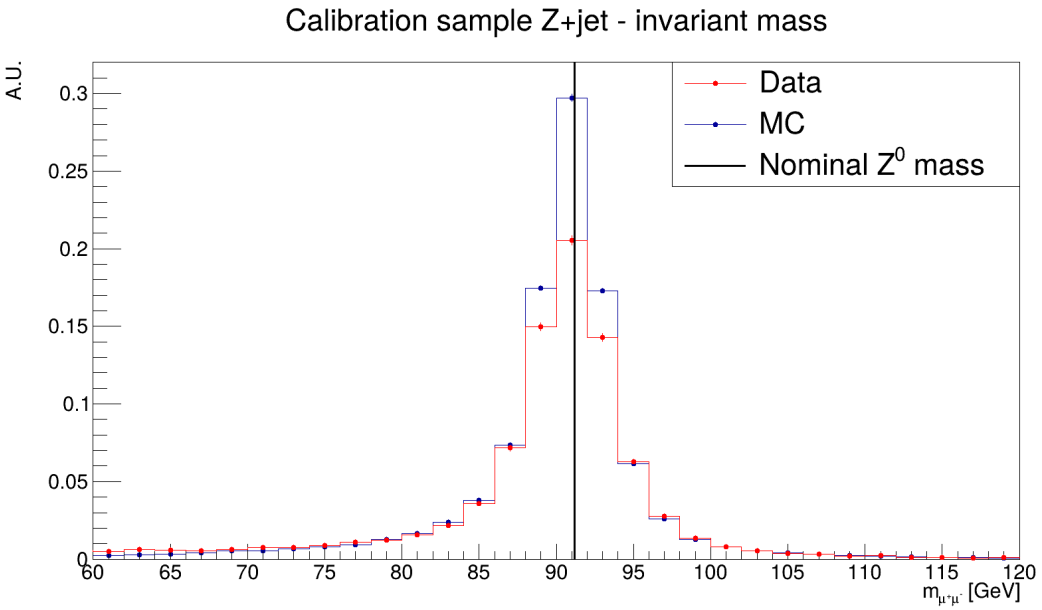

Figure 6.3: Invariant mass of the process $Z^0 \to \mu^+\mu^-$, showing a peak at the nominal Z^0 boson mass, $m_{Z^0} = 91.1880$ GeV. This confirms the negligibility of the backgrounds.

To generate the MC Z+jet sample, the first step is the simulation of pp collisions using `Pythia8` to produce a Z^0 boson alongside a jet. The Z^0 boson decays into a pair of muons, and the resulting particles are passed through a detailed detector simulation with `Geant4`. The simulated events undergo digitization and are filtered through the LHCb trigger system to select relevant events, with the same trigger lines employed for the experimental sample. The events are then reconstructed using the same software as the real data, and corrections are applied to ensure the MC sample accurately reflects the conditions of the 2016 data.

6.3.1 Events selection

The involved trigger lines with the corresponding cuts for the Z+jet sample are reported in table 6.6, and the fiducial region parameters are listed in table 6.7. The range of acceptance for the jet, $2.2 < \eta(jet) < 4.2$, does not correspond to the full LHCb coverage, $2 < \eta < 5$, and this is chosen to ensure that the jet cones are entirely inside the fully instrumented LHCb acceptance. Also, the transverse momentum of the jet must be greater than 20 GeV, which is a threshold value set in LHCb that enables to preserve the majority of signal events.

Trigger stage	Trigger line
L0	`L0MuonEWDecision_TOS`
HLT1	`Hlt1SingleMuonHighPTDecision_TOS`
HLT2	`Hlt2Global_TOS`

Table 6.6: Run 2 trigger lines for muons for the Z+jet data sample.

Quantity	Cuts
Muon transverse momentum	$p_T(\mu^{\pm}) > 20\,\text{GeV}$
Muon pseudorapidity	$2 < \eta(\mu^{\pm}) < 4.5$
Jet transverse momentum	$20\,\text{GeV} < p_T(jet) < 100\,\text{GeV}$
Jet pseudorapidity	$2.2 < \eta(jet) < 4.2$
Z-jet azimuthal separation	$\Delta\phi_{Z+jet} > 2.8$

Table 6.7: Fiducial region for the Z+jet data sample.

The number of events before and after the cuts reported in table 6.7 are listed in the following table 6.8.

Sample	Before cuts	After cuts
Data	190849	13262
MC	374981	34808

Table 6.8: Number of events contained in dijet sample before and after the fiducial cuts.

6.4 Jet Energy Scale and Jet Energy Resolution

Referring to the considerations elaborated in section 3.1.6, two factors are computed in the context of this analysis to compensate differences between Monte Carlo and experimental data: K^*, connected to the *Jet Energy Scale* (JES), and F^*, connected to the *Jet Energy Resolution* (JER). They do not have a fixed theoretical analytical expression, but they must be obtained by comparing the experimental recombination-based jet transverse momentum with the Monte Carlo one. A MC jet is the result of the clustering algorithm performed on the simulated partons of a pp collision.

To measure them, two variables are extracted from the aforementioned calibration samples. The first one, R_Z, is employed for the Z+jet calibration sample, and is expressed as:

$$R_Z = \frac{p_T(jet)}{p_T(Z^0)} \tag{6.4}$$

while the second one is named R_{12}, and is used for the dijet calibration sample:

$$R_{12} = \frac{p_T(jet_1) - p_T(jet_2)}{p_T(jet_1) + p_T(jet_2)} \tag{6.5}$$

R_Z and R_{12} have been studied because they embed the balancing of the Z^0-jet and jet_1-jet_2 in the transverse plane. In order to not introducing any bias due to the p_T-ordering of the jets, to derive

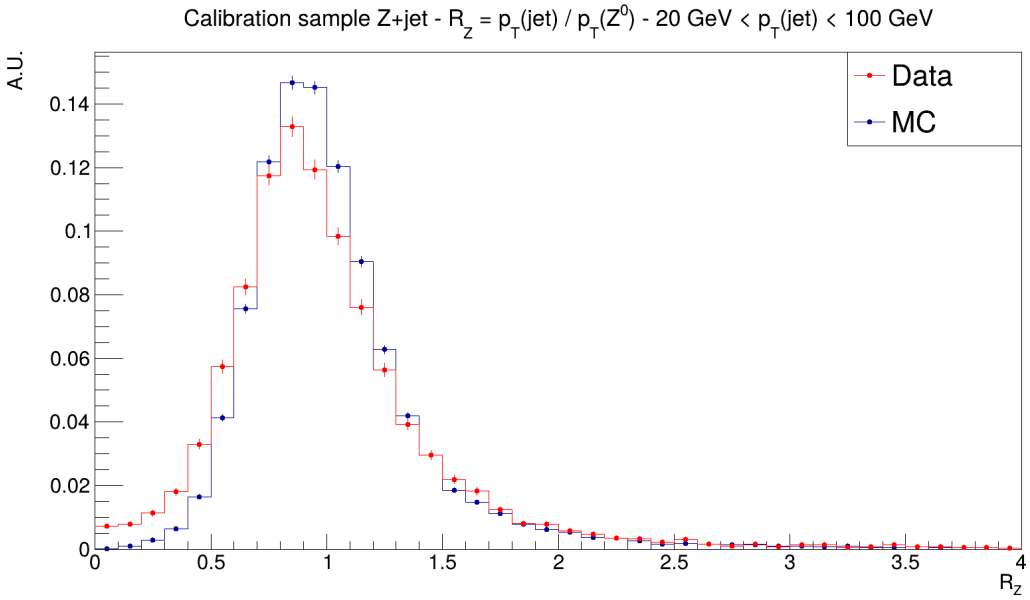

Figure 6.4: R_Z experimental and MC distributions in bins of $p_T(jet)$ in the range 20 GeV $< p_T(jet) <$ 100 GeV.

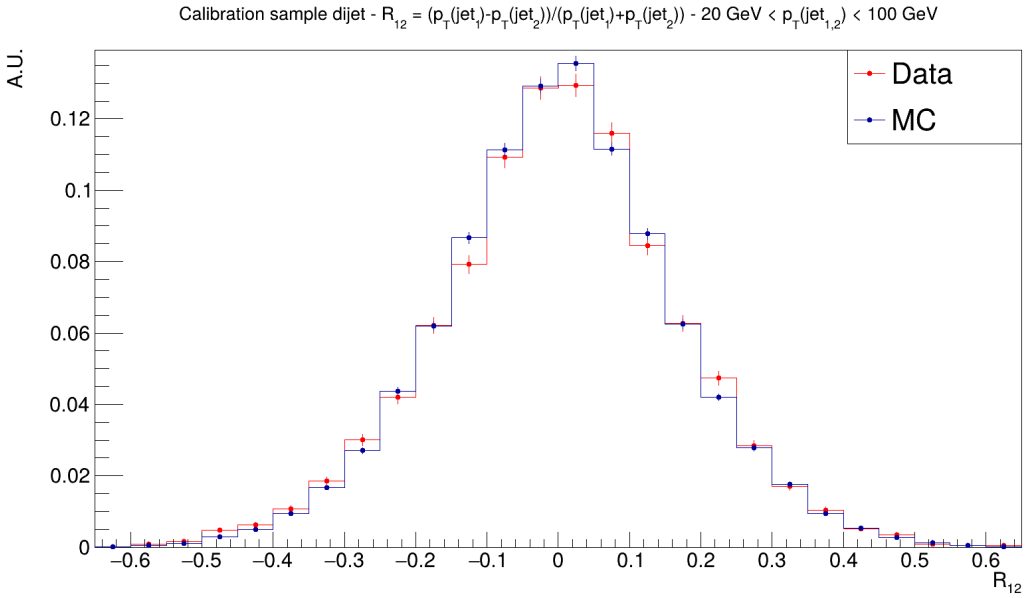

Figure 6.5: R_{12} experimental and MC distributions in bins of $p_T(jet_{1,2})$ in the range 20 GeV $< p_T(jet_{1,2}) <$ 100 GeV.

R_{12} the two leading jets in an event are assigned the label jet_1 or jet_2 randomly. In figures 6.4 and 6.5, the plots of the normalized R_Z and R_{12} distributions in the whole studied transverse momentum range are shown.

K^* is quantified in bins of the jet transverse momentum as a multiplicative factor for the MC transverse momentum. Therefore, it can only be derived from the Z+jet sample, as in R_{12} the factor - applied simultaneously to both jets - would cancel out the one appearing in the denominator, making it impossible to estimate K^* from the dijet calibration sample. Instead, F^* is accounted for into a gaussian smearing factor for the MC or the experimental jet transverse momentum, depending on which R_Z/R_{12} distribution is the narrowest. F^* has been determined with both Z+jet and dijet samples.

In the following sections 6.4.1 and 6.4.2, a detailed description of the procedures implemented to obtain K^* and F^* from the two analyzed calibration samples is provided.

6.4.1 JES- and JER-related factors from Z+jet sample

As anticipated, from the Z+jet sample both K^* and F^* are measured. In this sample a cut on the azimuthal separation of the Z^0 and jet is set: $\Delta\phi_{Z+jet} > 2.8$ (see figure 6.6). Thus, only events with back-to-back produced Z^0 and jet are kept, and so the R_Z distribution is ideally peaked at 1. Thereby, a $2 \to 2$ event topology is ensured.

