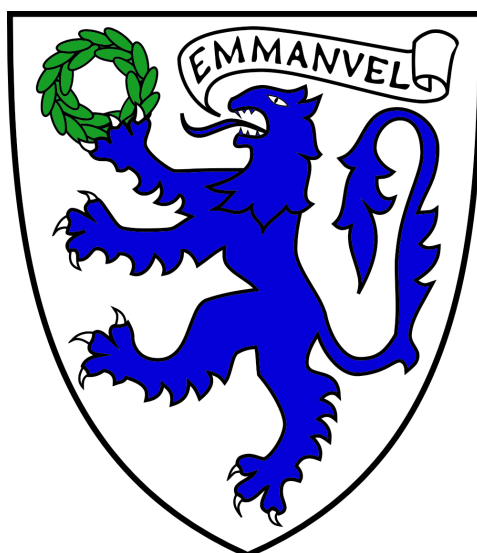


Z boson and associated jet production at the LHCb experiment

William James Barter
of
Emmanuel College,
University of Cambridge



This dissertation is submitted
for the degree of Doctor of Philosophy
June 2014



Abstract

The LHCb experiment at the Large Hadron Collider (LHC) is a fully instrumented forward-arm spectrometer initially designed to study the decays of heavy flavour particles produced close to the beam-line ($2.0 < \eta < 4.5$). This unique precision coverage of the forward region at the LHC means that the experiment is also able to make measurements of massive electroweak vector boson production which are complementary to those made at the general purpose detectors. Such measurements are sensitive to the inner structure of the proton, probing parton distribution functions at low values of Bjorken- x . These measurements also provide precision tests of perturbative quantum chromodynamics and the Standard Model.

This thesis considers the production of Z bosons in association with jets at LHCb in events where the Z bosons decay directly to two muons. Jet reconstruction at LHCb makes use of a particle flow algorithm which enables the majority of the jet content to be measured using the LHCb tracking systems. This algorithm enables jets to be reconstructed with a transverse momentum resolution of about 11-13%. The measurement of the Z+jet production cross-section is made above two jet transverse momentum thresholds: 10 and 20 GeV, using data collected by the LHCb experiment in 2011 corresponding to an integrated luminosity of about 1 fb^{-1} . The fiducial region for these measurements is defined by requiring that both the Z boson and the jet lie within the LHCb detector's acceptance. Measurements are also made of the ratio of these cross-sections to the inclusive Z boson production cross-section, and of the differential cross-sections with respect to kinematic variables associated with the Z+jet events. These results show good agreement with theoretical predictions based on the Standard Model, evaluated by matching fixed order calculations using POWHEG to PYTHIA for simulation of the parton shower development and hadronisation. The thesis concludes by considering the impact of related future measurements.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation has not been submitted for another qualification to this or any other university, and does not exceed the word limit for the respective degree committee.

William James Barter

*In loving memory of my Grandmother,
Lucy Caroline Simmons*

Acknowledgements

There are many people who I would like to thank for their support, help and advice throughout my studies. In particular, my supervisor, David Ward, has been a tremendous source of fruitful discussion and ideas. His patience and guidance have been instrumental during my research. I also wish to thank Val Gibson and James Stirling for making the time to answer my questions on the LHCb experiment, perturbative quantum chromodynamics and parton distribution functions.

I have also made use of the opportunity to discuss predictions of proton-proton collisions at the LHC with Keith Hamilton, Robert Thorne and Bryan Webber. These conversations have been instrumental in my studies. I have also relied on colleagues on the LHCb experiment, both within Cambridge and throughout the collaboration. The Cambridge LHCb group have always been willing to help with the details of compiling and debugging code. Likewise, the Electroweak Working Group and the Tools for Jet Reconstruction Working Groups within the LHCb collaboration have provided useful feedback. In particular, I would like to thank Jonny Anderson, Simone Bifani, Victor Coco, Susan Haines, Chris Jones, Ronan McNulty, Katharina Müller, Murilo Rangel, Tara Shears, Simon Wright and many others for their support. Others in the Cambridge High Energy Physics group have also been instrumental in my studies. John Hill and Steve Wotton's support of the group's computational infrastructure have made this thesis possible. John Chapman, John Marshall and Sarah Williams have always provided a useful sounding board and have been happy to discuss approaches over tea or coffee. I would also like to thank the Emmanuel College graduate community for their friendship. These studies were made possible by funding from the STFC and Emmanuel College, Cambridge.

Finally I wish to thank my friends and family. Without their support, their offers of tea, and their common sense, I would not have been able to write this thesis. My parents have made me who I am: my studies would not have amounted to anything without their support, love, and encouragement.

Preface

This thesis describes my work on various aspects of the LHCb research programme. I present this review of the content so that those reading are aware of where I have concentrated my studies over the last few years.

The main body of the thesis begins with a short introduction, followed by a discussion of the theoretical framework that underpins our current understanding of particle physics. This review is based on the work of others, which is cited.

I then discuss the LHCb detector and data collection. Parts of this chapter therefore also review the work of others. However, within the LHCb ‘QCD, Electroweak and Exotica’ Working Group I played a major role in the event selection, and was responsible for the implementation of the software trigger lines and offline selection requirements. This work is therefore presented in some detail.

The chapter on jet reconstruction is a summary of the work of the LHCb ‘Tools for jet reconstruction’ working group. I made major contributions to the work of this group. I was jointly responsible for the initial development of the LHCb particle flow algorithms, which was subsequently further developed by others. I was also responsible for studies of the jet reconstruction performance at LHCb.

The LHCb jet reconstruction is then put to use in the final analysis chapters. These chapters detail the LHCb measurement of forward Z+jet production, published in JHEP in January 2014. I was the lead analyst within the collaboration for this measurement, which was the first published LHCb jet measurement.

I conclude the thesis by considering related measurements of interest that can be made in the coming years.

William James Barter

Contents

1	Introduction	3
1.1	Co-ordinates at hadron colliders	6
2	Theoretical overview	9
2.1	The Standard Model of particle physics	9
2.1.1	Light, matter, and nuclear decays	12
2.1.2	The strong interaction	17
2.2	Theoretical predictions of Standard Model processes	19
2.2.1	The factorisation theorem	20
2.2.2	Determining the final state	21
2.2.3	Jets	23
2.2.4	Parton distribution functions	27
2.3	Computational programs for predictions at hadron colliders	31
3	The LHCb experiment at the LHC	35
3.1	The LHC	35
3.2	The LHCb experiment	38
3.2.1	Beampipe	42
3.2.2	Vertex Locator	42
3.2.3	The LHCb magnet and tracking detectors	47
3.2.4	Ring imaging Cherenkov sub-detectors	50
3.2.5	Electromagnetic calorimeter and associated subsystems	55
3.2.6	Hadronic calorimeter	57
3.2.7	Muon chambers	58
3.2.8	The LHCb trigger	61
3.2.9	Detector readout and management	68
3.2.10	Recorded integrated luminosity and running conditions	68
3.3	Event Reconstruction and the LHCb software environment	72
3.3.1	Track reconstruction at LHCb	72

3.3.2	Particle Identification at LHCb	74
3.3.3	The LHCb software environment	75
3.3.4	Simulating proton-proton interactions at LHCb	76
3.4	Offline Data Storage	76
4	Jet Reconstruction at LHCb	79
4.1	The particle flow algorithm	80
4.2	Jet identification selection	86
4.3	Jet energy scale	87
4.4	Jet reconstruction performance	92
4.5	Summary of jet reconstruction at LHCb	100
5	Production of Z bosons in association with jets at LHCb	101
5.1	Overview of electroweak production measurements in the forward region	102
5.2	Event selection	106
5.2.1	Event samples	108
5.2.2	Selection requirements	109
5.2.3	Kinematic distributions of selected events	111
5.3	Background	117
5.4	Z boson detection and reconstruction efficiencies	121
5.4.1	Muon identification	122
5.4.2	Muon track reconstruction efficiency	129
5.4.3	Muon trigger efficiency	135
5.4.4	Global event cut efficiency	136
5.5	Summary	139
6	Measuring the Z+jet production cross-section at LHCb at $\sqrt{s} = 7$ TeV	145
6.1	Defining the measurement	145
6.2	Jet transverse momentum resolution and jet detection efficiencies	147
6.2.1	Unfolding the jet p_T distribution	149
6.2.2	Jet reconstruction and selection efficiency	155
6.3	Cross-section measurement	157
6.3.1	Final State Radiation (FSR)	160
6.4	Systematic uncertainties	161
6.4.1	Z boson reconstruction and selection systematic uncertainties . . .	162
6.4.2	Jet energy scale	162
6.4.3	Jet energy resolution	169

6.4.4	Unfolding	170
6.4.5	Background and Fake Jets	170
6.4.6	Jet reconstruction and selection efficiency	171
6.4.7	Final State Radiation	175
6.4.8	Luminosity	175
6.5	Cross-checks	176
6.5.1	Measuring the Z production cross-section	176
6.5.2	Magnet polarity and charge asymmetries	176
6.5.3	Closure tests	178
6.6	Results	178
6.7	Summary	190
7	Related measurements for the coming years	191
7.1	Z+jet measurements using data recorded in 2012	191
7.2	Other measurements at LHCb	196
8	Conclusions	201
	Bibliography	209
	List of Figures	221
	List of Tables	227

“Look deep into nature, and then you will understand everything better.”
— Albert Einstein

Chapter 1

Introduction

“When it comes to the world around us, is there any choice but to explore?”

— Lisa Randall

Few scientific experiments have caught the public imagination like the Large Hadron Collider (LHC). The first day of operation was marked across the globe by mass media coverage. In the United Kingdom the event dominated the news, and particle physics has been catapulted to the front of the public perception of science.

The first data-taking period at the LHC has just finished, and the detectors and machine are now being prepared for higher energy collisions. Analysis of the initial data has led to many interesting scientific discoveries: the most obvious success is the discovery of a particle with properties consistent with those of the Standard Model Higgs boson by both the ATLAS [1] and the CMS [2] experiments. There have also been exciting developments away from the LHC. Recent results have shown that the neutrino sector contains considerable mixing [3] between all three neutrino flavours. Regardless of future measurements, the success of the Standard Model can be described as the crowning achievement of modern particle physics: it has provided accurate predictions for a range of physics processes at many different experiments. Neither direct nor indirect detection measurements at the LHC or elsewhere have seen conclusive evidence for supersymmetry

or other models of new physics; for the last forty years the Standard Model has passed every test¹.

Despite this success there are still major questions to be answered: we have yet to detect a particle which could account for astrophysical observations of dark matter, and the Standard Model does not account for gravity (though this does not provide problems for predictions at current experiments, since gravity is extremely weak at TeV energy scales). In addition, our current knowledge of the Standard Model suggests a degree of fine-tuning; over the coming years the LHC will provide stringent tests of naturalness. There is still much to be understood about the Standard Model and about nature at the TeV scale.

Precision measurements seek to study the Standard Model in detail: some directly extract Standard Model parameters (for example, the weak mixing angle) from data and test whether the measured values agree with other constraints. Others attempt to compare well measured quantities at colliders (for example, production cross-sections) with precise theoretical predictions. Such measurements can inform us about the behaviour of the Standard Model, testing both its structure and the assumptions and approaches taken in evaluating predictions. We discuss one such measurement here.

This thesis documents the measurement of forward Z boson and associated jet ('Z+jet') production, in events where the Z boson directly decays to two muons, at the LHCb experiment, based on data collected in 2011 corresponding to an integrated luminosity of about 1 fb^{-1} . These studies were also presented in Ref. [4], and were the first published results from LHCb to include jet reconstruction. This measurement provides a test of perturbative quantum chromodynamics (pQCD) in the forward region. It also provides sensitivity to the parton distribution functions (PDFs) that describe the inner structure of the proton, and therefore also tests the accuracy of different PDF parametrisations and the DGLAP equations that govern their evolution. Such studies build on earlier measurements of inclusive W and Z boson production at LHCb [5, 6].

The motivation for the Z+jet measurement can also be viewed from another angle: measurements of well-predicted Standard Model quantities provide useful benchmark calibrations of the experiment performance. If the measurement agrees with the Standard Model then there will be increased confidence in future work using the same techniques. This thesis therefore also discusses related analyses that can be carried out in the near

¹Neutrino mixing and neutrino masses were not originally included within the Standard Model, and some therefore argue that they are the first evidence for physics beyond the Standard Model. However, neutrino masses can be (and have been) incorporated into the theory.

future. We also consider the prospects of repeating the measurement using the larger sample of $\sqrt{s} = 8$ TeV data recorded by the LHCb experiment in 2012.

The thesis is organised as follows: chapter 2 introduces the theoretical background to the events studied, starting with the mathematical structure of the Standard Model, and following with an overview of relevant phenomenology at the LHC. Chapter 3 then documents the LHC and the LHCb detector that we use for the measurement. We pay special attention to the selection of events to be studied, as the author was responsible for this within the LHCb QCD, Electroweak and Exotica Working Group. We also discuss the event reconstruction techniques at LHCb, before turning to look at the LHCb jet reconstruction algorithms and their performance in chapter 4. LHCb uses a particle-flow approach to reconstruct jets, taking as much information about the final state from the tracking sub-systems. Only the reconstruction of stable neutral particles relies on calorimeter information. Chapters 5 and 6 together outline the analysis of the Z+jet events. We introduce the detailed motivation for the analysis in chapter 5, along with the event selection. This chapter also presents studies of selected events: we determine their purity, compare them to simulated data, and determine the efficiency with which a Z boson decaying directly to two final state muons is reconstructed. Chapter 6 introduces corrections for the resolution of the jet transverse momentum (p_T), and documents a measurement of the Z+jet production cross-section (and related differential distributions and quantities) above two jet transverse momentum thresholds²: 10 and 20 GeV. These results are compared to Standard Model theoretical predictions, which are based on fixed order matrix element calculations evaluated using POWHEG [7–11] that are matched to PYTHIA [12] for simulation of the parton shower development and hadronisation. There is also a thorough discussion of the systematic uncertainties associated with the measurement. Chapter 7 examines the potential for related future measurements, and chapter 8 concludes the thesis.

Before we turn our attention to the theoretical description of the Standard Model and the phenomenology of particle collisions at the LHC, we will briefly discuss the conventions used for co-ordinates at hadron colliders within this thesis (and within particle physics in general).

²Throughout this thesis we use natural units, where $\hbar = c = 1$.

1.1 Co-ordinates at hadron colliders

It is desirable to introduce the notation used for measurements at hadron colliders and the commonly used co-ordinate system, as we will make reference to various co-ordinates throughout this thesis. The specific co-ordinate system for the LHCb experiment will be discussed in Chapter 3.

The colliding beams are often defined to travel in the $\pm z$ direction. The plane to which this direction is normal is the plane transverse to the beams, and the component of a particle's momentum in this plane is usually labelled the transverse momentum, p_T . The azimuthal angle in this plane is defined to be ϕ . Whilst the particular value of $\phi = 0$, corresponding to the x -axis, is peculiar to an experiment, differences in the value of ϕ are well defined without reference to particular Cartesian axes. With this information, a particle's momentum in the x and y directions may be found. To specify the direction a particle travels in, one more piece of information is needed - how far 'forward' the particle is travelling. This is typically measured at hadron colliders using the rapidity variable, given by,

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (1.1)$$

where E is the energy of the particle. This variable is, for massless particles, equivalent to the pseudorapidity, given by,

$$\eta \equiv \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right) \equiv -\ln(\tan(\theta/2)), \quad (1.2)$$

where $\mathbf{p}$ is the particle's 3-momentum, and θ is the angle the particle makes to the z -axis.

The rapidity variable is useful as particle production at hadron colliders is typically reasonably flat as a function of rapidity (until there is a kinematic cut-off at large $|y|$). Differences in rapidity are also invariant under longitudinal Lorentz boosts.