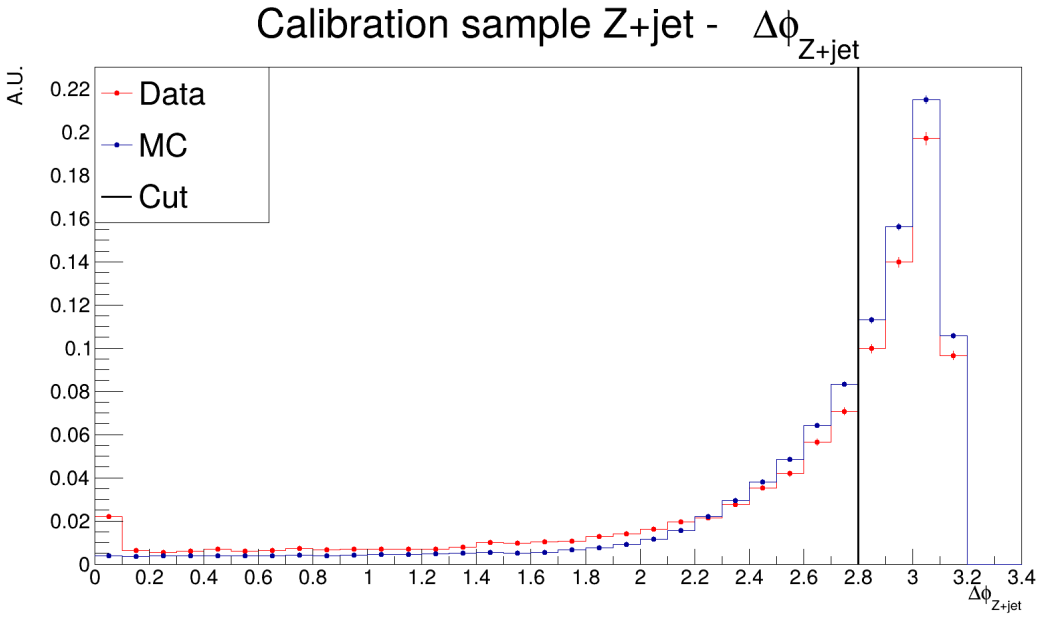

Figure 6.6: Azimuthal separation $\Delta\phi$ between the Z^0 boson and the jet: the cut at 2.8 is applied to ensure $2 \to 2$ event topology.

K^* is evaluated in bins of $p_T^{MC}(jet)$ between 20 GeV and 100 GeV with 5 GeV width as a multiplicative factor applied to the MC jet transverse momentum such that the mean value of the R_Z MC distribution matches the one extracted from data:

$$p_T^{MC}(jet) \to K^* \times p_T^{MC}(jet) \quad : \quad \langle R_Z^{MC}(K^*) \rangle = \langle R_Z^{data} \rangle \tag{6.6}$$

The method to obtain K^* is the following:

1. 10 different values of the multiplicative factor, K, are inserted in equation 6.6, computing the corresponding $\langle R_Z^{MC}(K) \rangle$;
2. the points in the K-$\langle R_Z^{MC}(K) \rangle$ plane are interpolated with a line of the type $\langle R_Z^{MC}(K) \rangle = mK + q$;
3. the intersection between the fit line and the horizontal line representing $\langle R_Z^{data} \rangle$ gives the K^* value that was searched for:
$$K^* = \frac{\langle R_Z^{data} \rangle - q}{m} \tag{6.7}$$
4. from the intersection between the lines representing the error bar of $\langle R_Z^{data} \rangle$, placed at $\langle R_Z^{data} \rangle + e_{\langle R_Z^{data} \rangle}$ and $\langle R_Z^{data} \rangle - e_{\langle R_Z^{data} \rangle}$, and the fit line, also an estimate of the interval associated to the K^* estimate is extracted:
$$\Delta K^* = \frac{2e_{\langle R_Z^{data} \rangle}}{m} \tag{6.8}$$

These steps are repeated for each $p_T^{MC}(jet)$ bin. In plots in figures 6.7 and 6.8 the procedure for extracting K^* in 2 of the 16 bins is shown. The behavior of $\langle R_Z \rangle$ as a function of K is linear and very regular in all $p_T^{MC}(jet)$ bins: the mean values of the R_Z distribution increase along with the K's.

In figures 6.9 and 6.10 the resulting K^*, ΔK^*, and $\Delta K^*/K^*$ as functions of $p_T^{MC}(jet)$ are plotted. In table 6.9, the values of K^* for each $p_T^{MC}(jet)$ bin are reported. The errors are calculated by propagation. The K^* values increase as the $p_T^{MC}(jet)$ of the considered bin gets larger, from a minimum at around 0.91 to a maximum of 1.26. In the first 7 bins (i.e. for $20\,\text{GeV} < p_T^{MC}(jet) < 55\,\text{GeV}$) $K^* < 1$, implying that $\langle R_Z^{MC} \rangle$ is higher than $\langle R_Z^{data} \rangle$, and the effect of K^* is shifting the mean value of the MC R_Z distribution towards a smaller value. Vice versa, in the remaining bins (i.e. for $60\,\text{GeV} < p_T^{MC}(jet) < 100\,\text{GeV}$) $K^* > 1$, so $\langle R_Z^{MC} \rangle$ is lower than $\langle R_Z^{data} \rangle$, and K^* moves the mean value of the MC R_Z distribution towards a larger value. As far as ΔK^* and $\Delta K^*/K^*$ are concerned, they both increase with $p_T^{MC}(jet)$, and their error bands increase as well. This happens because the number of events per bin decreases as the MC jet transverse momentum increases.

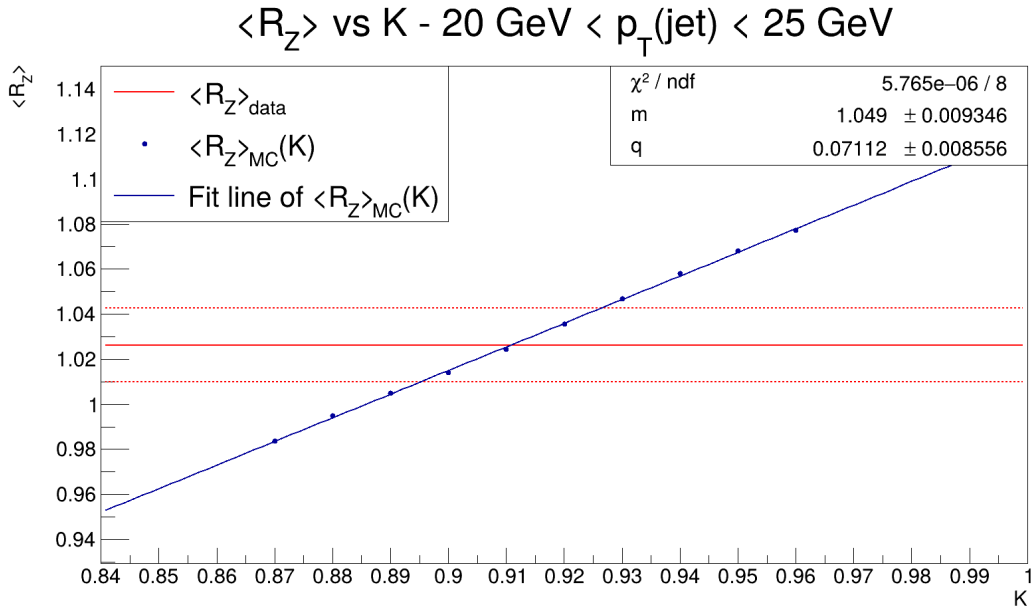

Figure 6.7: Extraction of K^* from the comparison between experimental and MC $\langle R_Z \rangle$ in bin $20\,\text{GeV} < p_T(jet) < 25\,\text{GeV}$.

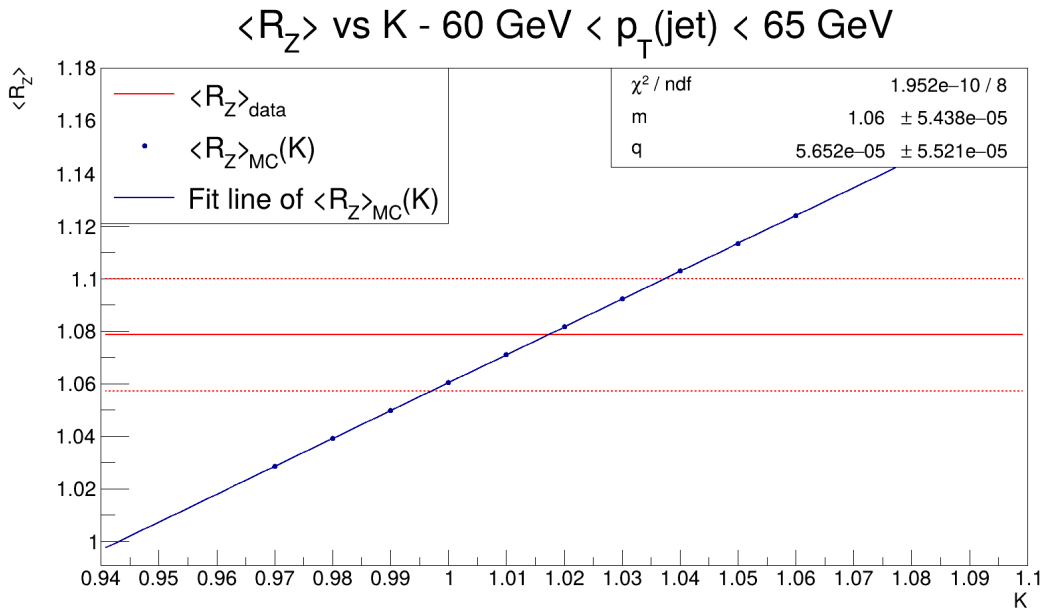

Figure 6.8: Extraction of K^* from the comparison between experimental and MC $\langle R_Z \rangle$ in bin $60\,\text{GeV} < p_T(jet) < 65\,\text{GeV}$.

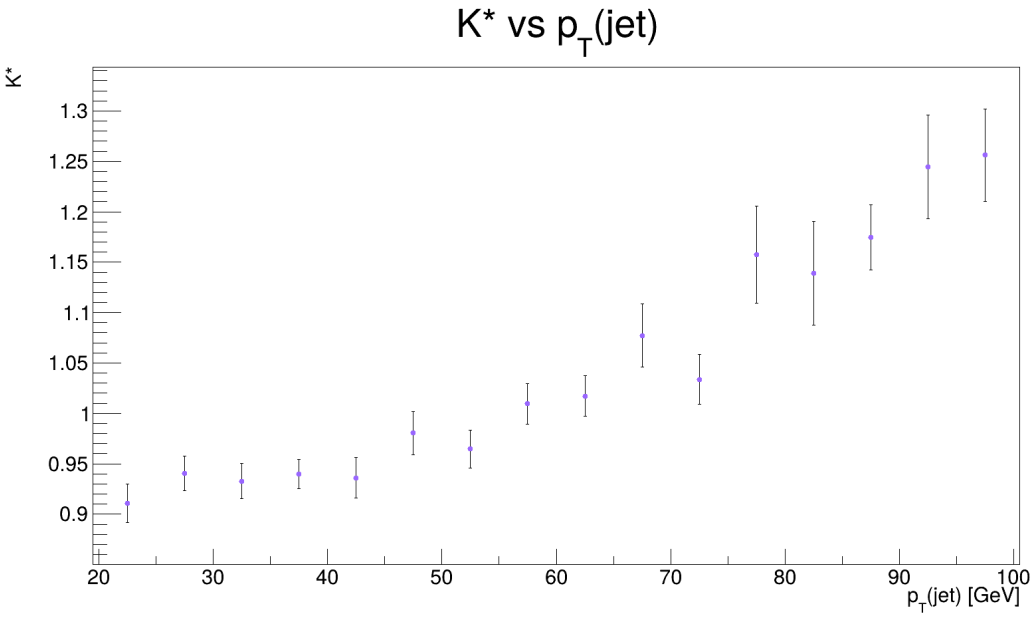

Figure 6.9: K^* corrections as functions of $p_T^{MC}(jet)$.

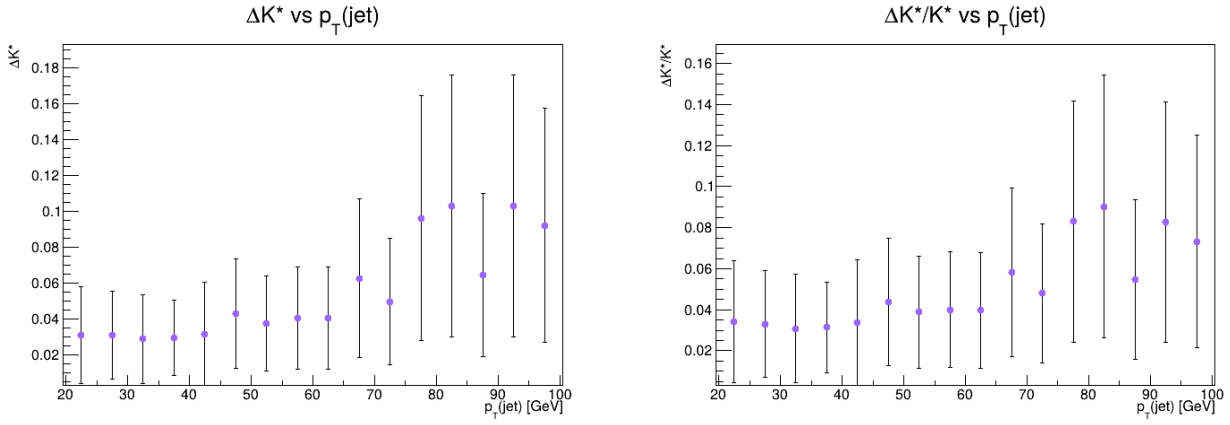

Figure 6.10: ΔK^* and $\Delta K^*/K^*$ as functions of $p_T^{MC}(jet)$.

$p_T(jet)$ [GeV] bin	K^*
$[20, 25]$	0.91 ± 0.02
$[25, 30]$	0.94 ± 0.02
$[30, 35]$	0.93 ± 0.02
$[35, 40]$	0.94 ± 0.01
$[40, 45]$	0.94 ± 0.02
$[45, 50]$	0.98 ± 0.02
$[50, 55]$	0.96 ± 0.02
$[55, 60]$	1.01 ± 0.02
$[60, 65]$	1.02 ± 0.02
$[65, 70]$	1.08 ± 0.03
$[70, 75]$	1.03 ± 0.02
$[75, 80]$	1.16 ± 0.05
$[80, 85]$	1.14 ± 0.05
$[85, 90]$	1.17 ± 0.03
$[90, 95]$	1.24 ± 0.05
$[95, 100]$	1.26 ± 0.05

Table 6.9: K^* values associated to JES correction in the studied bins of jet transverse momentum.

As far as the F^* determination is concerned, it is measured in bins of $p_T^{MC}(jet)$ between 20 GeV and 100 GeV with 10 GeV width as a smearing factor, F^*, for the MC jet transverse momentum such that the width of the R_Z MC distribution matches the one extracted from data:

$$p_T^{MC}(jet) \to \text{Gauss}\Big(p_T^{MC}(jet), F^* \times p_T^{MC}(jet)\Big) \quad : \quad \sigma_{R_Z}^{MC}(F^*) = \sigma_{R_Z}^{data} \tag{6.9}$$

With this operation, for each $p_T^{MC}(jet)$ value a random fluctuation in introduced, which is drawn from a gaussian distribution centered on the original $p_T^{MC}(jet)$ value and with a standard deviation proportional to $F^* \times p_T^{MC}(jet)$. The smeared $p_T^{MC}(jet)$ distribution is broader than the non-smeared one, and any sharp features or peaks in the original $p_T^{MC}(jet)$ distribution are less pronounced, resulting in a smoother smeared distribution. Also, individual $p_T^{MC}(jet)$ values become less deterministic, reflecting the oscillations included by the smearing. This variability mimics the effects of noise and of experimental resolution and uncertainties.

The steps performed to get F^* are enumerated here:

1. 100 different values of the smearing factor, F, are inserted in equation 6.9, computing the corresponding $\sigma_{R_Z}^{MC}(F)$;
2. the points in the F-$\sigma_{R_Z}^{MC}(F)$ plane are interpolated with a line of the type $\sigma_{R_Z}^{MC}(F) = mF + q$;
3. the intersection between the fit line and the horizontal line representing $\sigma_{R_Z}^{data}$ gives the F^* value that was searched for:
$$F^* = \frac{\sigma_{R_Z}^{data} - q}{m} \tag{6.10}$$
4. from the intersection between the lines representing the error bar of $\sigma_{R_Z}^{data}$, placed at $\sigma_{R_Z}^{data} + e_{\sigma_{R_Z}^{data}}$ and $\sigma_{R_Z}^{data} - e_{\sigma_{R_Z}^{data}}$, and the fit line, also an estimate of the interval associated to the F^* estimate is extracted:
$$\Delta F^* = \frac{2e_{\sigma_{R_Z}^{data}}}{m} \tag{6.11}$$

These steps are repeated for each $p_T^{MC}(jet)$ bin. In plots in figures 6.11 and 6.12 the procedure for extracting F^* in 2 of the 8 bins is shown. The behavior of $\sigma_{R_Z}^{MC}$ as a function of F is linear in all $p_T^{MC}(jet)$ bins: the width of the R_Z distribution increase along with the F's. There 100 $\sigma_{R_Z}^{MC}(F)$ points exhibit a moderate fluctuation with respect to a regular linear trend, and this can be explained by

the fact that the gaussian smearing is a randomly-driven operation, so the random number generator introduces an additional variability.

In figures 6.13 and 6.14 the resulting F^*, ΔF^*, and $\Delta F^*/F^*$ as functions of $p_T^{MC}(jet)$ are plotted. In table 6.10, the values of F^* for each $p_T^{MC}(jet)$ bin are reported. The errors are calculated by propagation. The F^* values increase as the $p_T^{MC}(jet)$ of the considered bin gets larger, from a minimum at around 0.11 to a maximum of 0.57. In all bins F^* is well below 1, meaning that $\sigma_{R_Z}^{MC}$ is larger than $\sigma_{R_Z}^{data}$, so F^* shrinks the width of the MC R_Z distribution. Both ΔF^* and $\Delta F^*/F^*$ points are close to each other, with the values and the error bands of $\Delta F^*/F^*$ decreasing due to the increasing F^* value in higher $p_T^{MC}(jet)$ bins. The error bands of F^* and ΔF^* increase, as the number of events per bin decreases as the MC jet transverse momentum increases.

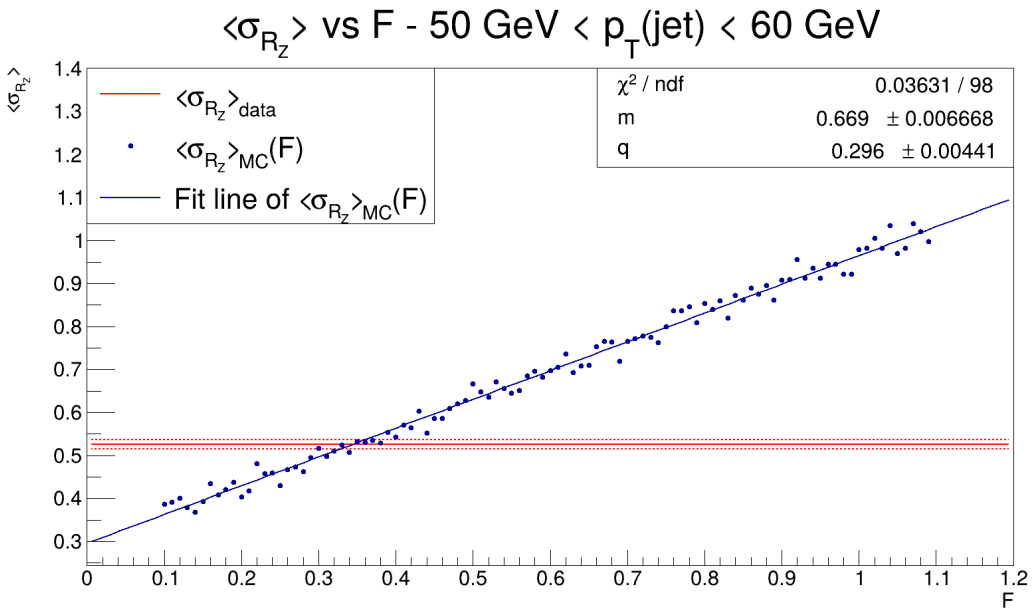

Figure 6.11: Extraction of F^* from the comparison between experimental and MC σ_{R_Z} in bin 50 GeV $< p_T(jet) <$ 60 GeV.

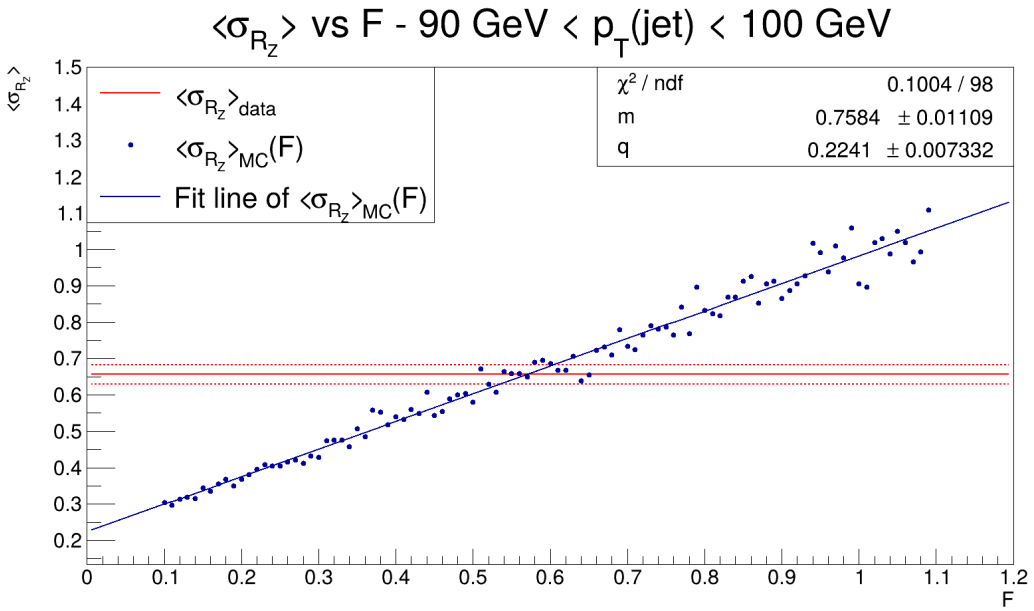

Figure 6.12: Extraction of F^* from the comparison between experimental and MC σ_{R_Z} in bin 90 GeV $< p_T(jet) <$ 100 GeV.