In addition, the η and ϕ variables provide a space which contains angular information about the direction of travel of particles, and the distance measure $\Delta R^2 = \Delta\eta^2 + \Delta\phi^2$ provides a reasonable variable to determine the separation of two particles.

The values of rapidity are often used to label certain areas of phase space. The central region refers to the region where hard particle production is favoured (for kinematic reasons), roughly $|\eta| < 1.5$. The forward region can be thought of as complementary,

occupying the region with roughly $|\eta| > 2$. Energy-flow increases in the forward region, as the proton remnants scatter through small angles in particle collisions.

Chapter 2

Theoretical overview

“One never notices what has been done; one can only see what remains to be done.”

— Marie Curie

This chapter discusses some of the theoretical concepts that are used throughout the rest of the thesis. We begin by looking back and summarising our best theory of the sub-atomic world, the Standard Model of Particle Physics, which was formulated in the latter half of the twentieth century. This is followed by more detailed discussions on the electroweak and strong interactions. We then focus on how these theories give rise to predictions for observables at proton-proton collider experiments.

2.1 The Standard Model of particle physics

The Standard Model of Particle Physics [13–16] is summarised in the following section, based on summaries presented in Refs. [17–21]. The Standard Model consists of a set of fundamental particles and the rules that govern their interactions. These rules describe the interactions of three of the four fundamental forces of nature: the electromagnetic force, the weak force, and the strong force. The mathematical formulation of these forces is based on gauge symmetries within group theory. Specifically, the Standard Model is invariant under $SU(3) \otimes SU(2)_L \otimes U(1)$ transformations. The Standard Model does not describe the gravitational interaction however; there exists no quantum theory of gravity which provides experimentally verifiable predictions. Classical gravity, being

much weaker than the three other forces, is not expected to affect processes at the LHC (or similar energy scales). Some modifications to the Standard Model to include gravity make predictions where gravitational interactions can affect LHC processes but so far there has been no experimental evidence for these extensions to the Standard Model [22]. In addition, the Standard Model contains no natural candidate for dark matter. For these reasons, the Standard Model is not a complete or universal theory of nature. Instead, some theoretical physicists advocate viewing it as an effective theory that is valid at relatively low energy scales [23].

In the Standard Model, the forces are mediated by the exchange of fundamental bosons, summarised in Table 2.1. The electromagnetic force is mediated by the photon, the weak force is mediated by the $W^\pm$ and Z bosons, and the strong force is mediated by the gluon. There is one other boson predicted by the Standard Model. The spin-0 Higgs boson is a direct consequence of the presence of the Higgs field. This field provides mass to the particles of the Standard Model whilst maintaining gauge invariance through the Electroweak symmetry breaking mechanism. The ATLAS and CMS detectors at CERN recently announced the discovery of a new particle with properties consistent with those of the Standard Model Higgs boson [1, 2].

The other particles of the Standard Model are the 12 fermions. These come in two varieties: the leptons and the quarks, each arranged in three generations. In the leptons, each generation consists of one charged lepton and one neutrino. Each quark generation consists of a positively charged up-type quark and a negatively charged down-type quark. The quarks are distinct from the other fermions in the Standard Model in that they feel the strong interaction. This will be discussed in more detail in Section 2.1.2. Every fermion in the Standard Model also has a partner anti-fermion, which has the same mass and spin, but opposite charge. The non-neutral fermions come in two further varieties: left- and right-handed chiral states. Only the left-handed chiral states interact with the $W^\pm$ boson, and the left-handed and right-handed states interact with the Z boson with different strengths. The leptons and quarks are summarised in Tables 2.2 and 2.3.

Force	Particle	Mass / GeV	Spin	Q/e
Electromagnetism	γ	0	1	0
Weak	$W^\pm$	80.385 ± 0.015	1	± 1
	Z	91.1876 ± 0.0021	1	0
Strong	g	0	1	0
Higgs	H	125.9 ± 0.4	0	0

Table 2.1: The fundamental bosons in the Standard Model, and their mass, spin, and electric charge (Q). The particle associated with the strong force, the gluon, is theoretically defined to have zero mass, but experimental measurements are consistent with a mass up to a few MeV. Likewise, the photon is defined to have zero mass and charge, and the upper bound on its mass is measured to be 1×10^{-24} GeV and the upper bound on the magnitude of its charge is measured to be $1 \times 10^{-35} e$. From Ref. [24].

Generation	Particle	Mass / MeV	Spin	Q/e
1	e	$0.510998928 \pm 0.000000011$	1/2	-1
	ν_e	-	1/2	0
2	μ	$105.6583715 \pm 0.00000035$	1/2	-1
	ν_μ	-	1/2	0
3	τ	1776.82 ± 0.16	1/2	-1
	ν_τ	-	1/2	0

Table 2.2: The flavour eigenstates of the fundamental leptons in the Standard Model, and their mass, spin, and electric charge (Q). Anti-particles are not shown, and have the same mass and spin as their particle partners, but have the opposite electric charge. Neutrinos are measured to undergo flavour oscillations as they propagate. This measurement implies that the neutrinos have non-zero masses, and that the mass eigenstates are distinct from the flavour eigenstates. These masses have not yet been measured, but are known to be less than 2 eV. There are also precision measurements of the differences in the squared masses of the neutrino mass eigenstates. The values in the table have been taken from Ref. [24].

Generation	Particle	Mass / MeV	Spin	Q/e
1	u	$2.3^{+0.7}_{-0.5}$	1/2	+2/3
	d	$4.8^{+0.5}_{-0.3}$	1/2	-1/3
2	c	1275 ± 25	1/2	+2/3
	s	95 ± 5	1/2	-1/3
3	t	$173070 \pm 520 \pm 720$	1/2	+2/3
	b	4180 ± 30	1/2	-1/3

Table 2.3: The mass eigenstates of the fundamental quarks in the Standard Model, and their mass, spin, and electric charge (Q). Anti-particles are not shown, and have the same mass and spin as their particle partners, but have the opposite electric charge. As free quarks are never observed, their masses can be defined in different ways. The mass given here is that in the $\overline{\text{MS}}$ renormalisation scheme. The top quark decays before hadronising, so a direct mass measurement is possible. The number quoted for the top quark mass is that found from direct measurements. From Ref. [24].

2.1.1 Light, matter, and nuclear decays

Quantum electrodynamics

Quantum electrodynamics (QED) is the quantum theory of light and its interaction with matter [25]. It is an Abelian gauge field theory, invariant under local $U(1)$ transformations. The charge of the theory is the electric charge.

The theory starts with the Lagrangian¹ given by

$$\mathcal{L}_{\text{QED}} = \bar{\psi}i(\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.1)$$

where the covariant derivative

$$D_\mu = \partial_\mu + ieA_\mu \quad (2.2)$$

acts on some spin 1/2 field ψ (with $\bar{\psi} = \psi^\dagger \gamma^0$), describing a charged quark or lepton of mass m , with the photon field A_μ interacting with this particle. The propagation of the photon field is described by the term

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (2.3)$$

¹The Dirac gamma matrices are represented by γ^μ , with the index μ running from 0 to 3. We also assume the Einstein summation convention.

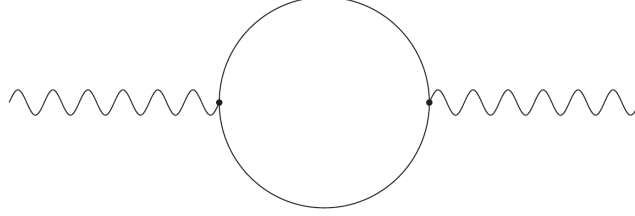


Figure 2.1: Loops such as the one shown here mean that renormalisation is needed in order to keep predictions from QED finite. Renormalisation subsumes the contributions of such loops into the definition of fundamental quantities, like the value of the electromagnetic coupling strength.

This Lagrangian is invariant under the transformations

$$\psi \rightarrow \psi e^{-ie\phi(x)}, \quad (2.4)$$

$$A^\mu \rightarrow A^\mu + \partial^\mu \phi(x). \quad (2.5)$$

The photon field must be massless, as a term representing a massive propagator ($v^2 A_\mu A^\mu$) breaks the gauge invariance.

From this Lagrangian the interactions within the theory can be calculated, using the Feynman rules to calculate the amplitudes and probabilities for various processes to occur. For the simplest processes the lowest order term in the electric charge is $\mathcal{O}(e^2)$; it is conventional to replace the charge by a coupling strength, $\alpha = e^2/4\pi$. The lowest order terms in the probability for each interaction are called leading order (LO). The next-to-leading-order (NLO) terms contain an extra power of α .

At leading order, calculations in QED give finite results. However, the theory diverges at higher orders, as loops like those shown in Figure 2.1 can be added in throughout any Feynman diagram, and the net effect of these loops is to render the calculation divergent. However, using renormalisation, the effect of these diagrams can be summed and removed from the theory by allowing quantities like the coupling strength to change with the energy of the process being studied. At higher energies the electromagnetic charge appears to increase due to this ‘running’.

Studies of QED have shown its power to describe electromagnetism at the quantum level to high levels of precision. The spin magnetic moment of an electron, $\boldsymbol{\mu}_s$, is usually parametrised by $\boldsymbol{\mu}_s = -g_s \mu_B \mathbf{S}$, where $\mathbf{S}$ is the electron spin and μ_B is the Bohr magneton,

with the parameter g_s relating the spin to the magnetic moment. The Dirac equation predicts $g_s = 2$, but higher order corrections in QED lead to a predicted value [26] of $(g_s - 2)/2 = 1\,159\,652\,181.78(0.77) \times 10^{-12}$, where the figure in parentheses indicates the theoretical uncertainty. This QED prediction agrees with the measured value [27] of $(g_s - 2)/2 = 1\,159\,652\,180.73(0.28) \times 10^{-12}$ at the level of nine parts in ten billion. QED is arguably the most precisely tested theory of nature.

The electroweak interaction

The theory of QED can be extended to account for the weak force - the force that describes nuclear decays.

We begin with the electroweak Lagrangian given by

$$\mathcal{L}_{\text{EW}} = \mathcal{L}_{\text{Bosons}} + \mathcal{L}_{\text{Fermions}} + \mathcal{L}_{\text{Higgs}}, \quad (2.6)$$

where the separate terms deal with the propagation of the vector boson fields, the interaction of the vector boson fields with the fermions (quarks and leptons), and the mechanism by which particles in the Standard Model acquire mass without breaking gauge invariance.

Consider first $\mathcal{L}_{\text{Bosons}}$, given by

$$\mathcal{L}_{\text{Bosons}} = -\frac{1}{4}W_{\mu\nu}^i W^{i\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu}, \quad (2.7)$$

$$W_{\mu\nu}^i = \partial_\nu W_\mu^i - \partial_\mu W_\nu^i - g_W \varepsilon^{ijk} W_\mu^j W_\nu^k, \quad (2.8)$$

$$B_{\mu\nu} = \partial_\nu B_\mu - \partial_\mu B_\nu, \quad (2.9)$$

where the B field behaves similarly to QED (and is invariant under $U(1)$), and the W field is invariant under $SU(2)$, with structure constants given by ε^{ijk} . The indices i, j and k run from 1 to 3.

The term $\mathcal{L}_{\text{Fermions}}$ describes how these massless fields couple to the fermions, and how the fermion fields propagate. The left-handed (Ψ_L) and right-handed (ψ_R) fermions are treated differently. Right-handed fermions are singlets, whereas left-handed fermions are

placed in doublets with the corresponding fermion of the same generation (for example up-type quarks are paired with down-type quarks). The Lagrangian is given by

$$\mathcal{L}_{\text{Fermions}} = \sum_{\text{fermions}} \bar{\Psi}_L \gamma^\mu D_\mu^L \Psi_L + \bar{\psi}_R \gamma^\mu D_\mu^R \psi_R, \quad (2.10)$$

with the covariant derivatives defined as

$$D_\mu^L = \partial_\mu + ig_W \frac{\tau_i}{2} W_\mu^i + ig_B Y B_\mu, \quad (2.11)$$

$$D_\mu^R = \partial_\mu + ig_B Y B_\mu. \quad (2.12)$$

This Lagrangian is invariant under $U(1) \otimes SU(2)_L$ transformations. The structure of the Lagrangian means that left-handed fermions interact with both the W and B fields whereas the right-handed fermions only interact with the B field. Explicit fermionic mass terms are forbidden through gauge symmetries. The matrices τ_i are the generators of $SU(2)$, of which the Pauli spin matrices are one representation, and the coupling strength of the fermions to the W fields is g_W . The B field couples with strength g_B to particles carrying weak hypercharge Y . The four vector fields (W^1, W^2, W^3, B) are all massless.

The final part of the electroweak Lagrangian enables the fermions and bosons to acquire mass, whilst maintaining the overall gauge invariance under $U(1) \otimes SU(2)$ transformations, by introducing a complex scalar (Higgs) field, Φ (an $SU(2)$ doublet with $\tilde{\Phi} = -i(\Phi^\dagger \tau_2)^T$), which breaks the electroweak symmetry by acquiring a non-zero vacuum expectation value [28–33]. This vacuum expectation value enables the fermions to acquire mass through Yukawa terms, and, through the coupling of the bosons to the scalar field, the bosons also acquire mass. Perturbations around this vacuum expectation value couple to both bosons and fermions, and are seen as a spin-0 particle which couples to mass, the Higgs boson. A particle with similar properties has recently been observed at the LHC [1, 2]. The Lagrangian for these interactions, assuming massless neutrinos, is given by

$$\mathcal{L}_{\text{Higgs}} = \mathcal{L}_{\text{HB}} + \mathcal{L}_{\text{HL}} + \mathcal{L}_{\text{HQ}}, \quad (2.13)$$

$$\mathcal{L}_{\text{HB}} = (D_\mu^L \Phi)^\dagger (D^{L\mu} \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (2.14)$$

$$\mathcal{L}_{\text{HL}} = -g_N \bar{\Psi}_{NL} \psi_{NeR} \Phi + \text{h.c.}, \quad (2.15)$$

$$\mathcal{L}_{\text{HQ}} = -\bar{\Psi}'_{ML} G_{MN} \psi_{Nd'R} \Phi - \bar{\Psi}'_{ML} \tilde{G}_{MN} \psi_{Nu'R} \tilde{\Phi} + \text{h.c.}. \quad (2.16)$$

In the terms $\mathcal{L}_{\text{HL}}$ and $\mathcal{L}_{\text{HQ}}$ the hermitian conjugate is taken, in order to ensure that the Lagrangian is hermitian. These terms also include sums over up-type and down-type quarks, and electron-type fermions. The couplings here are proportional to the fermion masses, and must be set individually through the coefficients G_{MN} and g_N . These parameters account for the majority of the Standard Model parameters. The W^1 and W^2 bosons acquire the same mass. These fields can be combined to form the physical W^+ and W^- bosons, which couple only to left-handed particles which carry weak isospin. There are also two electrically neutral spin-1 boson fields, the W^3 and the B . These mix to form new combinations,

$$Z^\mu = \cos(\theta_W) W^{3\mu} - \sin(\theta_W) B^\mu, \quad (2.17)$$

$$A^\mu = \sin(\theta_W) W^{3\mu} + \cos(\theta_W) B^\mu, \quad (2.18)$$

where one combination, the Z field, acquires mass (and the consequences of this, including a longitudinal polarisation), whereas the other, the A field, remains massless and corresponds to the photon. The amount of mixing is defined by the Weinberg (or weak) mixing angle, θ_W [14], an experimental parameter that must be measured (or derived from measurements of other Standard Model parameters). Like the underlying fields, the Z boson couples with different strengths to left-handed and right-handed particles. The photon couples equally to left-handed and right-handed particles, and maintains a residual $U(1)$ local symmetry.