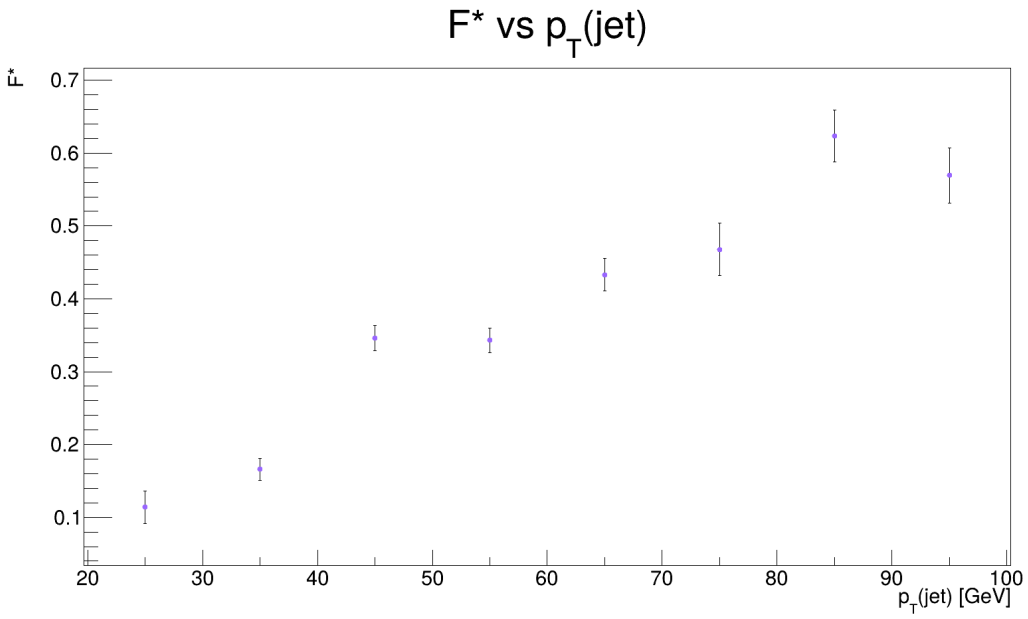

Figure 6.13: F^* corrections as functions of $p_T^{MC}(jet)$.

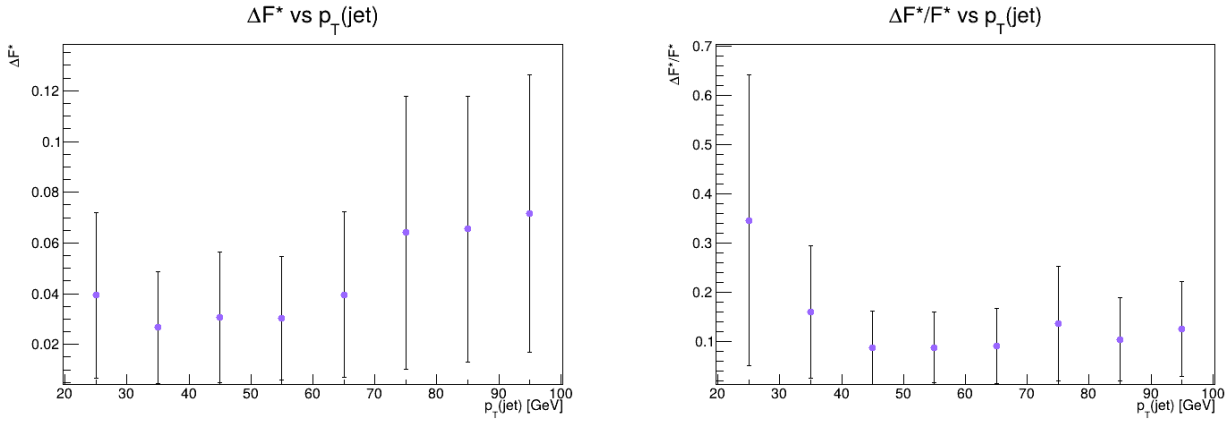

Figure 6.14: ΔF^* and $\Delta F^*/F^*$ as functions of $p_T^{MC}(jet)$.

$p_T(jet)$ [GeV] bin	F^*
$[20, 30]$	0.11 ± 0.02
$[30, 40]$	0.17 ± 0.02
$[40, 50]$	0.35 ± 0.02
$[50, 60]$	0.34 ± 0.02
$[60, 70]$	0.43 ± 0.02
$[70, 80]$	0.47 ± 0.04
$[80, 90]$	0.62 ± 0.04
$[90, 100]$	0.57 ± 0.04

Table 6.10: F^* values associated to JER correction in the studied bins of jet transverse momentum.

6.4.2 JER-related factor from dijet sample

As already examined in section 6.4, with the dijet calibration sample only F^* can be measured, as the K^* factor cancels out in numerator and denominator. Also for this sample, a cut on the azimuthal separation of the Z^0 and jet is set: $\Delta\phi_{jet_{1,2}} > 2.8$ (see figure 6.15). Thus, only events with back-to-back produced leading jet pairs are kept, and so the R_{12} distribution is ideally peaked at 0. Thereby, a $2 \to 2$ event topology is ensured.

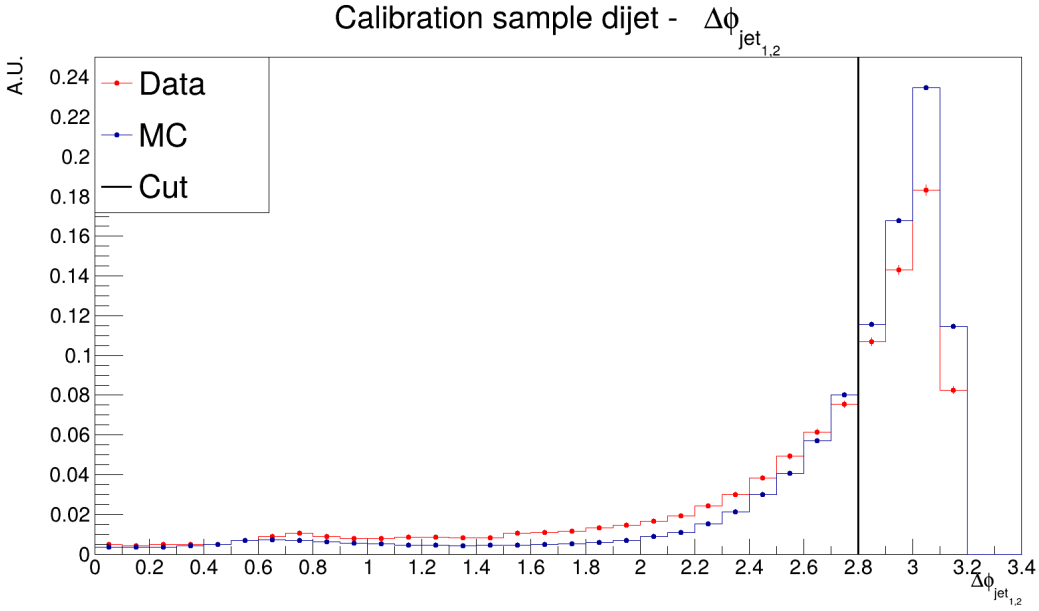

Figure 6.15: Azimuthal separation $\Delta\phi$ between the jet_1 and jet_2: the cut at 2.8 is applied to ensure $2 \to 2$ event topology.

The JER determination is similar to the one adopted for the Z+jet sample. It is measured in bins of $p_T^{MC}(jet_{1,2})$ between 20 GeV and 100 GeV with 10 GeV width, but the important difference with respect to the Z+jet sample is that in this case every p_T^{MC} bin shows a distribution broader than the experimental one, p_T^{data}. Therefore, JER is quantified as a smearing factor, F^*, for the data jet transverse momentum such that the width of the R_{12} data distribution matches the MC one:

$$p_T^{data}(jet_{1,2}) \to \text{Gauss}\Big(p_T^{data}(jet_{1,2}), F^* \times p_T^{data}(jet_{1,2})\Big) \quad : \quad \sigma_{R_{12}}^{data}(F^*) = \sigma_{R_{12}}^{MC} \tag{6.12}$$

Accordingly, the stages which lead to the F^* determination are:

1. 100 different values of the smearing factor, F, are inserted in equation 6.12, computing the corresponding $\sigma_{R_{12}}^{data}(F)$;
2. the points in the F-$\sigma_{R_{12}}^{data}(F)$ plane are interpolated with a line of the type $\sigma_{R_{12}}^{data}(F) = mF + q$;
3. the intersection between the fit line and the horizontal line representing $\sigma_{R_{12}}^{MC}$ gives the F^* value that was searched for:
$$F^* = \frac{\sigma_{R_{12}}^{MC} - q}{m} \tag{6.13}$$
4. from the intersection between the lines representing the error bar of $\sigma_{R_{12}}^{MC}$, placed at $\sigma_{R_{12}}^{MC} + e_{\sigma_{R_{12}}^{MC}}$ and $\sigma_{R_{12}}^{MC} - e_{\sigma_{R_{12}}^{MC}}$, and the fit line, also an estimate of the interval associated to the F^* estimate is extracted:
$$\Delta F^* = \frac{2e_{\sigma_{R_{12}}^{MC}}}{m} \tag{6.14}$$

These steps are repeated for each $p_T^{data}(jet_{1,2})$ bin. In plots in figures 6.16 and 6.17 the procedure for extracting F^* in 2 of the 8 bins is shown. The behavior of $\sigma_{R_{12}}^{data}$ as a function of F displays the

same behavior in all $p_T^{data}(jet_{1,2})$ bins: the width of the R_{12} distribution increases along with the F's, and the 100 $\sigma_{R_{12}}^{data}(F)$ points exhibit an increasing dispersion as F increases. The linear trend is lost, so a linear fit is performed locally around F^*. This significant fluctuation can be explained by the randomness of the gaussian smearing, and also by the fact that in dijet sample data and MC R_{12} distributions are already well compatible, as proven by figure 6.5.

In figures 6.18 and 6.19 the resulting F^*, ΔF^*, and $\Delta F^*/F^*$ as functions of $p_T^{data}(jet_{1,2})$ are plotted. In table 6.11, the values of F^* for each $p_T^{data}(jet_{1,2})$ bin are reported. The errors are calculated by propagation. The F^* values oscillate around an imaginary horizontal line placed roughly at $F^* = 0.016$, from a minimum at around 0.005 to a maximum of 0.028. In all bins F^* is far below 1, further proving that data and MC R_{12} distributions are very similar. $F^* \ll 1$ means that $\sigma_{R_{12}}^{data}$ is slightly larger than $\sigma_{R_{12}}^{MC}$, so F^* narrows the width of R_{12}^{data}. Both ΔF^* and $\Delta F^*/F^*$ points are very close to each other.

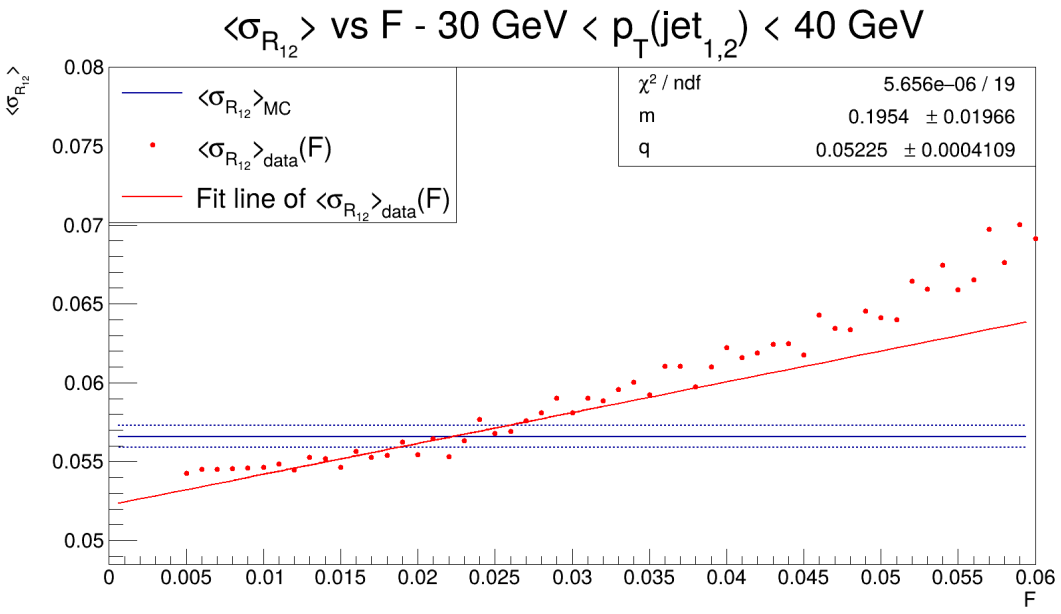

Figure 6.16: Extraction of F^* from the comparison between experimental and MC $\sigma_{R_{12}}$ in bin 30 GeV $< p_T(jet_{1,2}) <$ 40 GeV.

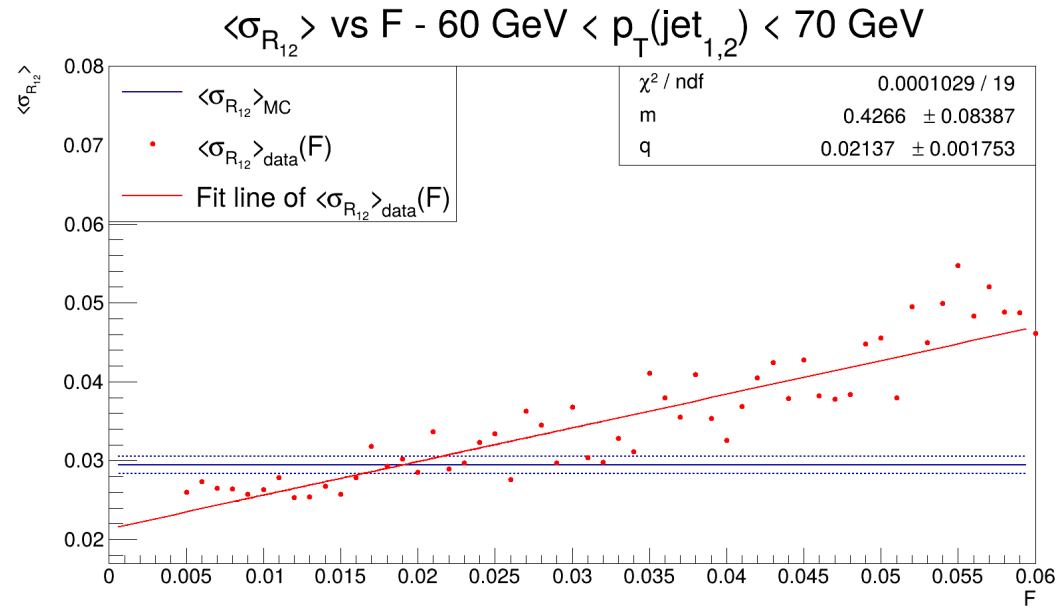

Figure 6.17: Extraction of F^* from the comparison between experimental and MC $\sigma_{R_{12}}$ in bin 60 GeV $< p_T(jet_{1,2}) <$ 70 GeV.

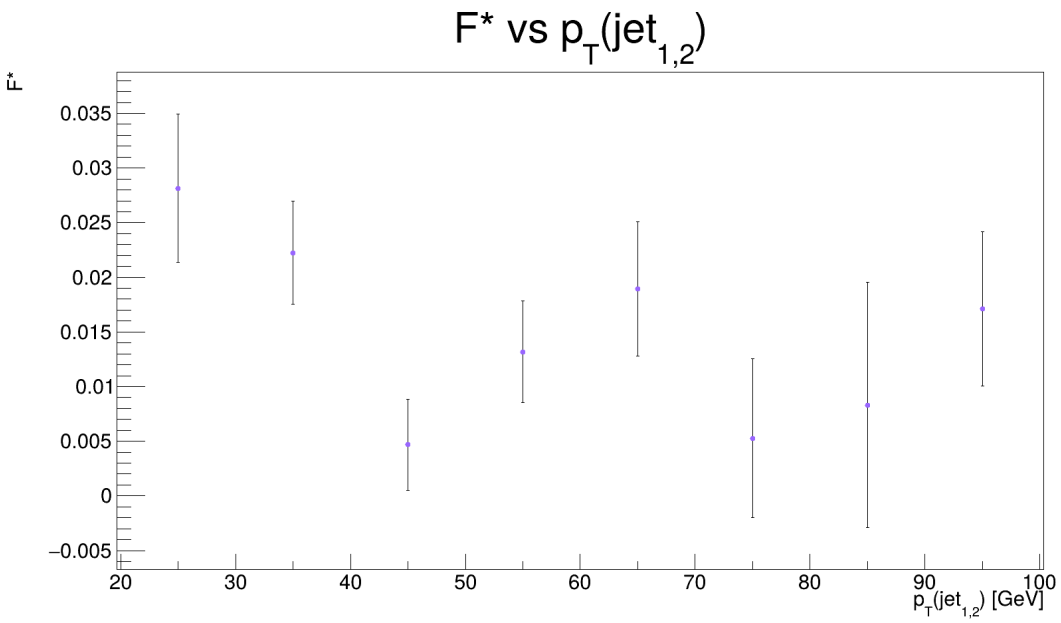

Figure 6.18: F^* corrections as functions of $p_T^{MC}(jet_{1,2})$.

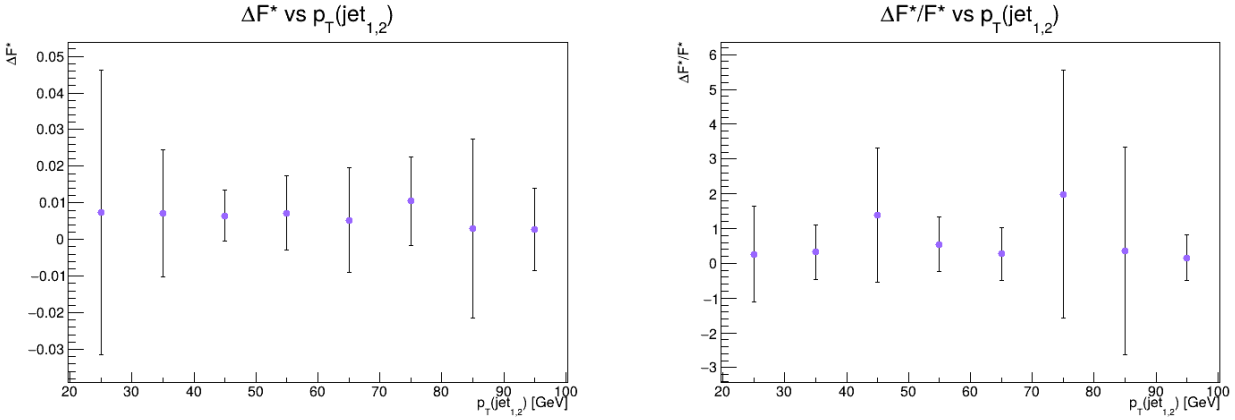

Figure 6.19: ΔF^* and $\Delta F^*/F^*$ as functions of $p_T^{MC}(jet_{1,2})$.

$p_T(jet_{1,2})$ [GeV] bin	F^*
$[20, 30]$	0.028 ± 0.007
$[30, 40]$	0.022 ± 0.005
$[40, 50]$	0.005 ± 0.004
$[50, 60]$	0.013 ± 0.005
$[60, 70]$	0.019 ± 0.06
$[70, 80]$	0.005 ± 0.007
$[80, 90]$	0.01 ± 0.01
$[90, 100]$	0.017 ± 0.007

Table 6.11: F^* values associated to JER correction in the studied bins of jet transverse momentum.

6.5 Efficiency determination

Once the K^* and F^* factors associated with JES and JER are computed as specified in the previous sections, they are used to correct the transverse momentum Monte Carlo distributions of the two jets in the dijet sample. The F^* values derived from the dijet sample have been chosen over the Z+jet sample ones because they are more precise. They are derived through the gaussian smearing of $\sigma_{R_{12}}$ distribution extracted from data, and then applied, following a typical and conservative approach, to the MC transverse distribution. In figure 6.20 the comparison between the uncorrected and corrected distributions is plotted.

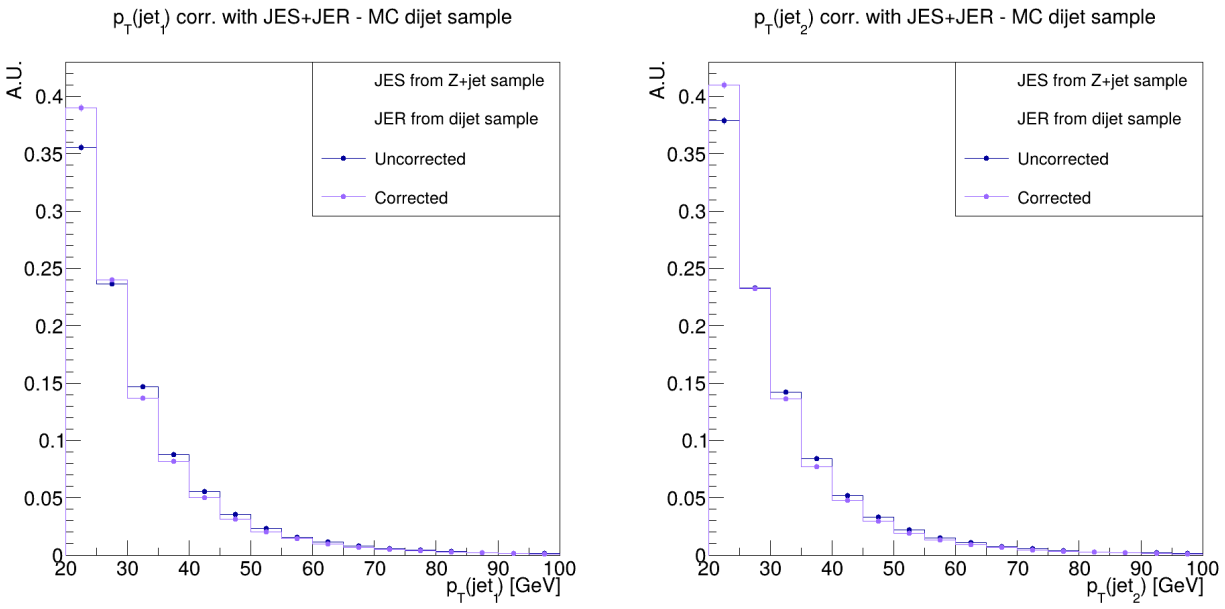

Figure 6.20: Comparison of the distributions of $p_T(jet_1)$ (left) and $p_T(jet_2)$ (right) with and without K^* and F^* corrections.