There is one more subtlety we must consider before we continue our discussion of the Standard Model - the reason that the quark fields are primed above in equation (2.16). There is no reason why the eigenstates of the fermions when they interact with the weak force should be the same as the mass eigenstates of the fermions; there is no reason that a top quark must decay to a bottom quark and a W^+ boson, and not, for example, to a strange quark and a W^+ boson. Allowing for this possibility in the Standard Model introduces the CKM matrix² [34, 35]. This unitary matrix allows for mixing between the

²If neutrino masses are included, then a similar matrix, the PMNS matrix, is necessary.

flavour and mass eigenstates across the generations and is described by three rotation angles and one complex phase. The presence of this phase allows for $\mathcal{CP}$ violation. Experimentally it has been seen that the diagonal elements of the CKM matrix are close to unity; a top quark is more likely to decay to a bottom quark than a down quark or a strange quark. A primary goal of the LHCb experiment [36] is to perform precision measurements of the CKM matrix, to better understand differences between matter and anti-matter.

2.1.2 The strong interaction

Our most fundamental theory that describes the strong interaction is quantum chromodynamics (QCD). Like QED, the theory describes the interactions of matter by means of particle exchange. In the case of QCD, particles with colour charge (like quarks) interact with each other by gluon exchange. Unlike QED, the gluons themselves carry colour charge, and so also interact with each other by emitting and absorbing gluons. There are three different types of colour charge, conventionally labelled red, green and blue. Antiparticles carry the anti-colour. Quarks are associated with a single colour; gluons are made up of a colour and an anti-colour (or a superposition of such pairs). The presence of the colour quantum number means that to first order (neglecting differences in couplings and masses) a particle is three times more likely to couple to a particular quark than a particular lepton.

QCD is a gauge field theory, with the theory invariant under local non-Abelian $SU(3)$ transformations. The Lagrangian of the theory is given by

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_a(i\gamma^\mu D_\mu^{ab} - m\delta^{ab})\psi_b - \frac{1}{4}F_{\mu\nu}^A F_A^{\mu\nu}, \quad (2.19)$$

with the covariant derivative given by

$$D_\mu^{ab} = \delta^{ab}\partial_\mu + ig_s G_\mu^A T_A^{ab}, \quad (2.20)$$

and the gluon field G_μ^A propagating term being given by

$$F_{\mu\nu}^A = \partial_\mu G_\nu^A - \partial_\nu G_\mu^A - g_s f_{ABC} G_\mu^B G_\nu^C. \quad (2.21)$$

The extra term in this tensor in comparison with QED highlights the non-Abelian nature of QCD. This Lagrangian is symmetric under $SU(3)$ gauge transformations.

The field ψ describes the quarks in the theory, carrying a colour charge, with the quantum number of the colour charge given by a and b . As there are three colours ($N_C = 3$), these indices run from 1 to 3. The coupling strength of the theory is g_s , and the structure constants are f_{ABC} , where the indices A, B, C run from 1 to $N_C^2 - 1 = 8$. The generators of the gauge group are T_A , and can be represented as eight 3×3 matrices (the Gell-Mann matrices). Like QED, the strong interaction is associated with a coupling strength, $\alpha_S = g_s^2/4\pi$.

QCD, like QED, contains divergences when particle exchanges are evaluated beyond leading order. As with QED, these divergences can be removed by renormalising the theory. The coupling strength of the strong interaction varies with the energy of the process being considered, Q , as

$$\alpha_S(Q^2) = \frac{\alpha_S(Q_0^2)}{1 + \beta\alpha_S(Q_0^2) \ln(\frac{Q^2}{Q_0^2})}, \quad (2.22)$$

$$\beta = \frac{11N_C - 2N_f}{12\pi}, \quad (2.23)$$

so for 3 colours and 6 quark ($N_f = 6$) flavours, as in the Standard Model, $\beta > 0$ and the strong interaction runs to become weaker at higher energies.

At low energies, α_S is too large for a perturbative approach to be valid. It is in this realm that we must turn to soft (or non-perturbative) QCD. The large value of this coupling strength in low energy QCD means that particles we measure in nature only exist in colour charge neutral ('colourless') states. This is known as confinement. We measure particles that are made up of colour-anti-colour pairs (mesons) or bound states of the three different (anti)colours (baryons). The different states of matter that contain particles which feel the strong force are labelled collectively as hadrons.

At higher energies, the value of α_S is sufficiently small ($\alpha_S(M_Z) = 0.1184 \pm 0.0007$ [24]) to allow the use of perturbative approaches. In this regime, we make predictions using perturbative QCD (pQCD). A key point of high energy QCD is that quarks that are bound in baryonic states may be considered as free particles, as their coupling to each

other is sufficiently small [37, 38]. The demonstration of this ‘asymptotic freedom’ won the Nobel prize in 2004.

2.2 Theoretical predictions of Standard Model processes

Putting together the different pieces of the Standard Model allows for the construction of a gauge field theory that is invariant under $SU(3) \otimes SU(2)_L \otimes U(1)$ transformations. The different terms contained within the Lagrangian give rise to different interactions, allowing the probability of any fundamental process to be written down using perturbative approaches. As discussed previously, a consequence of considering theoretical predictions beyond leading order is that the coupling strengths of the different interactions vary with the energies being probed, as the theories must be renormalised to provide meaningful predictions. This running means that for the strong interaction, the perturbative approach only becomes valid at high energies. In some cases, there seems to be an obvious choice for the renormalisation scale (for example, for interactions on the Z pole, it can make sense to set the renormalisation scale to the Z mass). However, the theory does not provide a correct scale to use, and the use of any scale in calculations is technically valid.

If a calculation were possible at all orders in the perturbative expansion, then the calculation would show no dependence on the renormalisation scale; the effects of the renormalisation scale can be viewed as effects of truncating the full perturbative expansion. In most cases, changing the renormalisation scale has a smaller effect in calculations at higher orders in the perturbative expansion than at lower orders. Since it is known that the calculation will show no dependence on the renormalisation scale were a complete calculation to be made, then the change in prediction when the renormalisation scale is varied provides information about the size of effects associated with higher order terms. This can be used to set an uncertainty from missing higher order terms in any theoretical prediction calculated using perturbation theory. The effect of varying scales is usually only studied in pQCD; perturbative series in QED tend to converge sufficiently quickly that the effects of changing the renormalisation scale can typically be neglected.

However, there are other noticeable effects which arise from higher orders in QED. Consider the Feynman diagram shown in Figure 2.2a. The final state particles, the muons, carry electrical charge, and can radiate photons, as shown in Figure 2.2b. These photons are real (as opposed to virtual, off-shell) particles, and the effect of their presence

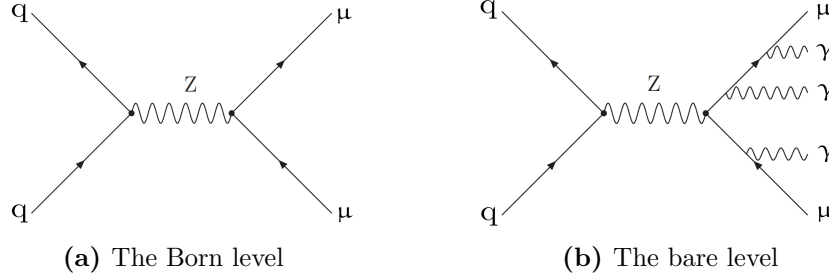


Figure 2.2: Feynman diagrams at the Born level in QED do not have photon radiation from the final state particles, as shown on the left, whereas at the bare level (the level at which experiments reconstruct particles) the final state particles have emitted radiation. An example of such radiation is shown on the right.

can be directly measured. This causes a difference between results defined at the Born level (before such radiation) and the bare level (after the radiation).

2.2.1 The factorisation theorem

We can calculate the cross-section for some hard process, for example $q\bar{q} \rightarrow Z \rightarrow \mu^+\mu^-$, using perturbative methods. However, the experiments at the LHC and the Tevatron collide hadrons, not partons. We therefore need to find some method for getting predictions for the proton collisions, rather than simply predictions for the parton collisions.

Here we make use of the factorisation theorem [20] and asymptotic freedom. We can essentially calculate the cross-section for each possible initial state, and then weight this cross-section by the probability that the protons contain initial-state partons of the required energy. These probabilities are parton distribution functions (PDFs). This approach is legitimate providing that the interactions between different partons in the same proton are small, which is the case at the high energies where asymptotic freedom applies. We can express the factorisation theorem mathematically as [20]:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu^2) f_j(x_2, \mu^2) \sigma_{ij}(p_1, p_2, \alpha_s(\mu^2), Q^2, \mu^2) \quad (2.24)$$

where $\sigma(P_1, P_2)$ is the cross-section for two hadrons with momenta P_1 and P_2 , and where we sum over all possible initial-state parton contributions, i and j . These partons have

momenta $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$, so the integrals are over all possible parton momenta. $\sigma_{ij}(p_1, p_2, \alpha_s(\mu^2), Q^2, \mu^2)$ is the cross-section for the interaction of these partons, and in pQCD depends upon the renormalisation and factorisations scales (which for simplicity here are set equal to each other, μ^2). The parton distribution functions, f , are dependent upon the scale, μ^2 , and the momentum fraction of the partons.

The factorisation scale is the scale which separates long and short range physics. Long range physics (low momentum parton emission) can be described by the structure of the colliding hadrons, and can therefore be absorbed into the definition of the proton through the parton distribution functions. Higher energy parton emission is considered to be part of the hard process. The actual cross-section we measure in nature does not depend upon this scale. However, we will find different predictions for cross-sections by varying this scale. Choosing a realistic factorisation scale is therefore an important element of finding realistic theoretical predictions. It is often set to either the vector boson mass (for inclusive vector boson production) [7], or the transverse momentum of the vector boson in the event (for inclusive vector boson + jet production) [11]. When higher orders terms are considered within a calculation the dependence of the calculated cross-section upon the factorisation scale diminishes, similarly to the way the dependence on the renormalisation scale reduces.

We shall discuss other elements of QCD before returning to the crucial role that PDFs play in hadron collider physics.

2.2.2 Determining the final state

In addition to the primary hard process that we study we must also consider the effect of a potentially large number of additional partons also being emitted in a proton-proton collision. For high energy partons (within pQCD) this effect is typically small, but at lower energies and momenta (where α_s becomes larger) the effects of soft (or collinear) parton emission can become important. This can be treated in different ways.

In beam-beam collisions, at leading order, vector bosons produced by parton collisions should have zero transverse momentum (neglecting the intrinsic transverse momentum of the colliding partons). Higher order diagrams include the emission of partons which can generate transverse momentum, and include terms that are of order $\alpha_s^m \ln^m(p_T^2/M^2)$ for all $m > 0$ [39–41], where p_T is the transverse momentum of the vector boson and M is the relevant energy scale, typically the vector boson mass. In the limit that

$p_T \rightarrow 0$, these terms are divergent (and are therefore no longer small). Therefore fixed order predictions from pQCD are known to break down when predicting the transverse momentum distributions of vector bosons at low p_T . However, in the limit of soft gluon emission the effect of these large logarithmic terms can be summed at all orders and included in the calculations [39–41]. This can be used to modify results from fixed order calculations so that the vector boson p_T distribution (and related variables) can be modelled more accurately. We will briefly return to this approach when we consider the computational programs available for calculating predictions at the end of this chapter.

Instead of resumming the effect of soft gluon radiation, which is known for only a small subset of processes, the effect of large higher order terms can be handled through the use of a parton shower [12, 42, 43]. Parton showers provide predictions for the complete final state. The result of a calculation of some hard process is found, and then the parton shower is responsible for simulating the effect of higher orders in pQCD and the effect of the evolution of other objects in the final state (for example, remnants of protons in pp collisions). Missing terms in the perturbation expansion are accounted for by allowing particles to radiate additional partons and photons. The showering is carried out in a strict ordering (a necessity of the calculation) until no final state particles have energy above some threshold (PYTHIA [12, 42] and HERWIG++ [43] both apply a 1 GeV threshold). The evolution variable used for the ordering varies between different programs. PYTHIA [12, 42] considers p_T ordering, where the hardest partons are emitted first whereas HERWIG++ [43] considers angular ordering. Parton showers are only strictly accurate in the limit that the parton splitting is soft or collinear, and the contribution of splitting at large angles is therefore underestimated.

As parton showers simulate the effect of higher orders in pQCD, parton showers and NLO calculations overlap in what they simulate. Historically, this has led to problems interfacing calculations beyond LO with parton showers and avoiding the double counting of various higher order contributions. Recent advances in the study of parton showers has enabled this double counting to be subtracted when the predictions of hard processes are found at NLO [8, 9, 44]. This means that predictions for hard processes can be made at NLO using one program, and then interfaced with other programs to simulate the parton shower and subsequent clustering and hadronisation.

In the processes we have discussed so far we have considered final states which include partons. However, the partons are not observed in the final states that we see in physical events. At lower energies, the coupling strength of the strong force increases. Since the partons are coloured states, the force between the partons increases, and the partons

hadronise into colourless bound states (of hadrons). There are no clear theories of how this process works as it is a low energy process and cannot be predicted by pQCD. One approach is the Lund String Model [45]. In this model the energy associated with the separated partons becomes sufficiently large for new partons to be created out of the vacuum. This process continues until some cut-off energy scale is reached. At this stage bound colourless hadrons are formed out of the particles that have been produced, using measured fragmentation functions. Other models take a different approach. The cluster model [46] works by splitting all the final state gluons into quarks and antiquarks. Neighbouring quarks and antiquarks are then grouped into colourless clusters. The mass and quark content of these clusters determine the hadrons formed, through the use of fragmentation functions and hadronisation fractions.

The unstable final state hadrons then decay through various Standard Model processes into more stable particles. This process is quantum mechanical in nature, and is therefore fundamentally probabilistic. It is typically modelled using measured branching fractions.

In some events there may be no single hard process: there may be additional interactions between other partons in the initial state, as the proton remnants are coloured objects and will interact, primarily through the strong interaction. If multiple initial-state partons in the same proton play a significant role in the event then we tend to label the event as containing multiple interactions. If, on the other hand, the final state is simply populated with additional soft QCD radiation and associated hadrons then we classify this as being part of some underlying event that takes place in addition to the hard interaction. This separation between underlying event processes and multiple interactions is arbitrary, and is simply a way of labelling different parts of the event. Ultimately, as we cannot observe the free partons, we deal with models tuned to other experimental data. These models give rise to empirical predictions that match what we observe, hopefully to a high level of precision.

2.2.3 Jets

The final state partons shower, hadronise and decay to produce a set of particles with the same momenta as the partons; the particles observed in the final state of a proton-proton collision are the long-lived stable particles following such processes. Whilst individual particles often contain little direct information about the hard event that took place, if a sufficiently hard parton is emitted in the laboratory frame of reference, then the particles it decays to will tend to form a collimated beam, with higher momentum partons tending

to produce more collimated beams. We therefore introduce jet clustering algorithms for the final state that produce experimental observables which contain significant and useful information about the partons emitted in the hard process and the QCD processes they undergo in hadronising [47]. Ultimately, the purpose of defining jets is to reduce the complexity of the final state objects whilst retaining the physics information of interest.

There are some fundamental requirements for well-behaved jet clustering algorithms at hadron colliders. These were produced by a collaboration of experimental and theoretical particle physicists, in a statement known as the Snowmass Accord [48]. These requirements are that:

- the algorithms should be easy to implement theoretically and experimentally.
- the algorithms should be defined at all orders of pQCD, and should give finite results at all orders of pQCD. This is referred to as infrared safety. The most important concerns here are that the emission of extra soft particles or collinear splitting of partons should have no effect on the jets clustered.
- the algorithms should be applicable before and after hadronisation is considered, and should be as insensitive to hadronisation effects as is possible.

These are achieved through the choice of suitable clustering definitions.