To properly determine the cross section, it is necessary to calculate the efficiency from the Monte Carlo, as it enters the formula in equation 6.1. The efficiency is defined mathematically as the ratio of the number of selected events to the number of generated events. However, because Jet Energy Scale and Jet Energy Resolution differ between data and MC, the K^* and F^* corrections derived from the calibration data are applied to account for these differences. These corrections are then included in the measure of the cross section as systematic uncertainties in the form of a modified efficiency. In the following, let the subscript *corr* represent quantities corrected with K^* and F^*. The efficiency ε_{corr} is expressed as:

$$\varepsilon_{corr} = \frac{N_{corr}}{N_{tot}} \tag{6.15}$$

where N_{corr} is the number of selected MC events after the corrections inside the fiducial region $2.2 < \eta(jet) < 4.2$, $20\,\text{GeV} < p_T(jet) < 100\,\text{GeV}$. The numbers of events before and after the application of the cut $\Delta\phi_{jet_{1,2}} > 1$ cut are shown in table 6.5.

The errors assigned to the efficiencies are calculated in a conservative manner, and the procedure to get them is the same for both the corrections, so let X stand for K^* and F^* equivalently. To compute the contribution of one of the two corrections, for the examined one three values are considered: the central correction itself, X, and the results of the sum and subtraction of the error, e_X, to the central correction value, named X^+ and X^-. The other correction is included through its central value only. In formulae:

$$X \pm e_X \quad \rightarrow \quad \begin{cases} X^+ = X + e_X \\ X^- = X - e_X \end{cases} \tag{6.16}$$

Then, the $p_T^{MC}(jet)$ distributions of the two jets in the dijet sample are corrected also with X^+ and X^-, while keeping the central value of the other correction, obtaining other two numbers of selected events, respectively N_X^+ and N_X^-, with which two additional efficiencies are derived:

$$\varepsilon_X^+ = \frac{N_X^+}{N_{tot}} \tag{6.17}$$

$$\varepsilon_X^- = \frac{N_X^-}{N_{tot}} \tag{6.18}$$

They are exploited to compute the error interval relative to the "+" and "-" X efficiency determination as:

$$\Delta\varepsilon_X^+ = \left|\varepsilon_{corr} - \varepsilon_X^+\right| \tag{6.19}$$

$$\Delta\varepsilon_X^- = \left|\varepsilon_{corr} - \varepsilon_X^-\right| \tag{6.20}$$

In table 6.12, the four "+" and "-" error contributions are shown.

Error on ε_{corr}	Value $[\times 10^{-4}]$
$\Delta\varepsilon_{K^*}^+$	0.479
$\Delta\varepsilon_{K^*}^-$	1.84
$\Delta\varepsilon_{F^*}^+$	1.13
$\Delta\varepsilon_{F^*}^-$	1.51

Table 6.12: "+" and "-" error contributions from the two corrections taken individually.

Then, to obtain a single error to assign to the corrected efficiency, the four contributions have been symmetrized by firstly summing the two "+" and two "-" values in quadrature:

$$\begin{cases} \Delta\varepsilon_{corr}^+ = \sqrt{\left(\Delta\varepsilon_{K^*}^+\right)^2 + \left(\Delta\varepsilon_{F^*}^+\right)^2} \\ \Delta\varepsilon_{corr}^- = \sqrt{\left(\Delta\varepsilon_{K^*}^-\right)^2 + \left(\Delta\varepsilon_{F^*}^-\right)^2} \end{cases} \quad \rightarrow \quad \Delta\varepsilon_{corr} = \frac{\Delta\varepsilon_{corr}^+ + \Delta\varepsilon_{corr}^-}{2} \tag{6.21}$$

Eventually, the value of the corrected efficiency is:

$$\varepsilon_{corr} = \left(2.3^{+0.1}_{-0.2}\right) \times 10^{-3} = (2.3 \pm 0.2) \times 10^{-3} \tag{6.22}$$

6.6 Dijet cross section measurement

Remembering the proportionality relation between the number of events and the cross section, the latter results in:

$$\sigma_{jj}^{data} = \frac{N_{data}}{\varepsilon_{corr} \cdot \varepsilon_{GEC} \cdot pre_{HLT} \cdot pre_{strip} \cdot \mathcal{L}} \cdot PS \tag{6.23}$$

where N_{data} is the number of events of the experimental dijet sample after the $\Delta\phi > 1$ cut, reported in table 6.5, and PS is the phase space factor introduced in section 6.2.1. It is the result of the ratio between the generator-level acceptance of the fast simulation implemented in this thesis to the analogous acceptance of the LHCb software, and comes from the different cuts they applied in the generation stage.

The resulting experimental cross section obtained from the analysis of the dijet sample is:

$$\sigma_{jj}^{data} = (1.307 \pm 0.009\ (\text{stat.}) \pm 0.1\ (\text{syst.}) \pm 0.03\ (\text{lumi.})) \times 10^7\,\text{pb} = (1.3 \pm 0.1) \times 10^7\,\text{pb} \tag{6.24}$$

where the error contributions are statistical, systematic due to K^* and F^* corrections, and systematic coming from the integrated luminosity contribution - constant and independent of the specific study. It is worth noticing that the leading uncertainty is given by the jet momentum corrections.

6.7 Preliminary α_s extraction and uncertainty determination with LHCb

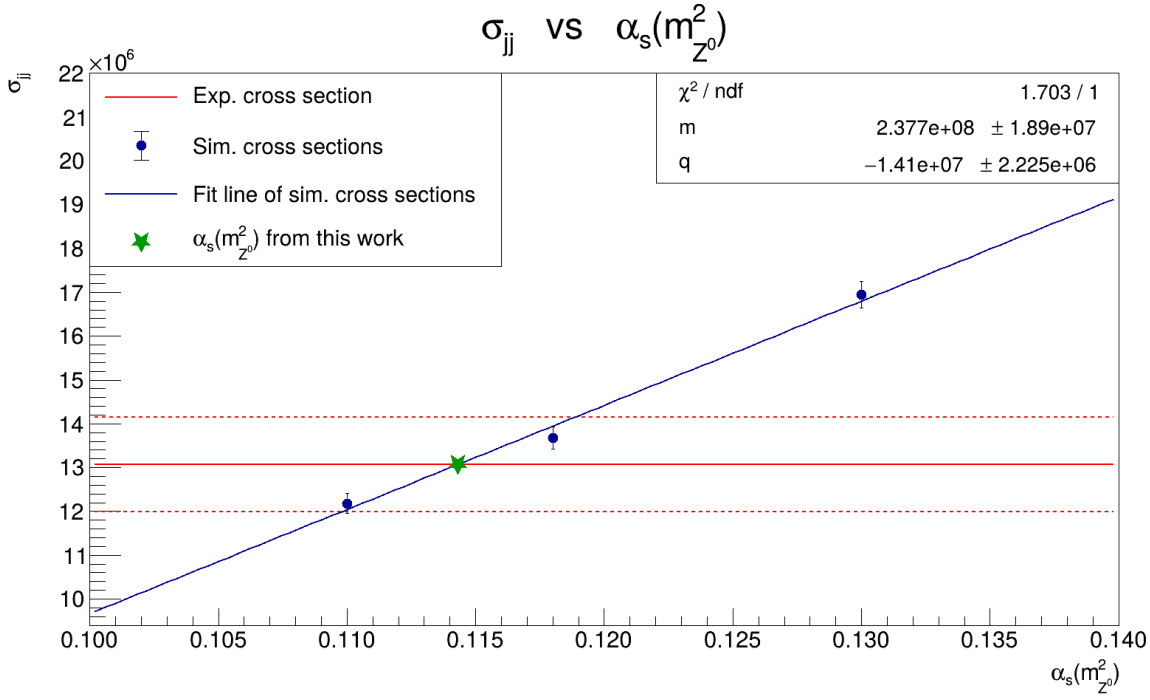

Figure 6.21: The blue points are the simulated weighted dijet cross sections σ_{jj}^{sim}, plotted as a function of α_s, and fitted with a line of the type $\sigma_{jj} = m\alpha_s + q$. The solid red line represents the measured dijet cross section σ_{jj}^{data}, whose error bar is limited by the two red dashed lines. The x coordinate of the intersection between the simulated cross sections blue fit line and the experimental cross section red horizontal line, marked with the green star, gives the estimate of the strong coupling constant obtained with the methods described in this work.

The building blocks for the determination of the strong coupling constant are all set. The simulated weighted dijet cross sections σ_{jj}^{sim} are plotted as a function of α_s in figure 6.21, and they are fitted with a line of the type $\sigma_{jj} = m\alpha_s + q$. This function is chosen empirically, and it is compatible within the errors of the simulated points. In the future, more data points will be needed to verify this trend. In the same plot, the experimental dijet cross section σ_{jj}^{data} measured as described in section 6.6 is added as a horizontal solid line, with its error bar limited by the two dashed lines. The x coordinate of the intersection between the simulated cross sections fit line and the experimental cross section horizontal line gives the estimate of the strong coupling constant, denoted with α_s^*:

$$\alpha_s^* = \frac{\sigma_{jj}^{data} - q}{m} \tag{6.25}$$

and the corresponding uncertainties are defined as the horizontal separation between the intersection of the dashed lines representing the error bar of σ_{jj}^{data}, placed at $\sigma_{jj}^{data} + e_{\sigma_{jj}^{data}}$ and $\sigma_{jj}^{data} - e_{\sigma_{jj}^{data}}$ (where $e_{\sigma_{jj}^{data}}$ represents the three contributions to the cross section uncertainty reported in equation 6.24, taken individually to reflect their corresponding contributions to the α_s uncertainty), and the fit line:

$$\Delta\alpha_s^* = \frac{2e_{\sigma_{jj}^{data}}}{m} \tag{6.26}$$

The final estimate of α_s at the reference scale corresponding to the Z^0 boson mass and of its statistical and systematic uncertainties results in (omitting the "*" superscript for clarity):

$$\alpha_s\left(m_{Z^0}^2\right) = 0.1143 \pm 0.0007\ (\text{stat.}) \pm 0.009\ (\text{syst.}) \pm 0.002\ (\text{lumi.}) = 0.114 \pm 0.009 \tag{6.27}$$

Since the errors of α_s are inferred from the measured dijet cross section, also here the main uncertainty contribution comes from K^* and F^* corrections. The statistical error is negligible, despite the applied trigger and stripping pre-scales.

The total α_s uncertainty amounts to about 7.9% of its value. Both the current world average [16] and the most precise single measurement of α_s to date, obtained by the ATLAS Collaboration in 2023 [27], give as a result $\alpha_s\left(m_{Z^0}^2\right) = 0.1180 \pm 0.0009$, the total uncertainty being about 0.76% of the central value. The uncertainty determined in this work is larger than those of the two results that have just been reported by about one order of magnitude. The uncertainty of the world average is dominated by hadronic final states of e^+e^- annihilations data, while the ATLAS single measurement is limited by the PDF uncertainty. Even so, the estimate of α_s extracted here is compatible with the present world average.

There are several studies that exploited final-state jets for determining α_s. For example, in the CMS analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ [94, 95] the estimate of the strong coupling constant at the Z^0 boson mass is $\alpha_s\left(m_{Z^0}^2\right) = 0.1188 \pm 0.0030$ for NLO computations, and $\alpha_s\left(m_{Z^0}^2\right) = 0.1170 \pm 0.0019$ for NNLO ones; in the ATLAS determination of the strong coupling constant from transverse energy-energy correlations in multijet events at $\sqrt{s} = 13\,\text{TeV}$ [96], α_s was obtained with two methods: the simultaneous fit to all transverse energy-energy correlation (TEEC) distributions across different kinematic regions, which yielded to $\alpha_s\left(m_{Z^0}^2\right) = 0.1175 \pm 0.0006\,(\text{exp.}) \pm 0.0034\,(\text{theo.})$, and the global fit to the asymmetry (ATEEC) distributions, which led to $\alpha_s\left(m_{Z^0}^2\right) = 0.1185 \pm 0.0009\,(\text{exp.})^{+0.0025}_{-0.0012}\,(\text{theo.})$. The CMS NLO value of α_s is dominated by the jet scales uncertainty, while both CMS and ATLAS calculations at NNLO produce an estimate of α_s that is controlled by theoretical uncertainties, related to the fit uncertainties in the used measurements in the case of CMS, and to variations in the modelling of parton showers and hadronization in the Monte Carlo simulations in the case of ATLAS. In these two studies, the total uncertainty lies in the 0.001 order of magnitude, the same one of the uncertainty calculated as presented in this thesis.