The most commonly used jet algorithms are those in the k_T family [49–52], which are sequential recombination algorithms. Historically these are computationally intensive, but modern techniques such as those implemented in FASTJET [53] have enabled a fast implementation. These use the following algorithm:

1. We define some distance measure, d_{ij} , between any two particles i and j . We also calculate the distance between each particle, i , and the beam. We label this d_{iB} .
2. We take the minimum distance. If this corresponds to a pair of particles, they are combined into one overall particle. This is typically done by summing their 4-momenta, and using this quantity to define a new particle. We then replace the merged particle with this new particle. If the minimum distance is between a particle and the beam, then we remove this particle from the list of particles and consider it to be a jet.
3. We then repeat the process on the new list of particles.

This process provides a list of jets in the event.

For the k_T family [49–52] the distance measures are given by

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad (2.25)$$

$$d_{iB} = p_{T,i}^{2p}, \quad (2.26)$$

where $p_{T,i}$ is the transverse momentum of particle i and ΔR_{ij} is the relative distance in $\eta - \phi$ space between the particles i and j , given by $\Delta R_{ij}^2 = \Delta \eta_{ij}^2 + \Delta \phi_{ij}^2$. The parameter R is the “cone size”. As we shall see, two particles inside this cone will tend to be clustered into the same jet. When the parameter $p = 1$, we have the k_T algorithm. For $p = 0$, we have the Cambridge-Aachen algorithm. For $p = -1$, we have the anti- k_T algorithm.

We show how the anti- k_T [52] algorithm functions by considering a few simple cases:

- If there are any two particles in the range $\Delta R_{12} < R$ then they will be clustered together to form a jet.
- The d_{ij} between a hard particle and a soft particle is determined by the momentum of the hard particle. If there are no other hard particles present within a distance $2R$ of the hard particle, then all soft particles within a radius R of the hard particle will be clustered with it to form a conical jet with radius R .
- If there is another hard particle in the range $R < \Delta R_{12} < 2R$ from the hard particle then there will be two jets present. If one of the particles is much harder, then it will form a conical jet, and the softer particle forms a non-conical jet. If the two hard particles have similar transverse momenta then neither jet will be conical.

If there is collinear splitting of a parton, k , into particles i and j , then the ΔR_{ij} value after this splitting will be 0, and the particle we form by recombining i and j is the same as the particle present before the collinear splitting, k . Collinear splitting of partons therefore has no effect on the clustered jets. In addition, we can consider the emission of a soft particle from a hard parton with non-zero angular separation. Such emissions do not modify the kinematic properties of the jet. These properties together make the algorithm infrared safe. As such it can be considered to be well defined at all orders in pQCD.

Jet shapes are defined by adding many soft particles to the event throughout $\eta - \phi$ space. By considering which particles are clustered into which jets, we can set up boundaries for the jets in $\eta - \phi$ space, giving a shape to the jets. The process of adding these soft partons to the event does not change the shape or kinematics of the jets in the event as the algorithm is infrared safe. The shape of jets reconstructed using different

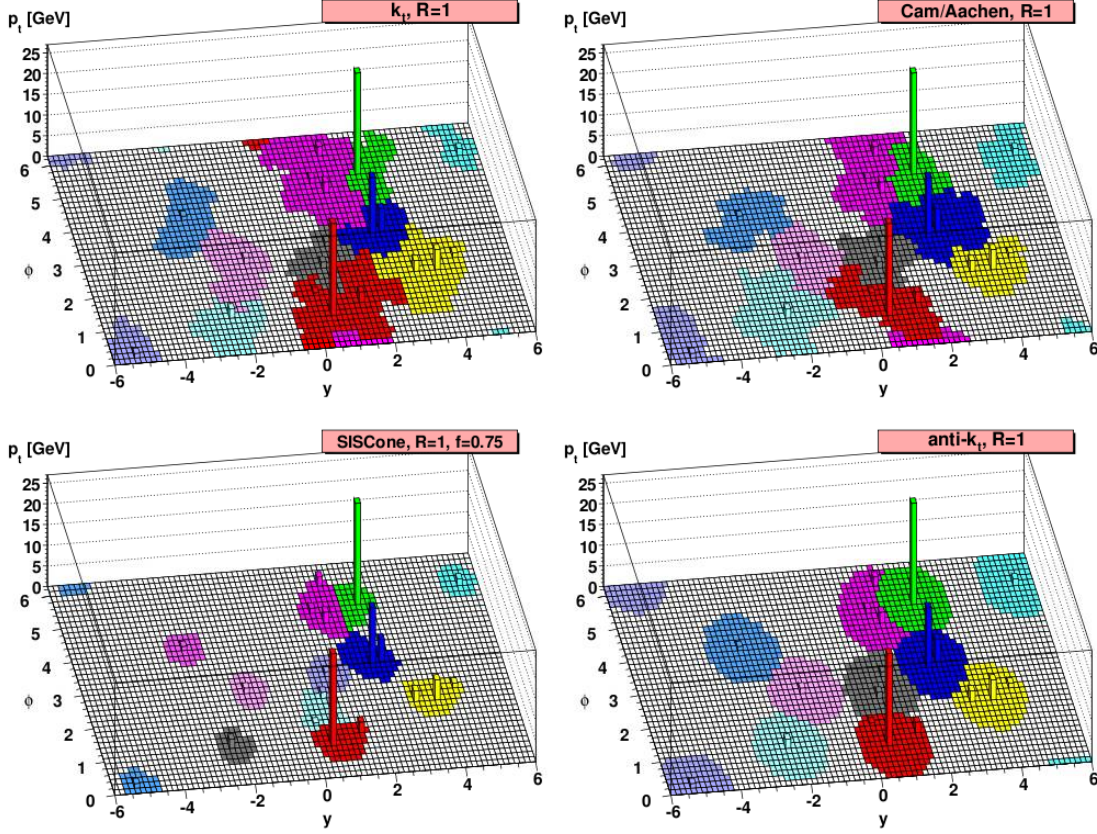


Figure 2.3: The shape of jets in the same event that are clustered using different algorithms. Taken from Ref. [52].

algorithms is compared in Figure 2.3. One of the nice features of the anti- k_T algorithm is that it tends to give more conical jets than the other algorithms in the k_T family [52].

We must also consider the differences between jets reconstructed at different stages of a proton-proton collision. In particular, we want to compare jets reconstructed at the parton level, created out of final state partons produced in a matrix element level calculation and any subsequent parton shower, to the jets we would reconstruct after considering hadronisation and fragmentation. This stage corresponds to the jets defined at the hadron level, and the jets correspond to those we would reconstruct in a perfect detector. There are different effects [52] that give rise to differences between jets reconstructed at the hadron level with respect to those reconstructed at the parton level:

- Underlying event: The underlying event will cause particle production through all $\eta - \phi$ space. When included within the clustering, this has the net effect of increasing the momentum of jets. The effect is smaller for smaller values of R .

- **Hadronisation:** As hadronisation occurs, the final state partons will cluster into particles which then undergo subsequent decays. This may well lead to particles formed from partons inside a jet at parton level falling outside a jet at hadron level. This reduces the momentum of jets at the hadron level with respect to those at the parton level. The effect decreases for larger cone sizes.

These processes will also generate mass for jets - even for jets which at the parton level consist of one massless parton³.

The size of these effects have been modelled and studied [54] and found to be greater for gluon-initiated jets than quark-initiated jets. In essence, as they carry more colour charge, the amount of radiation in gluon-initiated jets tends to be greater than the amount in quark-initiated jets. The sum of the squares of the typical sizes of these processes squared is shown in Figure 2.4 as a function of the radius parameter used. These effects also depend on the momentum and type of emitted parton. A broad minimum is seen in the region between $R = 0.5$ and $R = 0.7$.

2.2.4 Parton distribution functions

Parton distribution functions are, crudely, a measure of the probability of a certain parton having a given energy, and represent the modulus squared of the wavefunction of the inner structure of the proton. They are found as a function of the energy scale being considered, Q^2 , and the Bjorken- x value [56]. This variable was first introduced in deep inelastic scattering measurements, and is given by $x = Q^2/2p \cdot q$, where p is the parton's 4-momentum and q is the 4-momentum transfer in the interaction being considered. In the high energy limit, x can be viewed as the fraction of the proton's momentum carried by the parton [17]. Indeed, the x variable was already introduced in equation (2.24).

A naïve view of the proton treats it as simply being composed of two up quarks and one down quark. This approach is surprisingly powerful for calculating, for example, the magnetic moment of the proton, but it is not a complete picture. Due to gluon exchange inside the proton, quark and anti-quark pairs or gluons are constantly being created and annihilated. Consequently, the proton contains other contributions than those suggested by the simple picture. However, we can split these sources up, leaving

³Light quarks are often modelled as massless, as their mass is much lower than the energy scales being probed.

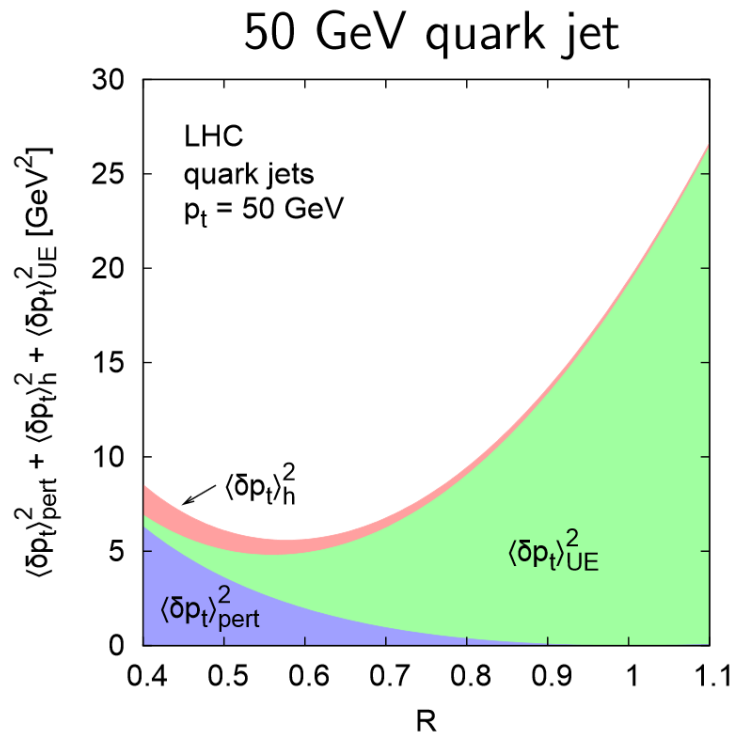


Figure 2.4: The sum of the squares of the expected differences in jet p_T between the parton and hadron levels, due to the effects of the underlying event and hadronisation. Also shown are effects related to processes described by higher-orders in pQCD. These processes cause differences in the jet p_T between jets formed from final state particles found in fixed order pQCD predictions and jets formed after any parton shower is considered. Taken from Ref. [55], in the style of Ref. [54]

valent contributions to the up and down distribution functions (u_v and d_v), such that

$$\int_{x=0}^{x=1} u_v(x) dx = 2, \quad (2.27)$$

$$\int_{x=0}^{x=1} d_v(x) dx = 1, \quad (2.28)$$

so that the proton contains two up valence quarks and one down valence quark. Then the sea contributions, from particle and anti-particle creation, can be written as

$$\bar{u}_s = u_s, \quad (2.29)$$

$$\bar{d}_s = d_s. \quad (2.30)$$

It is reasonable to make the approximation that $u_s = d_s$ as the difference in mass between up and down quarks is small, and the production of sea partons is dominated by QCD. There are also other sea contributions to the PDFs from the strange and charm partons. It is also normal to include a contribution to PDFs from bottom quarks, though four flavour PDFs are calculated neglecting this contribution (so all bottom particles are produced in the final state) [57]. There is also a gluon PDF. Indeed, the majority of proton-proton interactions at the Large Hadron Collider are due to gluon-gluon interactions, and over half the proton's momentum is carried by gluon constituents at LHC energy scales. Recent PDFs have considered the possibility of including a photon contribution, though this is not yet a standard approach [58]. Typical values of PDFs are shown in Figure 2.5.

The different PDFs are also constrained by the total momentum carried by the proton. This condition - a momentum sum rule - can be written as

$$\sum_i \int_{x=0}^{x=1} x f_i(x) dx = 1, \quad (2.31)$$

where the parton distribution associated with a species of particle, i , is given by $f_i(x)$. Other sum rules can also be considered, which constrain other properties of the proton [59].

The factorisation scale, discussed earlier, determines the energy at which emission is considered to occur within the hard interaction or within the proton. This split is, of course, completely arbitrary, and is not due to any physics, but due to computational

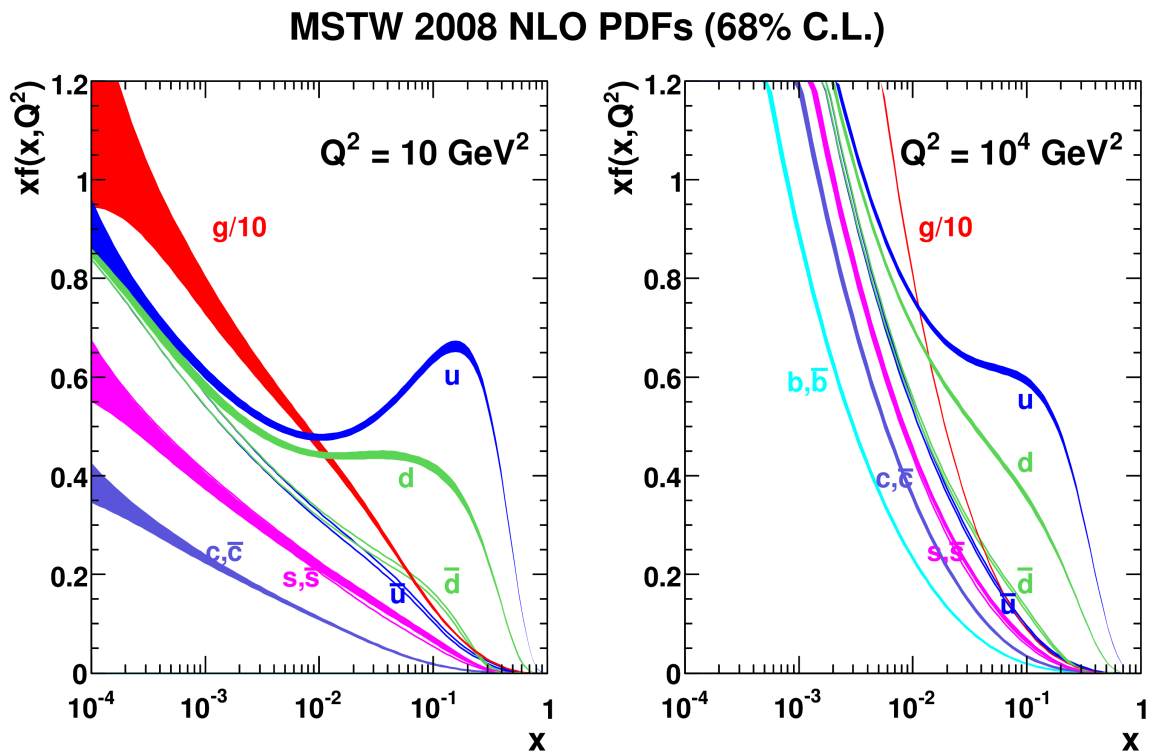


Figure 2.5: PDFs at different values of x and Q^2 . From Ref. [57]

approaches. Therefore, the complete theoretical prediction at all orders in pQCD will be independent of the factorisation scale choice. This fact enables the PDF at one energy scale to be ‘evolved’ to find the PDF at another, different energy scale, through the DGLAP equations [20, 60–62]. PDFs can be found at different orders in pQCD, to match onto hard process calculations at different orders. The orders differ by the number of splittings they consider for the different partons inside the proton.

PDFs can be calculated directly from lattice QCD. However, this approach currently yields large uncertainties in the distributions. PDFs used to make predictions of cross-sections at the LHC are therefore calculated by performing global fits to data taken at fixed target experiments, and to collider data recorded at HERA and the Tevatron. These fits are performed by different groups: MSTW [57] (originally MRST [63]), CTEQ [64], NNPDF [65] and others. However, the distributions are not particularly well constrained for small values of x . The LHCb experiment will play a key role in making measurements that can be used to probe, understand and potentially constrain the PDFs in this region [66]. This will be discussed in more detail later in this thesis (see Chapter 5). The study of parton distribution functions and the inner structure of the proton is a rich area of research within QCD and one where data taken at the LHC has an important role to play.

2.3 Computational programs for predictions at hadron colliders

As has been demonstrated previously, there are many different approaches to making predictions based on the Standard Model. A discussion of the main computational programs available follows. All programs discussed here make use of Monte Carlo methods, relying on random number generation to compute the various integrals and probabilities yielded by analytic calculations. The different programs take different approaches to simulate the various stages of events at hadron colliders discussed above.

There are standard programs for calculating cross-sections for processes involving electroweak boson production at hadron colliders, FEWZ [67] and MCFM [68]. These produce predictions at the parton level, which are generally accurate for angular distributions like rapidity, but do not provide accurate predictions at low transverse momenta [41]. These predictions are now available at NNLO for vector boson production. Programs

like RESBOS [39, 41, 69, 70] allow for resummation of soft gluon emission at the level of next-to-next-to-leading logarithms (NNLL), providing corrections to NLO distributions for vector boson production. These predictions tend to provide more accurate predictions at low transverse momentum, but are only available for a few processes, and the predictions are inclusive in the number of jets.

Programs like PYTHIA⁴ [12, 42] and HERWIG++⁵ [43] provide LO predictions of the hard process being simulated. They then run parton showers and clustering models to provide a complete event simulation. The evolution variables and clustering approaches differ between the two programs, so a comparison is expected to provide a reasonable estimate of the differences between theoretical approaches. Other programs, like ALPGEN [71, 72] and SHERPA [73] also provide similar predictions. In addition, there are programs like PHOTOS [74], which are dedicated to the study of QED FSR.

Relatively recently it has become possible to interface higher order predictions of a hard process with a parton shower program without double counting. For example, the POWHEG [7–10] generator enables parton level predictions of a hard process at NLO to be interfaced with either PYTHIA [12, 42] or HERWIG [75, 76] for subsequent modelling of a parton shower, clustering and hadronisation. This has many advantages. The calculation of the hard process at NLO provides more accuracy in predictions of the overall cross-section and angular variables than LO calculations, whereas the interface with a parton shower enables a complete event simulation, the construction of hadron level quantities, and accurate predictions at low transverse momenta.

There is one final stage that is often associated with Monte Carlo simulation of events. It is often useful to apply a simulation of a detector experiment to the final state generated by a complete event generator. Such studies can be used to determine the potential for a detector to make particular measurements, to compare data taken at an experiment directly with theory without unfolding detector effects, or to analyse the detector performance. These simulations are often performed using sophisticated models of the detector geometry and response, and include simulation of secondary particle interactions with the machinery. This is often implemented using GEANT4 [77, 78].

The different Monte Carlo generators have various parameters that can be set to govern the different parts of the simulation (for example, different parameters control

⁴There are two versions of PYTHIA: PYTHIA 6 and PYTHIA 8. PYTHIA 8 is an update of the original fortran version to C++, and is the currently maintained code.

⁵HERWIG++ is a C++ implementation of the original HERWIG code. However it has been extended beyond HERWIG as it is now the maintained code from HERWIG collaborators.

the implementation of the parton shower and hadronisation models). These parameters are tuned to match data, but can be varied, and such variations can be used to find the sensitivity of the computation to different parts of the simulation. For predictions at the LHC, the Monte Carlo generators have been tuned on measurements at previous experiments such as LEP. Different tunings are often studied in measurements of the underlying event.

Chapter 3

The LHCb experiment at the LHC

“The test of all knowledge is experiment. Experiment is the sole judge of scientific truth.”

— Richard Feynman

In this chapter we discuss the Large Hadron Collider Beauty (LHCb) experiment at the Large Hadron Collider (LHC), and the computing and software environments used by the experiment. The LHC machine is the highest energy collider yet constructed, colliding proton beams at multi-TeV energy scales. The LHCb detector is a fully instrumented forward spectrometer primarily designed to search for physics beyond the standard model in the decays of particles containing b quarks [36]. These measurements will mainly search for an enhanced rate of processes that are very rare within the Standard Model, or for $\mathcal{CP}$ violation at levels that are inconsistent with other measurements of Standard Model parameters. Many of the details of the LHC machine and the LHCb detector are taken from Refs. [79] and [80]. This chapter also provides detail on the LHCb trigger and the LHCb offline data selection, as the author has worked on these extensively.

3.1 The LHC

The LHC machine [79, 81] is located at the European Organisation for Nuclear Research, CERN, just outside Geneva (see Figure 3.1), and is a 27 kilometre machine that acts as a particle accelerator and collider. The machine crosses the border between France and Switzerland in the circular tunnel that was originally constructed for the Large Electron

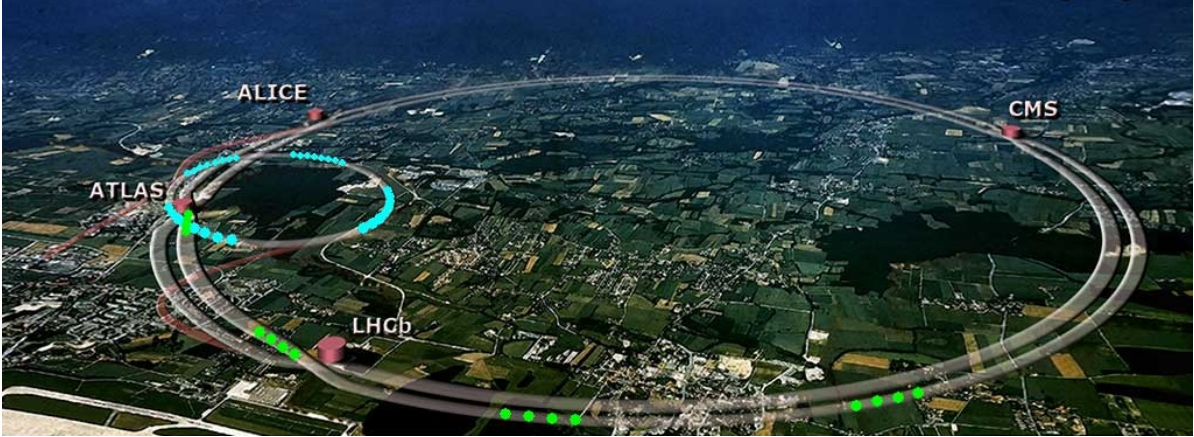


Figure 3.1: The CERN site. Source: CERN

Positron collider (LEP). The LHC is a synchrotron accelerator, and is used to accelerate and collide both protons and lead ions. Within this thesis we only consider the machine when colliding protons.

The LHC proton beams begin as hydrogen stored in gas bottles. The hydrogen from these bottles is ionised, and the protons are extracted. The protons are then accelerated through a series of particle accelerators (as shown in Figure 3.2), many of which were also responsible for important physics discoveries in the 20th century. The first acceleration stage occurs at the LINAC 2 machine [82, 83]. This delivers the protons to the Proton Synchrotron booster at an energy of 450 MeV, which in turn feeds the Proton Synchrotron (PS) with proton beams of 1.4 GeV. The PS accelerates the protons to 25 GeV, develops the LHC bunch train, and delivers the protons to the Super Proton Synchrotron (SPS). This accelerates the protons to 450 GeV, and creates the beam structure required for the LHC. This entire process takes around sixteen minutes [79].

The LHC then accelerates the proton beams over a period of about twenty minutes. Eight radio-frequency cavities are used by the LHC to accelerate each beam to its final energy¹ (3.5 or 4 TeV). This allows proton collisions with centre-of-mass energies of 7 or 8 TeV. The LHC beams consist of bunches of protons, designed to be separated by a 25 ns time interval. However most runs to date have collided beams with a bunch spacing of 50 ns. Each bunch contains roughly 1.15×10^{11} protons, and the design allows for 2808 bunches per beam. The LHC could contain more bunches per beam, up to a maximum of approximately 3600, but some bunches are left empty to facilitate the

¹The machine is currently being prepared to deliver beams with energies up to 7 TeV.

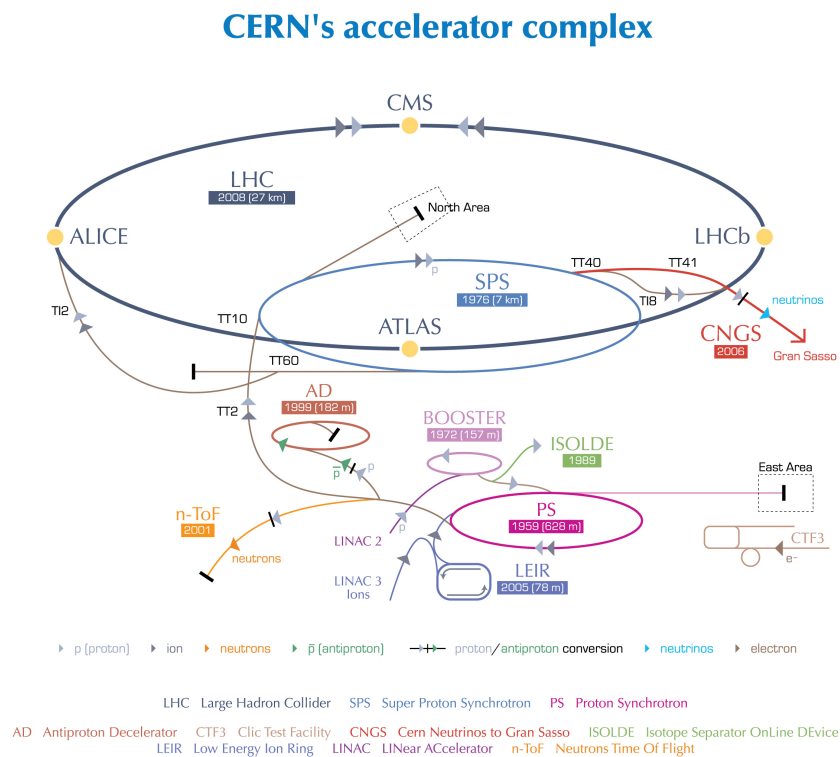


Figure 3.2: The CERN accelerator complex. The different accelerators which now accelerate protons before injection to the LHC themselves played an important role in discoveries in the latter half of the 20th century. Source: CERN

transfer of the beam during the injection stages. With the 50 ns bunch spacing each LHC beam in 2011 and 2012 typically contained around 1400 bunches.

The beam is controlled and accelerated using 1232 dipole magnets, which each produce a field of 8.4 T. The magnets are cooled to 1.9 K in order to produce the high magnetic field needed. This cooling is managed using super-fluid helium.

The LHC beams travel through Ultra High Vacuum conditions of better than 10^{-8} mbar. The beams are typically a few millimetres wide in the transverse plane, but are focused at the detectors to a width of around 14 microns using quadrupole magnets. The nominal operating luminosity is $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, however, the current operating luminosity is typically around $5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, yielding 20-30 proton-proton interactions in each bunch crossing. At the LHCb experiment, the luminosity is around $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$, yielding less than 2 proton-proton interactions per bunch crossing (on average). This lower luminosity is a deliberate design choice, since the presence of fewer proton-proton interactions in a bunch crossing means that LHCb can reconstruct secondary vertices within each proton-proton collision more efficiently. This also has the additional advantage that the sub-detectors at LHCb suffer less radiation damage. The reduced luminosity is achieved by de-focusing the beams, and is known as luminosity levelling.

The protons are collided at four main sites around the LHC ring. These sites house four different large detector experiments. The ATLAS [84] and CMS [85] detectors are ‘general purpose’, and are mainly designed to search for the direct production of new particles or the presence of new physics effects. The ALICE [86] detector is designed to probe collisions of lead ions, investigating quark-gluon plasma effects. The LHCb [80] detector is discussed in more detail in the following section.

3.2 The LHCb experiment

The LHCb detector [80] is a fully instrumented forward-arm spectrometer located at Point 8 of the LHC ring. The primary design purpose of the detector is to measure the production and decay of particles containing b quarks.

At the LHC, the dominant production mechanisms of b quarks are gluon-gluon splitting, quark-antiquark annihilation and gluon fusion [80]. The parton distribution functions for the initial states mean that the overall system tends to be boosted along the direction of the beamline; one colliding parton is at high x , and the other at low x .

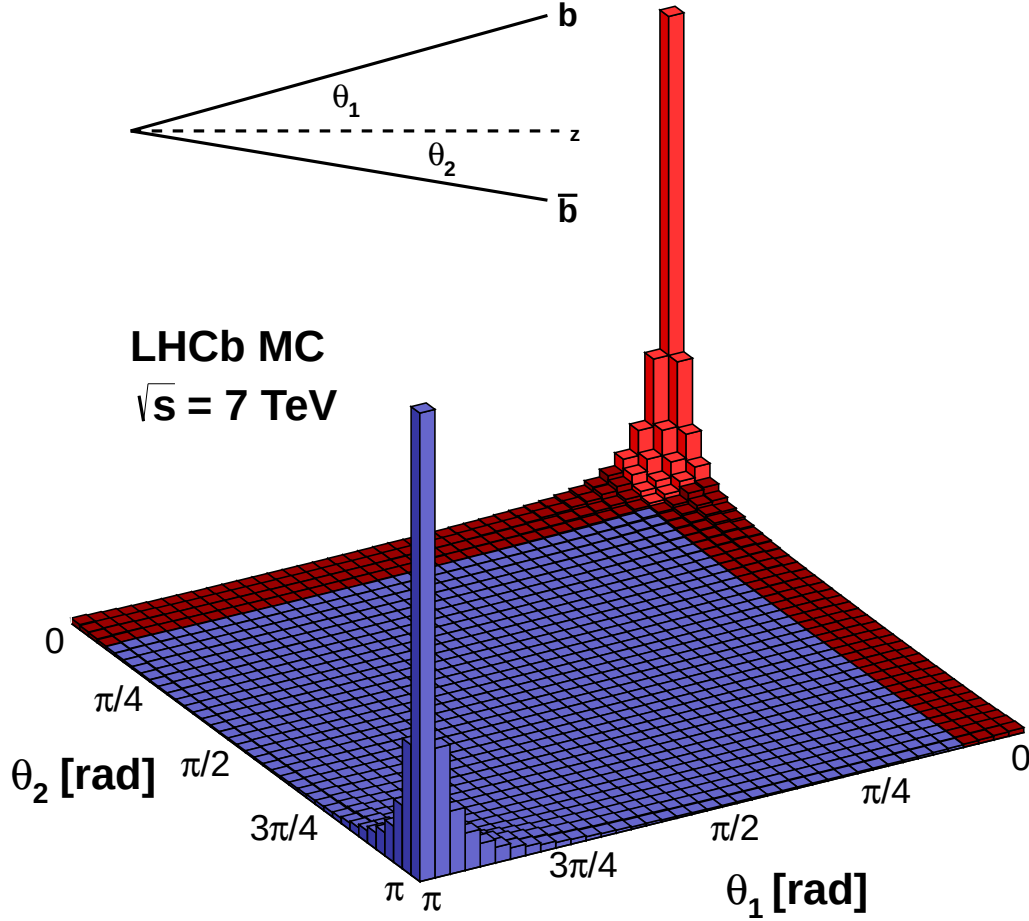


Figure 3.3: Production angles of $b\bar{b}$ pairs at the LHC. The red region shows the acceptance of the LHCb detector, with the light red region highlighting the events where both beauty particles are produced in the LHCb acceptance. Source: The LHCb collaboration

This means that the final state particles are produced close to the beam axis: production is disfavoured as the sum in quadrature of the b-quark mass and transverse momentum increases [87]. Production is therefore favoured in the forward and backward regions; the angles of the b and $\bar{b}$ are highly correlated, as shown in Figure 3.3.