Chapter 7

Future improvements

Several tweaks and additions can be considered to expand the work done for this thesis, with the aim of getting to a precise determination of the strong coupling constant with the LHCb detector. Here is a list of the most relevant possible advancements.

Enhanced simulation study. The fast simulation implemented in this work may be made more accurate in predicting the behavior of the analyzed processes by generating samples at least at Next-to-Leading Order as perturbative level, which should provide more realistic simulated results. Particular attention should be paid in matching the energy scale set at generator stage in `MadGraph5_aMC@NLO` to that of the fragmentation stage configured in `Pythia8`.

Differential measurements. To extract α_s in this work, only integral quantities have been studied. Increasing the statistics of the simulated samples could allow to investigate differential measurements. A list of observables for the differential cross section measurements includes, but is not limited to: the transverse momentum, pseudorapidity, azimuthal angle and mass of each single jet in an event; the combined previous values of all possible dijet systems; the mass of trijet systems. Differential measurements offer several advantages. They provide information about the behavior of processes as a function of specific variables, and this increased granularity can enhance the sensitivity to specific physical effects that may not be visible otherwise. Also, by focusing on specific kinematic variables, differential measurements can reduce the impact of certain systematic uncertainties, and this can lead to more precise determinations of physical quantities such as cross sections. Moreover, they allow for a better discrimination between various contributions to a process or between competing theoretical models, especially when exploring different kinematic regions, and this can permit the validation and refinement of theoretical frameworks. Lastly, differential measurements can reveal deviations from Standard Model predictions, potentially indicating the presence of new physics beyond it.

Multi-jet events. From the plot of the weighted simulated cross section as a function of events with $\geq n$ jets in figure 5.12, it is evident that also other bins besides the ≥ 2 analyzed here can be looked at. In particular, events with three hadronic jets can be examined. This will give the possibility to compute one of the multi-jet α_s determination observables: the ratio of the number of events with three jets to the number of events with two jets, R_{32}. As commented in section 5.3.6, this variable is suitable mainly because requiring one more jet in the final state leads to three-jet cross sections, which are sensitive to α_s^3, while two-jet cross sections are sensitive to α_s^2, and, compared to the three- and two-jet cross sections taken individually, R_{32} has the advantage that numerous uncertainties cancel in the ratio, at least partially. In fact, ratios of event rates, such as R_{32}, help reduce the sensitivity to experimental uncertainties related to luminosity measurements and detector effects. This makes such observable more robust for precision measurements of α_s. From a theoretical point of view, QCD predicts how the jet multiplicity distribution depends on α_s, and so, by comparing experimental measurements of the jet multiplicity ratio to these predictions, one can extract the value of α_s and test the validity of QCD. ATLAS and CMS have successfully performed R_{32} calculations to measure α_s, so it would be very useful to provide a measurement from LHCb, which works in complementary

kinematic ranges with respect to them, both to check whether LHCb is able to get close to their performances, and also with the ultimate goal to reduce the global α_s uncertainty.

$Z^0 \to \mu^+\mu^-$ analysis. Other approaches and physics channels can be studied in order to estimate the strong coupling constant. A remarkable example is offered in the paper published by the ATLAS Collaboration in 2023 [27], which obtained the most precise single measurement of α_s ever, compatible with the world average. In their work, ATLAS physicists studied the production of Z^0 bosons in pp data collected at a center of mass energy of 8 TeV. Z^0 bosons are typically produced when two quarks in the colliding protons annihilate, which is a process governed by the weak interaction. The strong force comes into play through the emission of gluons from the initial quarks. This process causes the Z^0 boson to recoil in the plane transverse to the beam axis: the size of this "kick" is measured as transverse momentum and depends on α_s. However, the Z^0 boson decays instantly after being produced, so its transverse momentum is measured via its decay products, which are two electrons or muons. Also considerations on the Z^0 boson rapidity are included. The comparison of the cross section of Z^0 boson production versus transverse momentum and rapidity with theoretical calculations, provided the precise determination of the strong coupling constant. This approach is favorable, because, since in this process the final-state particles are not subject to the strong interaction, the theoretical complications and uncertainties are reduced, gathering advisable features: large sensitivity to α_s relative to the experimental precision, and better accuracy of the theoretical predictions, enabled by the computation of some perturbative QCD corrections at high level. The $Z^0 \to \mu^+\mu^-$ LHCb data can be employed to estimate α_s. The LHCb detector, indeed, has demonstrated to be capable of reconstructing this Z^0 decay process with high precision, having trigger lines that identify muons with very high precision, so the $Z^0 \to \mu^+\mu^-$ samples are very good candidates to analyze for the purpose of extracting an α_s estimate.

Run 3 and Upgrades. The ongoing LHCb Run 3 and the planned setups for its Upgrades will enable a better jet measurement, especially of jet energy. Also, new dedicated trigger lines will be available for jet selection. Thus, it will be possible to repeat the analysis presented in this thesis, adapting it to Run 3 data samples in order to reach a more accurate determination of α_s and of its uncertainty.

LHCb electromagnetic calorimeter development. The LHCb electromagnetic calorimeter plays a crucial role in measuring the jet energy, and calorimetric clusters are fundamental inputs to the jet reconstruction algorithms in LHCb. New calorimetric clustering techniques exploiting Machine Learning methods should improve the measurement of the energy deposited in ECAL, and this consequently benefits the goodness of the reconstructed jets.

γ+jet sample. A direct consequence of the development of the electromagnetic calorimeter would reflect in the possibility of employing γ+jet samples, containing a high-energy photon produced together with a jet, in place of or in addition to the Z+jet sample to extract the K^* and F^* correction factors measured in this thesis. In fact, the present configuration of the electromagnetic calorimeter does not allow a precise reconstruction of high-energy photons, mainly because of the saturation of calorimeter cells.

Conclusions

The target of this thesis work was to estimate the value and the leading uncertainties of the strong coupling constant, that at present amount to $\alpha_s\left(m^2_{Z^0}\right) = 0.1180 \pm 0.0009$ [16]. For the first time, experimental and simulated data from the LHCb detector collected in 2016 during LHC Run 2 at a center of mass energy of 13 TeV and containing hadronic jets were exploited for this purpose. Among many other possible single- and multi-jet distributions, the selected observable to extract α_s and its uncertainties is the inclusive dijet cross section. It has been measured from a dijet sample, and two correction factors - related to Jet Energy Scale and Jet Energy Resolution - required to match the Monte Carlo reconstructed jet energy to the experimental one were calculated, and their effect has been quantified in the measured dijet cross section in the form of a corrected efficiency. This, in turn, propagated inside the determination of the α_s systematic uncertainty, as the three main contributions considered to obtain it are precisely the two correction factors and the integrated luminosity of 2016 data-taking, $\mathcal{L} = 1.6\,\text{fb}^{-1}$, which carries a constant relative uncertainty of 2%, independent of the specific process under study. The statistical error has been evaluated as well. The strong coupling constant itself has been extrapolated through the comparison between the experimental dijet cross section and the linear fit of three simulated dijet cross sections at different α_s values: 0.1100, 0.1180, 0.1300. The values of the inclusive dijet cross section and of the strong coupling constant estimated in this work read:

$$\sigma_{jj}^{data} = (1.307 \pm 0.009\ (\text{stat.}) \pm 0.1\ (\text{syst.}) \pm 0.03\ (\text{lumi.})) \times 10^7\,\text{pb} = (1.3 \pm 0.1) \times 10^7\,\text{pb}$$
$$\alpha_s\left(m^2_{Z^0}\right) = 0.1143 \pm 0.0007\ (\text{stat.}) \pm 0.009\ (\text{syst.}) \pm 0.002\ (\text{lumi.}) = 0.114 \pm 0.009$$

This first attempt at extrapolating the strong coupling constant with LHCb data is deemed successful. The derived value is compatible with the average value, and the primary outcome of the analysis presented here is the fact that the statistics is not the limiting factor to this approach, since the uncertainty coming from the corrections applied to the jets transverse momentum distributions is the leading contribution to the total α_s uncertainty. Therefore, there is no need to consider larger data samples or to combine different data-takings, but it is crucial to refine the techniques of correcting the jets p_T in order to obtain a more precise determination of the strong coupling constant with the methods employed in this thesis.

Bibliography

[1] S. L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[2] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, November 1967.

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

[4] S. L. Glashow, J. Iliopoulos and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, 2:1285–1292, October 1970.

[5] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [Electroweak Model and Constraints on New Physics].

[6] H. Georgi and S. L. Glashow. Unity of All Elementary-Particle Forces. *Phys. Rev. Lett.*, 32:438–441, February 1974.

[7] H. Fritzsch and P. Minkowski. Unified Interactions of Leptons and Hadrons. *Annals of Physics*, 93(1):193–266, 1975.

[8] R. L. Workman *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2022(8):083C01, August 2022 [Grand Unified Theories].

[9] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [Light Quarks (u,d,s), c Quark, b Quark, t Quark].

[10] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [Electron, Muon, Tau].

[11] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [Mass and Width of the W Boson, Z Boson].

[12] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [H].

[13] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [Quantum Chromodynamics].

[14] D. J. Griffiths. *Introduction to Elementary Particles*. Wiley-VCH, Second, Revised edition, 2008 [Section 8.6].

[15] A. Bettini. *Introduction to Elementary Particle Physics*. Cambridge University Press, Second edition, 2014 [Section 6.5].

[16] S. Navas *et al.* [Particle Data Group Collaboration]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, August 2024 [Physical Constants].

[17] M. Creutz. *Quarks, Gluons and Lattices*. Oxford University Press, 1983.

[18] A. S. Kronfeld *et al.* [USQCD Collaboration]. Lattice QCD and Particle Physics, July 2022.

[19] J. D. Bjorken. Asymptotic Sum Rules at Infinite Momentum. *Phys. Rev.*, 179:1547–1553, March 1969.

[20] J. D. Bjorken and E. A. Paschos. Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon. *Phys. Rev.*, 185:1975–1982, September 1969.

[21] H. Abramowicz, I. Abt, L. Adamczyk, M. Adamus, R. Aggarwal, S. Antonelli, P. Antonioli, A. Antonov, M. Arneodo, V. Aushev, Y. Aushev, O. Bachynska, A. Bamberger, A. Barakbaev, G. Barbagli, G. Bari, F. Barreiro, N. Bartosik, D. Bartsch, and Z. Zulkapli. Measurement of beauty production in deep inelastic scattering at HERA using decays into electrons. *European Physical Journal C*, 71:1–15, January 2011.

[22] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump and C.-P. Yuan. CT10 next-to-next-to-leading order global analysis of QCD. *Physical Review D*, 89(3), February 2014.

[23] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[24] L. N. Lipatov. The parton model and perturbation theory. *Sov. J. Nucl. Phys.*, 20(1):94–102, 1975.

[25] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nucl. Phys.*, B126:298–318, 1977.