The LHCb detector is consequently designed with sufficient coverage to detect many of the B hadrons that are pair-produced. In order to fit into the existing cavern at point 8 of the LHC tunnel the LHCb detector only probes the forward region (and

Year and Polarity	2011 MU	2011 MD	2012 MU	2012 MD
Crossing angle $(x - z) / \mu\text{rad}$	520	20	236	236
Crossing angle $(y - z) / \mu\text{rad}$	35	35	100	100

Table 3.1: Beam crossing angles at LHCb. The data taken in 2010 had the same crossing angle as that in 2011. MU and MD are abbreviations of the magnet up and magnet down polarities respectively.

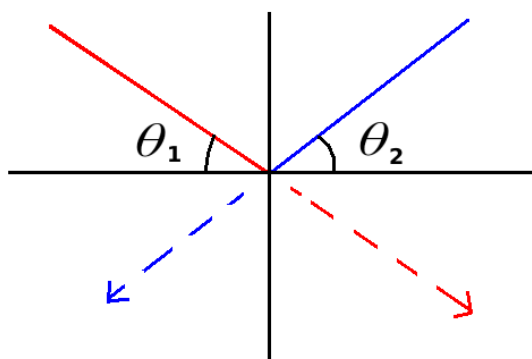


Figure 3.4: Schematic to show the beam crossing angles. In the plane shown here the beam crossing angle, θ , is given by $\theta = (\theta_1 + \theta_2)/2$.

not the backward region as well). The maximum angular acceptance of the detector is ± 300 mrad in the horizontal plane and 250 mrad in the vertical plane; the full angular coverage of the LHCb detector corresponds to approximately $2.0 < \eta < 4.5$, where the upper limit reflects the size of the beampipe. The LHCb detector structure is outlined in Figure 3.5, which also shows the co-ordinate system used at LHCb, with the z -axis corresponding to the beam direction², the y -axis in the vertical direction and the x axis constructed to create a right-handed co-ordinate system. The proton-proton interactions take place close to the origin of the co-ordinate system. The LHC beams are displaced before the collisions occur, and the collisions take place at small angles. This setup ensures that there is a localised interaction point. The crossing angles are also chosen to minimise the effect of the LHCb dipole magnet on the beams. They are shown in Table 3.1 and Figure 3.4, but are sufficiently small to be treated as negligible for the rest of this thesis. We now discuss the different subsystems of the LHCb detector and their use.

²This statement ignores the small change in beam direction due to the crossing angle.

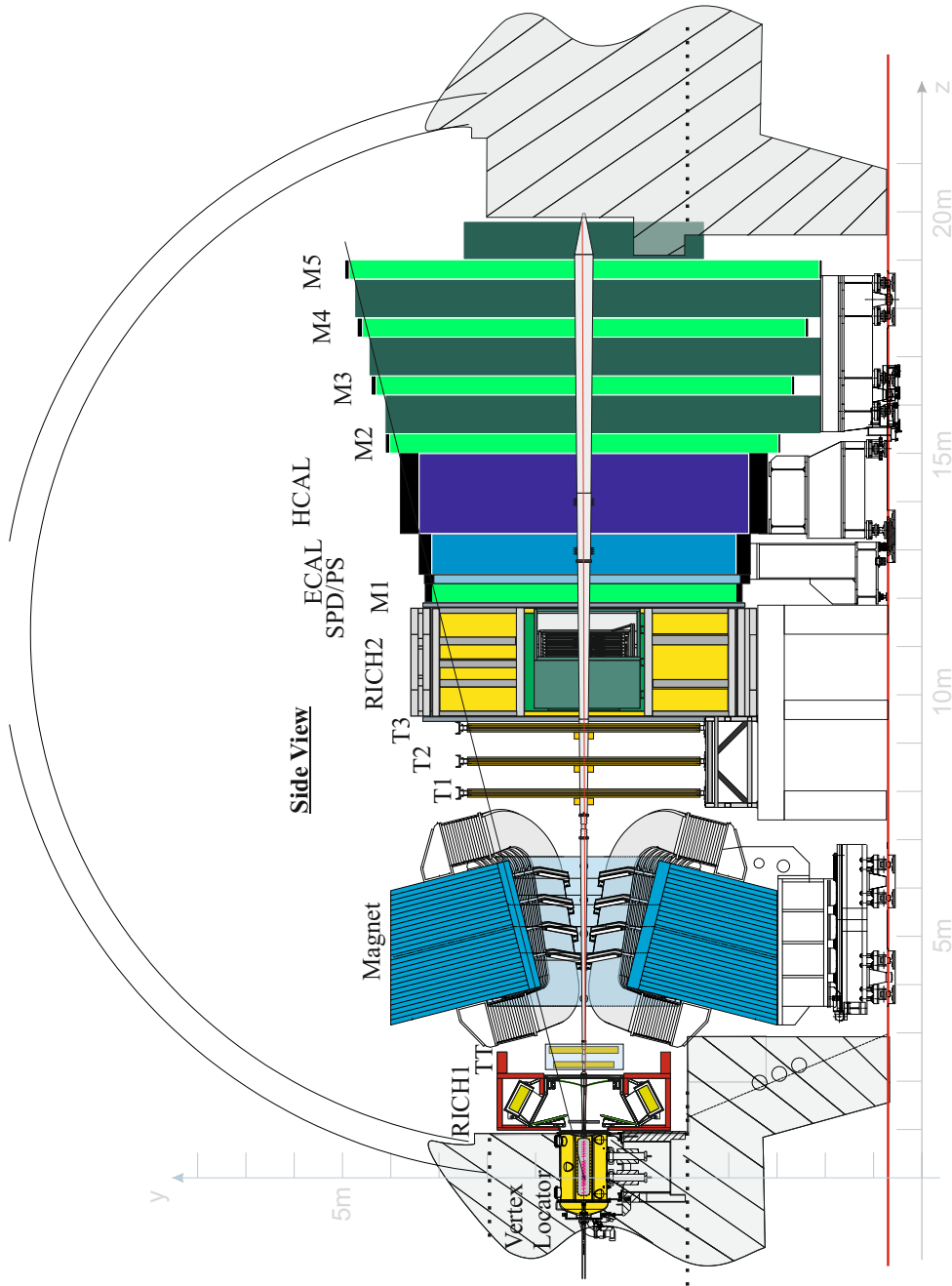


Figure 3.5: The LHCb detector. The different sub-detectors are discussed in detail in the text. We note two definitions here: the x axis is defined so that it goes into the page. We also refer to ‘downstream’ within the detector to refer to regions which are further away from the nominal interaction point in the direction of travel of particles passing through the detector, so that, for example, the HCAL is downstream of RICH 2. Source: The LHCb collaboration

3.2.1 Beampipe

The beampipe surrounds the LHC beam as it passes through the centre of the LHCb detector, and therefore occupies the high rapidity region. This means that particles from the interactions of interest will pass through the beampipe. The beampipe is therefore designed to contain a minimal number of interaction lengths since interactions would produce secondary particles which would pass into the LHCb detector, increasing occupancy. This would reduce the performance of the RICH detector and the tracking subsystems [80].

The beampipe consists of four main conical sections: the three closest to the interaction point are made of beryllium (corresponding to 12 metres of the 19 metre beampipe), the one furthest away is made of stainless steel. Beryllium was chosen for the region closest to the interaction point because of its high transparency to particles produced in proton-proton collisions. Whilst beryllium is the optimal material, its toxicity, fragility and cost are drawbacks which mean that stainless steel was chosen for the region with the lowest particle density. Flanges, bellows and the VELO exit window are made of aluminium alloys. The first cone is 1 mm thick, and includes a 25 mrad half-angle cone. The other Beryllium sections range from 1 mm to 2.4 mm thick, and are at half-angles of 10 mrad. The vacuum inside the LHCb beampipe is isolated from the main LHC vacuum, to allow interventions within the LHCb vacuum system.

3.2.2 Vertex Locator

The Vertex Locator (VELO) joins onto the beampipe, surrounding the location where the LHC beams cross. The VELO is designed to track the production of charged particles in the proton-proton collisions, enabling the reconstruction of individual interaction and decay vertices within the proton-proton events. This is particularly important at LHCb since many analyses search for secondary vertices where a B-hadron decays³. As well as allowing efficient vertex reconstruction, the VELO is also an important component within the LHCb tracking systems [80].

The VELO consists of 21 stations, each divided into two modules (left and right). Any track within the kinematic acceptance of the LHCb detector will pass through at least three stations. There are an additional 2 stations upstream of the nominal interaction

³A typical B hadron will fly a distance of $\mathcal{O}(1\text{ cm})$ before decaying.

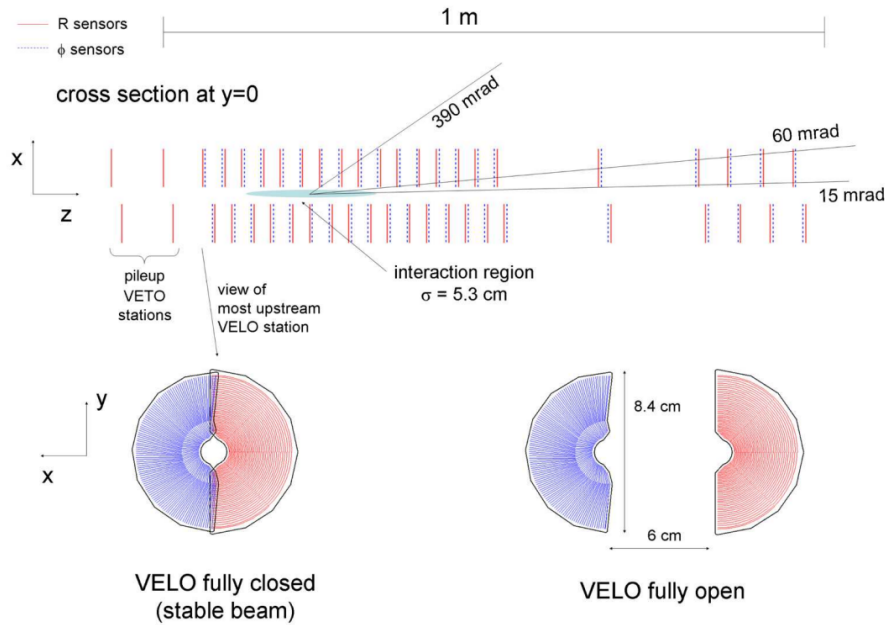


Figure 3.6: The layout of the VELO, showing the different stations, and the region of overlap when the VELO is fully closed. Source: The LHCb collaboration

point, perpendicular to the beam line. These form the pile-up veto system, and can be used to reject events containing too many proton-proton interactions. The structure of the VELO is shown in Figure 3.6. Each VELO station contains silicon tracking sensors designed to measure the radial and azimuthal positions of any track passing through it, and associated electronic chips which readout this information. This information is combined with information from the other sub-detectors to reconstruct tracks. In order to reduce ‘edge effects’ and improve knowledge of the detector alignment the left and right VELO modules are designed to overlap [80].

The VELO sits inside the LHC vacuum. However, to protect the vacuum from potential leaks of the gas in the modules, the two halves of the VELO are placed inside aluminium boxes. These boxes also shield the VELO, so that the radio frequency (RF) electromagnetic fields generated by the LHC beam do not affect the electronics in the modules. For this reason the aluminium boxes, shown in Figure 3.7, are called ‘RF boxes’. The boxes are corrugated on their inner faces to enable the two sides of the VELO to slot together and provide the overlap between the left and right modules. In this region the RF box is known as the RF foil. The RF foil makes the largest contribution of the VELO to the material budget; the entire VELO corresponds, on average, to 17.5% of a radiation length [80].

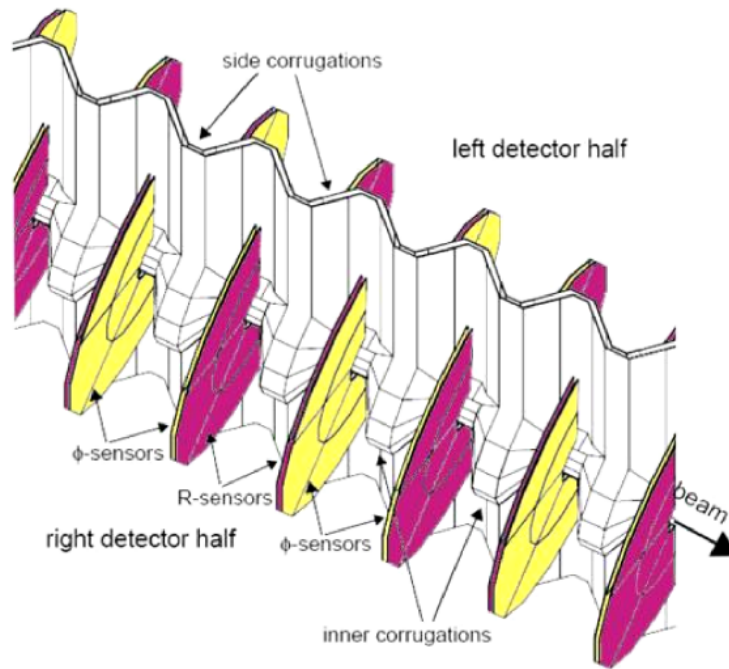


Figure 3.7: The RF foils and the modules within the VELO. Source: The LHCb collaboration

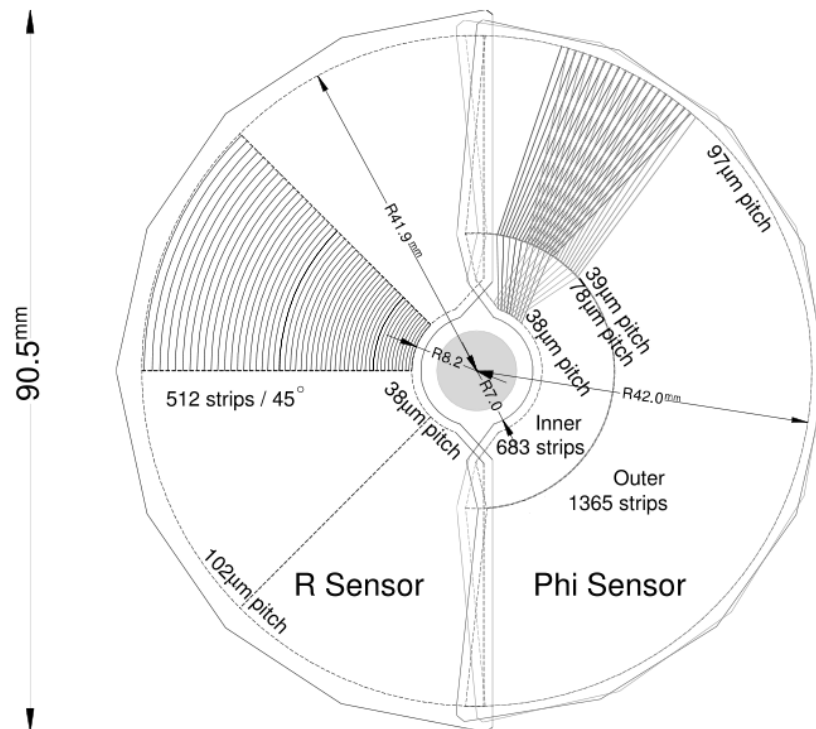


Figure 3.8: The R and ϕ (Phi) sensors in the VELO. The strip structure of two adjacent ϕ sensors is shown, along with the strip structure of one R sensor. Source: The LHCb collaboration

In order to make the best measurements of vertices and the impact parameters of tracks with respect to these vertices it is advantageous to have the VELO as close as possible to the interaction region. However, the beam spot size decreases between the beams being injected and stable running. The VELO is therefore designed to retract: during stable running the VELO is at a distance of 5 mm from the beam, whilst during the injection phase the VELO is retracted by 3 cm.

The VELO is designed to measure tracks within a cylindrical geometry, as studies have shown that this allows for faster track reconstruction. The sensors are split in two; half measure the radii of the hits (the distance from the beam-line) and the other half measure the azimuthal angles of the hits. The sensors are therefore labelled as R -sensors and ϕ -sensors respectively (shown in Figure 3.8). By considering the position of each sensor within the VELO, the third dimension is added. Both sensor types are 300 microns thick, with the active area extending from a radius of 8 mm to a radius of 42 mm, but their inner strip structure is very different.

The R -sensors are divided into 4 regions, each covering 45° and containing 512 strips. Each strip forms an arc of a circle with the centre being the nominal beamline. The decision to split the sensor into different regions reduces the capacitance and the occupancy of each strip. The pitch (the distance between adjacent strips) increases linearly with the radius of the sensor, from 38 microns (the minimum achievable) at the innermost radius to 102 microns at the outer radius. This ensures that measurements at each radius contribute with equal weight.

In the simplest case, the ϕ -sensor would simply consist of radial strips, with the pitch increasing linearly with radius, with the minimum allowed pitch at the inner radius. However, this would make the pitch too large at the outer radius, resulting in too high strip occupancies. As a solution, the ϕ sensor is divided into two regions. In the inner region, the system is set-up as described above. In the outer region, starting at a radius of 17.25 mm, the pitch is set to be half that of the inner region at the same radius. In addition, the strips do not run radially in each region. The strips run at an angle of $\sim 20^\circ$ to the radial direction in the inner region, then at an angle of $\sim 10^\circ$ in the outer region. The direction of this skew changes (from e.g. $+20^\circ$ to -20°) from one ϕ -sensor to the next.

The VELO is the part of the LHCb detector which lies closest to the collision point. It is therefore the part of the detector which suffers the most radiation damage, and so

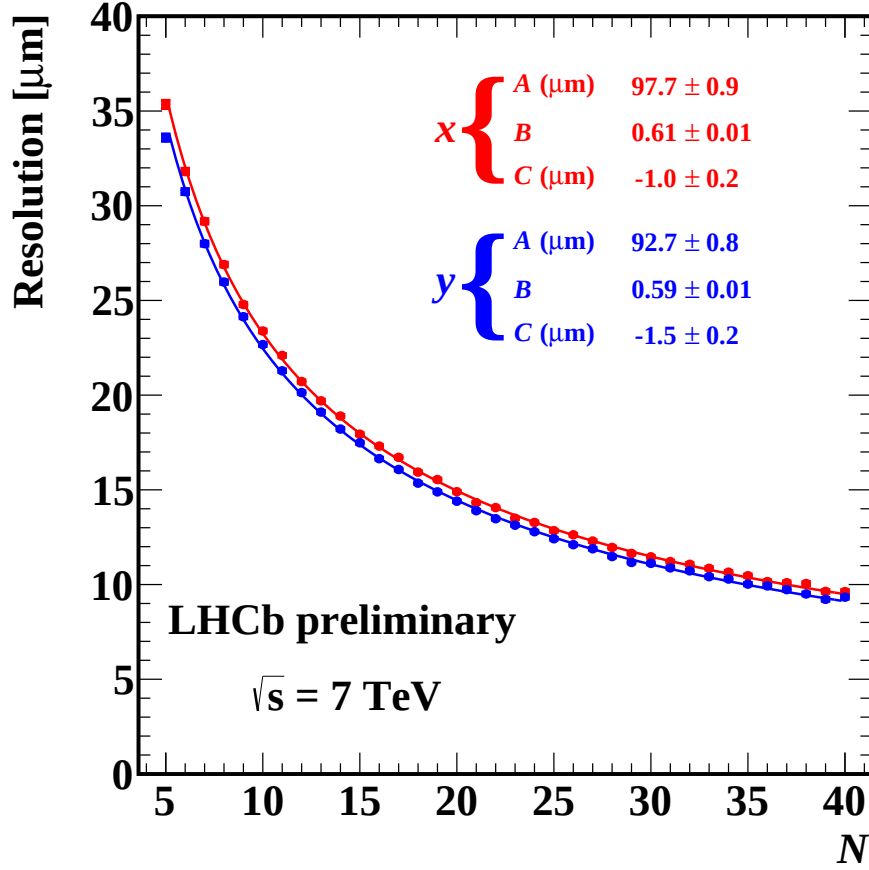


Figure 3.9: The position resolution of primary vertices at LHCb in the $x - y$ plane. The fit function used is $A/N^B + C$, where N is the number of tracks associated with the primary vertex. Source: The LHCb collaboration

performance degrades with time. There is a replacement VELO available should such damage become significant [80].

The position of vertices within the VELO can be measured with a typical resolution in the transverse plane of 10-20 microns (shown in Figure 3.9), and a typical resolution in the z direction of 50-100 microns, depending on the number of tracks which form the vertex. The VELO also plays a key role in determining the impact parameters of tracks with respect to these vertices, providing an impact parameter resolution of about 13 microns for high momentum tracks.

3.2.3 The LHCb magnet and tracking detectors

The LHCb tracking systems are the VELO, the TT and the T1, T2 and T3 stations [80]. The TT and the inner tracker (the IT) of the T1-3 stations are silicon trackers. The outer tracker (the OT) of the T1-3 stations is a drift-time detector. The LHCb dipole magnet is located between the TT and the T1 station, and also plays a pivotal role in tracking measurements at LHCb. Tracking measurements are used to reconstruct electrically charged particles as they traverse the detector, determining their momenta (and related quantities) [80].

The LHCb magnet

The LHCb dipole magnet is used to measure the momentum of charged particles (as the radius of curvature of a particle in a magnetic field is determined by the particle's momentum). The magnet covers the complete LHCb acceptance, with the field aligned in the $\pm y$ -direction. Uniquely at the LHC, the magnetic field polarity can be switched. This is useful for analyses of asymmetries as many systematic effects can be removed by considering the mean of measurements made with opposite field polarities. The magnet has a peak field strength of 1.1 T, and provides an integrated field strength of ~ 4 Tm for tracks of length 10 m. The peak field is achieved between the TT and the T1 stations, whilst the field inside the RICH detectors is as small as possible (< 3 mT), as discussed later. The coils used to achieve the magnetic field are shown in Figure 3.10. Each coil has a mass of 27 tons, and surrounds a yoke of mass 1500 tons [80].

In order to reach the required momentum resolution, the magnetic field was measured using an array of Hall probes to a relative accuracy of 4×10^{-4} , and the position of the peak B -field was measured to a precision of a few millimetres. The magnetic field is shown in Figure 3.11. The reproducibility of the magnetic field was also determined at this level. This is especially important at LHCb as hysteresis effects become important due to the periodic changing of the field polarity. In most regions of the detector, the agreement between the measured field and the calculated field is better than 1% [80].

The silicon trackers

The silicon trackers are located in the TT and the IT within the T1-3 stations. The detectors use silicon microstrips, with a strip pitch of about 200 microns. Each silicon

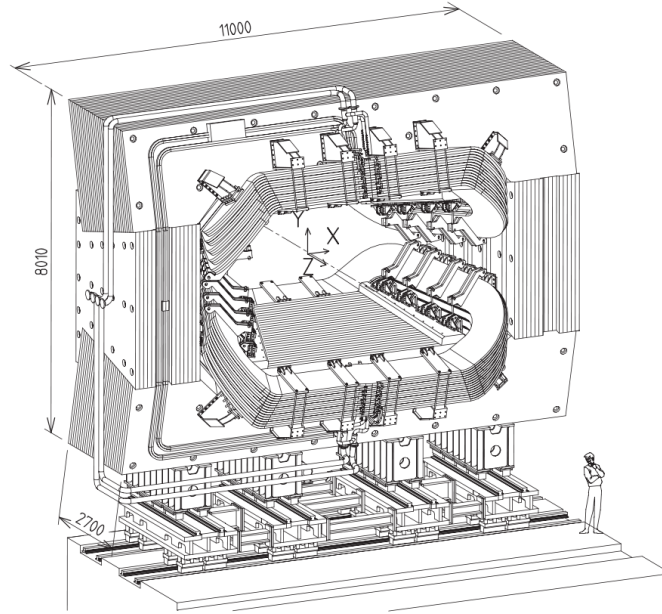


Figure 3.10: The LHCb magnet. Dimensions shown are in millimetres. Source: The LHCb collaboration

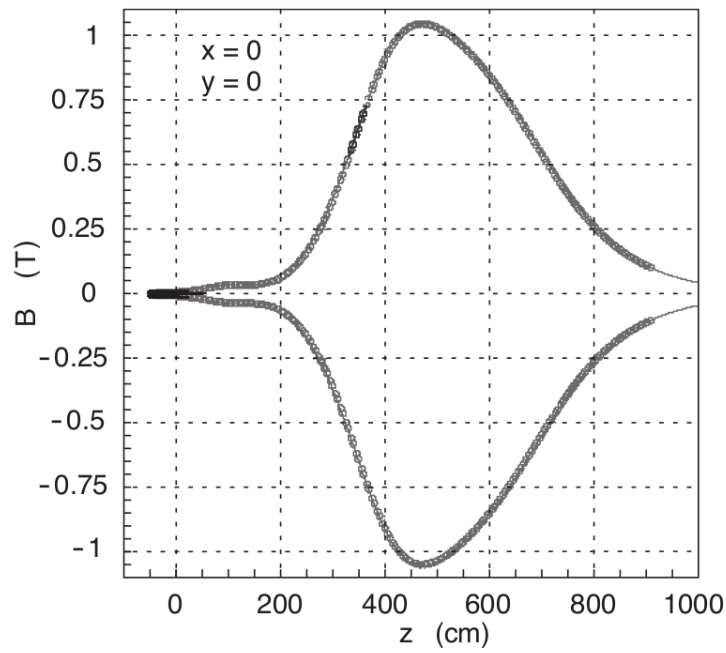


Figure 3.11: The magnetic field throughout the LHCb detector for the two magnet polarities used. The measured field is shown as points, with the line showing the field predicted by the detector model. Source: The LHCb collaboration

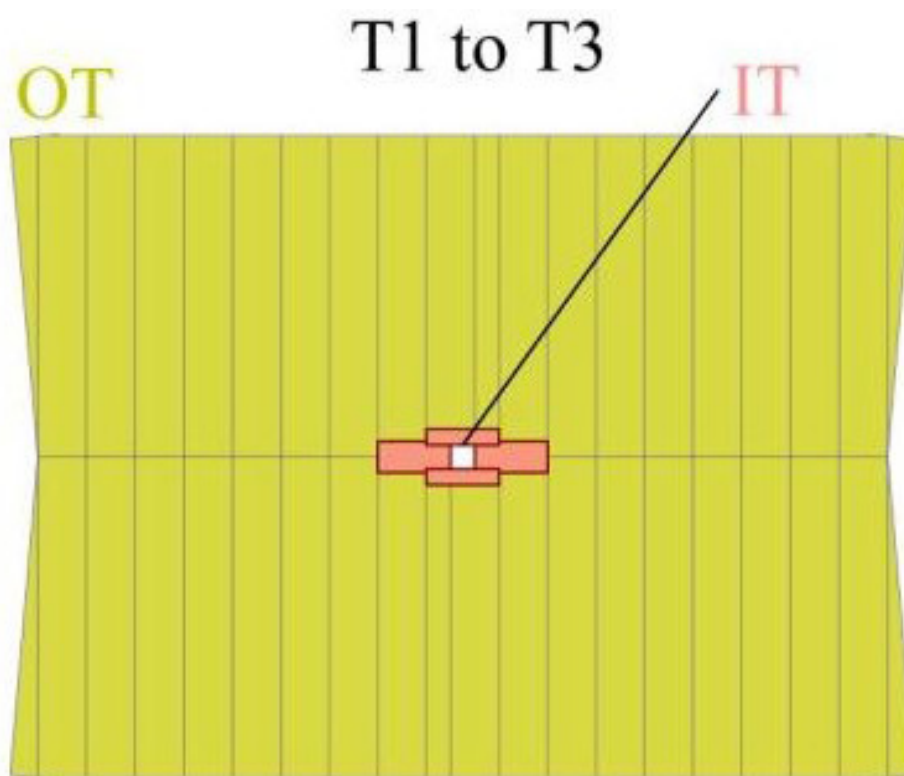


Figure 3.12: The coverage of the IT and OT within the T1-3 stations. Source: The LHCb collaboration

tracker consists of 4 layers. The strips in the second and third layers are rotated by $\pm 5^\circ$ with respect to those in the first and fourth layers. The TT is located between the RICH1 detector and the LHCb magnet. It is designed to detect low momentum particles that would otherwise be swept out of the LHCb detector acceptance by the dipole magnet. It is 150 cm wide and 130 cm high, and has an active area of 8.4 m². The TT corresponds to about 4% of a radiation length. The IT, located in the T1-3 stations, covers a 120 cm by 40 cm cross-shaped region with an active area of 4.0 m², and corresponds to about 3.5% of a radiation length. This region of the detector sees the highest density of particles passing through it: the IT covers only 2% of the detector acceptance but approximately 20% of tracks pass through it. It is for this reason that silicon technology is used [80].

The outer tracker

The outer tracker makes up the rest of the T1-3 tracking stations (as shown in Figure 3.12). It is designed as an array of individual, gas-tight Kapton straw-tube modules. Each module contains 4.9 mm diameter drift-tubes. The gas (70% Argon, 30% Carbon Dioxide) ensures a fast drift time (below 50 ns), and a resolution of better than 200 microns. As for the IT, there are 4 layers, with the middle layers arranged at $\pm 5^\circ$ to the vertical. The total active area of each station is 5971×4850 mm². Occupancies are smaller than 10% (with the IT covering the region with higher particle density) [80].

3.2.4 Ring imaging Cherenkov sub-detectors

The ring imaging Cherenkov (RICH) sub-detectors measure Cherenkov radiation emitted by charged particles that pass through them. This, along with knowledge of the particle momenta from the tracking systems, allows particle masses to be determined. The RICH sub-detectors provide particle ID for particles with momenta between 2 GeV and 100 GeV [80].

The sub-detector works on the principle that charged particles which travel through a medium faster than the speed of light in that medium will emit coherent radiation. The angle of this radiation with respect to the particle's direction of propagation, θ , is given by $\cos(\theta) = \frac{1}{\beta n}$ where β is the particle's speed and n is the refractive index of the medium. This gives rise to characteristic rings of light formed when a particle traverses the medium. Measurements of the radii of the rings will give the emission angle, which in

turn provides the particle's speed. The RICH sub-detectors rely on sophisticated pattern recognition algorithms to search for photons emitted by particles [80].

We consider the two RICH sub-detectors separately as their design is different, having been optimised to measure particles with different momenta.

RICH 1

The RICH 1 detector (shown in Figure 3.13) is located between the VELO and the TT. It contains aerogel and C_4F_{10} gas as the radiator media. Aerogel has a refractive index of 1.030 for wavelengths of 400 nm, whilst C_4F_{10} has a refractive index of 1.0014 at the same wavelength. The Cherenkov angles of these media as a function of particle species and momenta are shown in Figure 3.14. These radiators were chosen to provide efficient detection of particles with $1 < p < 60$ GeV; the particle ID performance seen in LHCb data is shown in Figure 3.15 [80].