[26] Y. L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[27] ATLAS Collaboration. A precise determination of the strong-coupling constant from the recoil of Z bosons with the ATLAS experiment at $\sqrt{s} = 8\,\text{TeV}$, 2023.

[28] M. A. Winn. Recent results on heavy flavour production at LHCb. *7th Workshop "Implications of LHCb Measurements and Future Prospects"*, CERN, Geneva, Switzerland, 8-10 Nov 2017.

[29] R. D. Ball, S. Carrazza, J. Cruz-Martinez, L. Del Debbio, S. Forte, T. Giani, S. Iranipour, Z. Kassabov, J. I. Latorre, E. R. Nocera, R. L. Pearson, J. Rojo, R. Stegeman, C. Schwan, M. Ubiali, C. Voisey, and M. Wilson. An open-source machine learning framework for global analyses of parton distributions, 2021.

[30] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. Also published by CERN Geneva in 2010.

[31] LHCb Collaboration. *LHCb magnet: Technical Design Report*. CERN, Geneva, 2000.

[32] LHCb Collaboration. *LHCb VELO (VErtex LOcator): Technical Design Report*. CERN, Geneva, 2001.

[33] J. Gassner, M. Needham and O. Steinkamp. Layout and Expected Performance of the LHCb TT Station. Technical report, CERN, Geneva, 2004.

[34] LHCb Collaboration. *LHCb inner tracker: Technical Design Report*. CERN, Geneva, 2002. Revised version number 1 submitted on 2002-11-13 14:14:34.

[35] LHCb Collaboration. *LHCb outer tracker: Technical Design Report*. CERN, Geneva, 2001.

[36] LHCb Collaboration. *LHCb RICH: Technical Design Report*. CERN, Geneva, 2000.

[37] LHCbRICH Collaboration. LHCb RICH1 Engineering Design Review Report. Technical report, CERN, Geneva, 2005.

[38] LHCbRICH Collaboration. LHCb RICH 2 Engineering Design Review Report. Technical report, CERN, Geneva, 2002. Revised version number 1 submitted on 2002-05-21 14:24:22.

[39] LHCb Collaboration. *LHCb calorimeters: Technical Design Report*. CERN, Geneva, 2000.

[40] S. Filippov, Y. Gavrilov, E. Guschin, V. Klubov, L. Kravchuk, S. Laptev, V. Postoev and A. Sadovskii. Design and Construction of the LHCb Scintillator Pad/Preshower Detector. Technical report, CERN, Geneva, 2000.

[41] S. Barsuk, A. Golutvin, V. Kochetkov, I. Korolko, S. Malyshev, V. Mayatski, V. Rusinov, A. Soldatov and E. Tarkovski. Design and construction of electromagnetic calorimeter for LHCb experiment. Technical report, CERN, Geneva, 2000.

[42] LHCb Calorimeter group. Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC, 2020.

[43] C. Coca, T. Preda, A. Rosca, I. Ajinenko, A. E. Dorokhov, R. I. Dzhelyadin, A. K. Konoplyannikov, V. Matveev, V. Novikov, O. P. Yushchenko and Y. Ranyuk. The hadron calorimeter prototype beam-test results. Technical report, CERN, Geneva, 2000.

[44] LHCb Collaboration. *LHCb muon system: Technical Design Report*. CERN, Geneva, 2001.

[45] LHCb Collaboration. *LHCb muon system: addendum to the Technical Design Report*. CERN, Geneva, 2003.

[46] LHCb Collaboration. *LHCb muon system: second addendum to the Technical Design Report*. CERN, Geneva, 2005. Submitted on 9 Apr 2005.

[47] LHCb Collaboration. *LHCb trigger system: Technical Design Report*. CERN, Geneva, 2003. Revised version number 1 submitted on 2003-09-24 12:12:22.

[48] S. Tolk, J. Albrecht, F. Dettori and A. Pellegrino. Data driven trigger efficiency determination at LHCb. Technical report, CERN, Geneva, 2014.

[49] LHCb Collaboration. *LHCb online system, data acquisition and experiment control: Technical Design Report*. CERN, Geneva, 2001.

[50] N. Tatsuya, E. Aslanides, A. Smith and W. Witzeling. Addendum to the LHCb Online System Technical Design Report. Technical report, CERN, Geneva, 2005. Revised version submitted on 2005-11-04 15:04:16.

[51] LHCb Starterkit. The LHCb data flow.

[52] LHCb Starterkit. Changes to the data flow in Run 2.

[53] G. Corti, G. Barrand, I. Belyaev, M. Cattaneo, P. Charpentier, M. Frank, P. Koppenburg, P. Mato-Vila, F. Ranjard and S. Roiser. Software for the LHCb experiment. *IEEE Trans. Nucl. Sci.*, 53:1323–1328, 2006.

[54] R. Brun and F. Rademakers. ROOT - An object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1):81–86, 1997. New Computing Techniques in Physics Research V.

[55] M. Clemencic, G. Corti, S. Easo, C. R. Jones, S. Miglioranzi, M. Pappagallo and P. Robbe (on behalf of the LHCb Collaboration). The LHCb Simulation Application, Gauss: Design, Evolution and Experience. *Journal of Physics: Conference Series*, 331(3):032023, December 2011.

[56] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, C. T. Preuss, T. Sjöstrand, P. Skands, M. Utheim and R. Verheyen. A comprehensive guide to the physics and usage of PYTHIA 8.3, 2022.

[57] D. J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152–155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

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

[59] W. Herr and B. Muratori. Concept of luminosity. *CAS - CERN Accelerator School: Intermediate Accelerator Physics*, pages 361–378, 2006.

[60] C. Barschel. Precision luminosity measurement at LHCb with beam-gas imaging. *RWTH Aachen University*, 2014. Presented 05 Mar 2014.

[61] LHCb Collaboration. Precision luminosity measurements at LHCb. *Journal of Instrumentation*, 9(12):P12005–P12005, December 2014.

[62] M. Ferro-Luzzi. Proposal for an absolute luminosity determination in colliding beam experiments using vertex detection of beam-gas interactions. *Nucl. Instrum. Methods Phys. Res., A*, 553(3):388–399, 2005.

[63] LHCb Collaboration. The LHCb Upgrade I. *Journal of Instrumentation*, 19(05):P05065, May 2024.

[64] LHCb Collaboration. LHCb PLUME: Probe for LUminosity MEasurement. Technical report, CERN, Geneva, 2021.

[65] LHCb Collaboration. LHCb SMOG Upgrade. Technical report, CERN, Geneva, 2019.

[66] LHCb Collaboration. Expression of Interest for a Phase-II LHCb Upgrade: opportunities in flavour physics, and beyond, in the HL-LHC era. Technical report, CERN, Geneva, 2017.

[67] E. M. Metodiev. The Fractal Lives of Jets.

[68] G. P. Salam. Towards jetography. *The European Physical Journal C*, 67(3–4):637–686, May 2010.

[69] W. Barter. Z boson and associated jet production at the LHCb experiment. *Cambridge University (main)*, 2014. Presented 27 May 2014.

[70] F. Beaudette. The CMS Particle Flow Algorithm, 2014.

[71] P. Astier, A. Cardini, R. D. Cousins, A. A. Letessier-Selvon, B. A. Popov and T. G. Vinogradova. Kalman Filter Track Fits and Track Breakpoint Analysis. *Nucl. Instrum. Methods Phys. Res., A*, 450(1):138–54, 2000.

[72] L. Sestini. Search for a Higgs boson decaying to a pair of b quarks in the forward region of pp collisions with the LHCb detector. *Padua University*, 2017. Presented 09 Mar 2017.

[73] LHCb collaboration. Study of forward Z + jet production in pp collisions at $\sqrt{s} = 7\,\text{TeV}$. *Journal of High Energy Physics*, January 2014.

[74] V. Salustino Guimarães, M. Rangel, V. Coco and C. Potterat. Energy-Momentum ratio of charged particles at LHCb. Technical report, CERN, Geneva, 2013.

[75] G. Sterman and S. Weinberg. Jets from Quantum Chromodynamics. *Phys. Rev. Lett.*, 39:1436–1439, December 1977.

[76] M. Cacciari, G. P. Salam and G. Soyez. FastJet user manual. *The European Physical Journal C*, 72(3), March 2012.

[77] M. Cacciari, G. P. Salam and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, page 063–063, April 2008.

[78] S. Catani, Y. L. Dokshitzer, M.H. Seymour and B.R. Webber. Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions. *Nuclear Physics B*, 406(1):187–224, 1993.

[79] S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Physical Review D*, 48(7):3160–3166, October 1993.

[80] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber. Better jet clustering algorithms. *Journal of High Energy Physics*, page 001–001, August 1997.

[81] M. Wobisch and T. Wengler. Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering, 1999.

[82] G. P. Salam and G. Soyez. A practical Seedless Infrared-Safe Cone jet algorithm. *Journal of High Energy Physics*, 2007(05):086–086, May 2007.

[83] S. Catani and D. Zeppenfeld. Jet Algorithms. In *1st Les Houches Workshop on Physics at TeV Colliders*, pages 132–140, May 2000.

[84] G. Lanfranchi, X. Cid Vidal, S. Furcas, M. Gandelman, J. A. Hernando, J. H. Lopez, E. Polycarpo and A. Sarti. The Muon Identification Procedure of the LHCb Experiment for the First Data. Technical report, CERN, Geneva, 2009.

[85] A. Sarti, S. Furcas, G. Lanfranchi and M. Palutan. Calibration Strategy and Efficiency measurement of the Muon Identification procedure at LHCb. Technical report, CERN, Geneva, 2010.

[86] R. Aaij, J. Albrecht, F. Dettori, K. Dungs, H. Lopes, D. Martinez Santos, J. Prisciandaro, B. Sciascia, V. Syropoulos, S. Stahl and R. Vazquez Gomez. Optimization of the muon reconstruction algorithms for LHCb Run 2. Technical report, CERN, Geneva, 2017.

[87] M. Xu, H. Yin and S. Farry. Measurement of high-p_T muon reconstruction efficiencies for studies of γ^*/Z production and angular coefficients with LHCb RunII data-sets. Technical report, CERN, Geneva, 2020.

[88] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, July 2014. Version 3.5.2.

[89] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr and G. Watt. LHAPDF6: parton density access in the LHC precision era. *The European Physical Journal C*, 75(3), March 2015.

[90] M. Dobbs and J. B. Hansen. The HepMC C++ Monte Carlo Event Record for High Energy Physics. Technical report, CERN, Geneva, 2000. Revised version number 1 submitted on 2001-02-27 09:54:32.

[91] CMS Collaboration. Measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section in pp collisions at $\sqrt{s} = 7TeV$ and first determination of the strong coupling constant in the tev range. *The European Physical Journal C*, 73(10), October 2013.

[92] LHCb Collaboration. Inclusive W and Z production in the forward region at $\sqrt{s} = 7\,\text{TeV}$. *Journal of High Energy Physics*, June 2012.

[93] LHCb Collaboration. Measurement of the cross-section for $Z \to \mu\mu$ production with 1 fb^{-1} of pp collisions at $\sqrt{s} = 7\,\text{TeV}$. *LHCb-CONF-2013-007*, April 2013.

[94] CMS Collaboration. Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$. Technical report, CERN, Geneva, 2021.

[95] CMS Collaboration. Addendum to: Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Journal of High Energy Physics*, December 2022.

[96] ATLAS Collaboration. Determination of the strong coupling constant from transverse energy-energy correlations in multijet events at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *Journal of High Energy Physics*, July 2023.