As charged particles traverse the RICH 1 detector they radiate Cherenkov photons uniformly along their tracks. These photons are then focused using spherical mirrors (with radii of curvature 2700 mm), producing rings on a detection plane which are then reflected onto the photon detectors using flat mirrors. This is necessary in order to keep the photon detectors out of the acceptance. The photons are readout using Hybrid Photon Detectors (HPDs). These are vacuum photon detectors in which a photoelectron is released by the incident Cherenkov photon. This photoelectron is then accelerated by a high voltage (10 to 20 kV), creating around 5000 electron-hole pairs. The readout electronics are able to distinguish individual photons [80] with high efficiency. Quartz plates separate the HPDs from the radiator gas, and the whole RICH 1 detector is enclosed using a 6 sided box made from 30 mm thick aluminium alloy plates. The thickness of RICH 1 is approximately 8% of a radiation length [80].

The RICH 1 detector is shielded from the magnetic field using iron plates which reduce the magnetic field in the HPD plane to a maximum value of 2.4 mT [80].

RICH 2

The RICH 2 detector is designed to detect higher momentum particles, and is located downstream of the magnet, between the last tracking station and the first muon station. It contains a CF_4 radiator, which provides efficient PID for particles with $15 < p < 100$ GeV

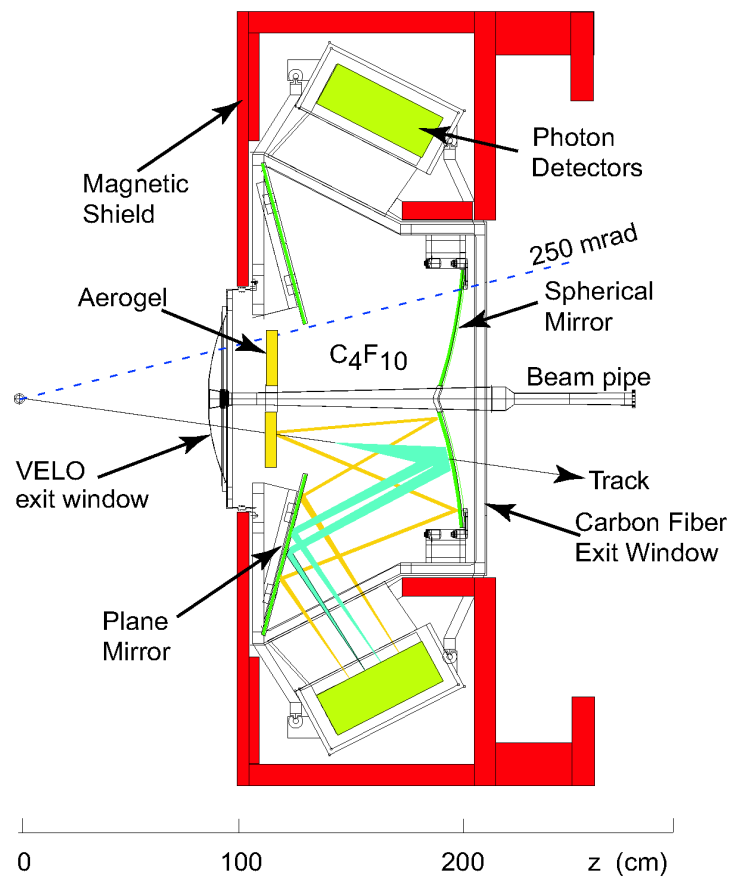


Figure 3.13: The design of the RICH 1 detector. The mirrors reflect Cherenkov radiation from charged particles in the LHCb acceptance into the photon detectors, with coherent radiation forming characteristic rings. The photon detectors are located outside the acceptance to reduce the material that particles in the detector pass through, and are shielded from the magnetic field. A schematic of a particle traversing the detector is shown in the figure. Source: The LHCb collaboration

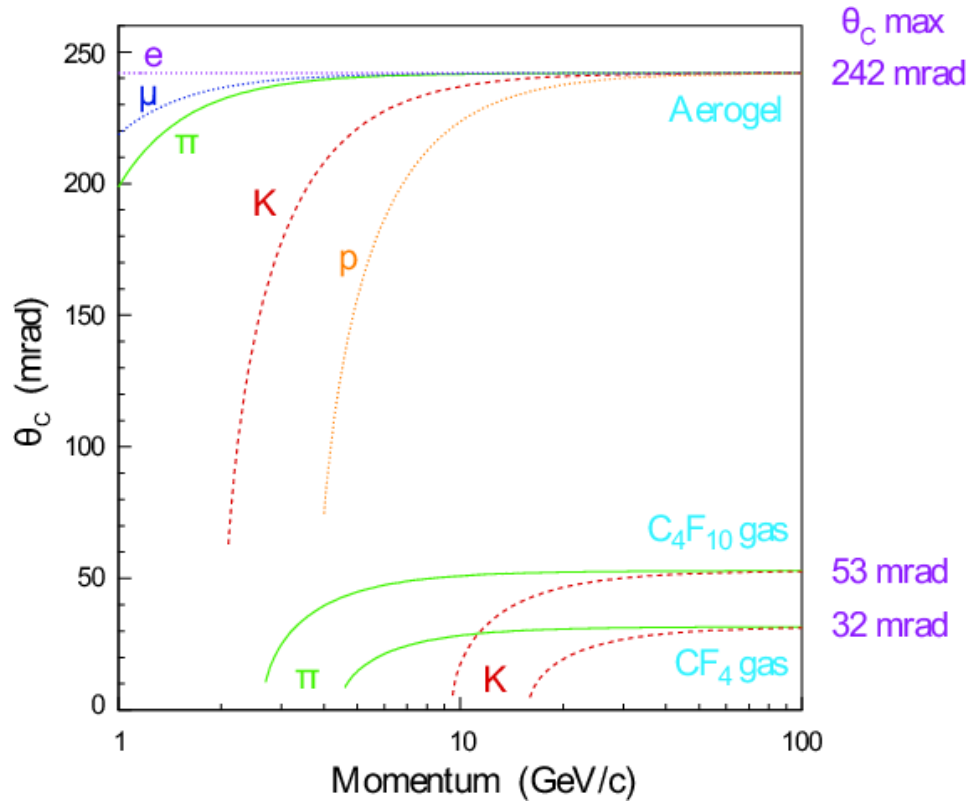


Figure 3.14: The Cherenkov angles of the different radiators used in RICH 1 and RICH 2 for different particle species. RICH 2 is designed to provide particle identification capability for higher momentum particles than RICH 1. Source: The LHCb collaboration

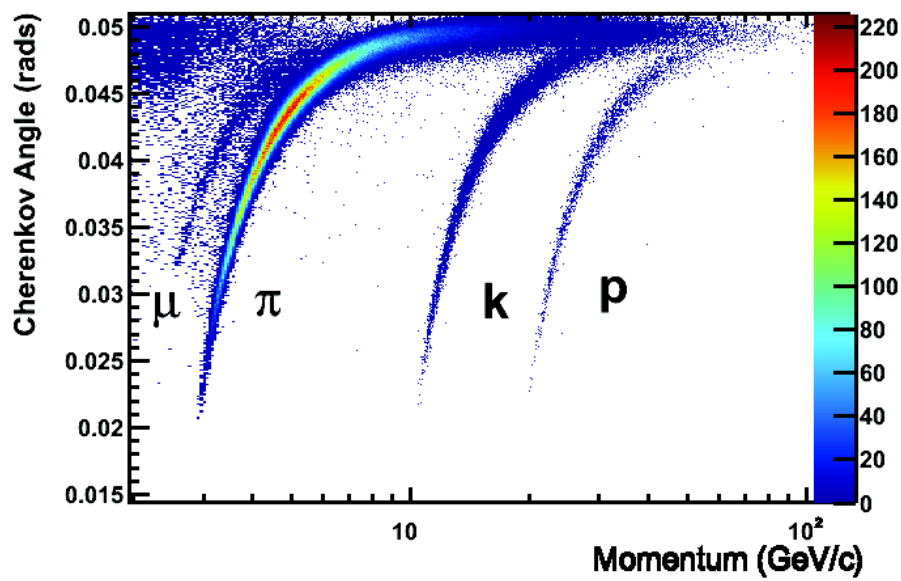


Figure 3.15: The performance of the RICH 1 sub-detector in LHCb data. The ability to distinguish different particles is clear. From Ref. [88].

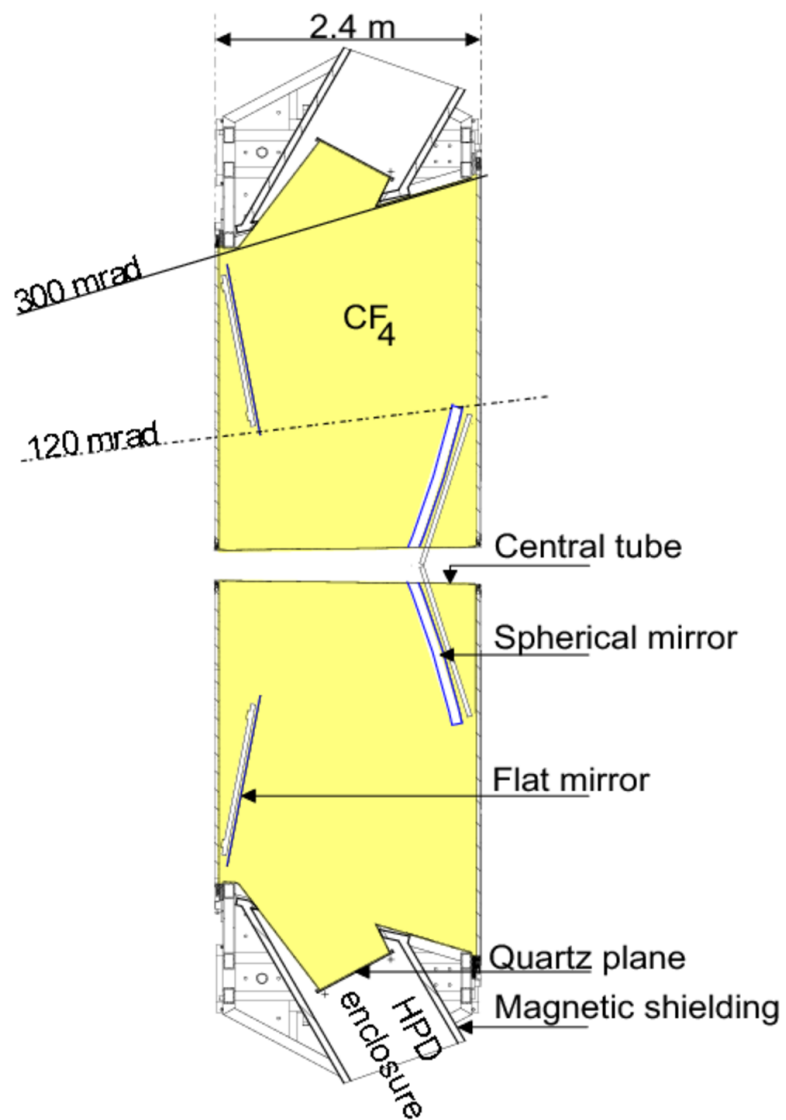


Figure 3.16: The design of the RICH 2 detector. The design is similar to that of RICH 1.
Source: The LHCb collaboration

(as shown in Figure 3.14). The refractive index of CF_4 is 1.0005 for light with a wavelength of 400 nm [80].

The RICH 2 detector has a reduced angular acceptance of ± 120 mrad in the horizontal plane and ± 100 mrad in the vertical plane, extending down to the beampipe. It is shown in Figure 3.16. It corresponds to a thickness of 15% of a radiation length.

As for RICH 1, spherical mirrors focus the Cherenkov radiation into rings, and flat mirrors direct this radiation onto the HPDs. Quartz plates separate the HPDs from the radiator gas, and the entire detector is contained within aluminium panels. The HPDs are shielded from the magnetic field using iron plates. The residual field is between 0.2 and 0.6 mT [80].

3.2.5 Electromagnetic calorimeter and associated subsystems

Here we discuss three related parts of the detector: the electromagnetic calorimeter (ECAL), the preshower (PRS) detector and the scintillator pad detector (SPD). This system is designed to determine the transverse energy of electron and photon candidates, with the different elements used to distinguish between species of particles. The system is also used to select events in the LHCb hardware trigger (discussed in Section 3.2.8) within a $4\ \mu\text{s}$ window [80].

The PRS and SPD are located before the ECAL, and are separated by a lead converter of 2.5 radiation lengths for electrons (only 1 % of pions will interact with the converter). This length was chosen as a compromise between energy resolution and trigger performance. The ECAL, located after the PRS, is 25 electron radiation lengths in depth. This was chosen to ensure that an entire electron shower is contained within the ECAL systems, which improves the energy resolution [80].

Electrons, being charged, will interact with the SPD, whereas photons will not. This enables photons and electrons to be distinguished. The photons tend to begin showering in the PRS, though some may reach the ECAL before converting into an electron-positron pair. The presence of the PRS in addition to the ECAL enables the background of charged pions to be distinguished from electrons and photons since, even if pions begin showering in the ECAL, they are very unlikely to shower within the PRS. The shower structure of different particles within the LHCb calorimeter system is shown in Figure 3.17.

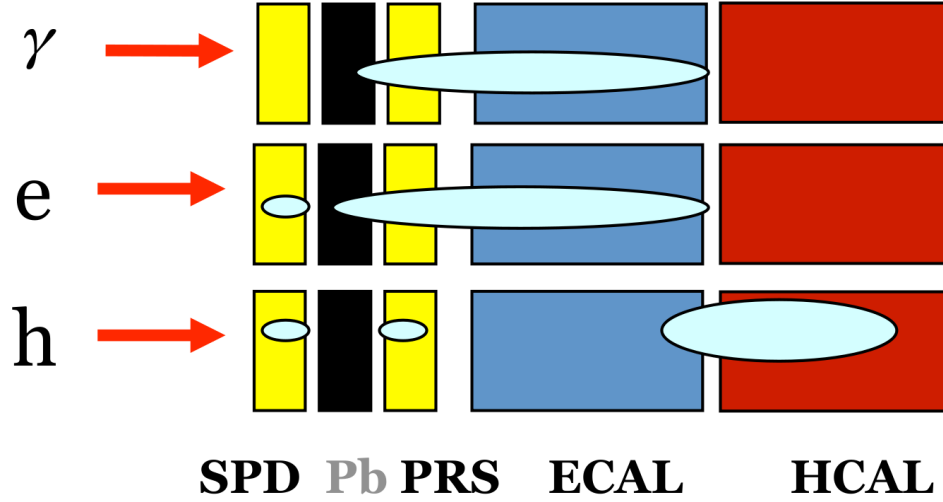


Figure 3.17: Schematic to show the shower structure within the LHCb calorimeters, from Ref. [89]. Different particle species produce different shower structures, allowing for efficient particle identification.

The PS, SPD and ECAL use a lateral segmentation with 3 different regions of the detectors, as the hit density varies by two orders of magnitude across the calorimeter surface. The structure is shown in Figure 3.18. The PRS and SPD contain 3072 cells in the inner region, with 3584 in the middle and 5376 in the outer regions. Each cell is attached to readout channels to determine the energy of electromagnetic particles as they shower [80].

The PRS and SPD are two almost identical planes of rectangular scintillator pads of high granularity. The sensitive area is about $7.6\text{m} \times 6.2\text{m}$. The PRS is roughly 0.45% smaller than the SPD, in order to cover the same angular acceptance.

The ECAL uses shashlik calorimeter technology, an alternating sampling scintillator and lead structure that is readout by plastic wavelength-shifting (WLS) fibres. The ECAL, like the PRS and SPD, has finer granularity in the region with the highest particle density. The energy resolution of an ECAL cell is about $\sigma_E/E = 9.0\%/\sqrt{E/\text{GeV}} \oplus 0.8\%$ [80] for photons or electrons.

Measurements of electrons at LHCb use information from both the tracking systems and the ECAL since the ECAL is used to correct the measured track momentum for the effect of bremsstrahlung radiation. The bremsstrahlung radiation appears in the detector as a neutral cluster which ‘points’ back to a track. At sufficiently low momenta the electrons will be bent by the magnet into a different ECAL cell from the radiation. At

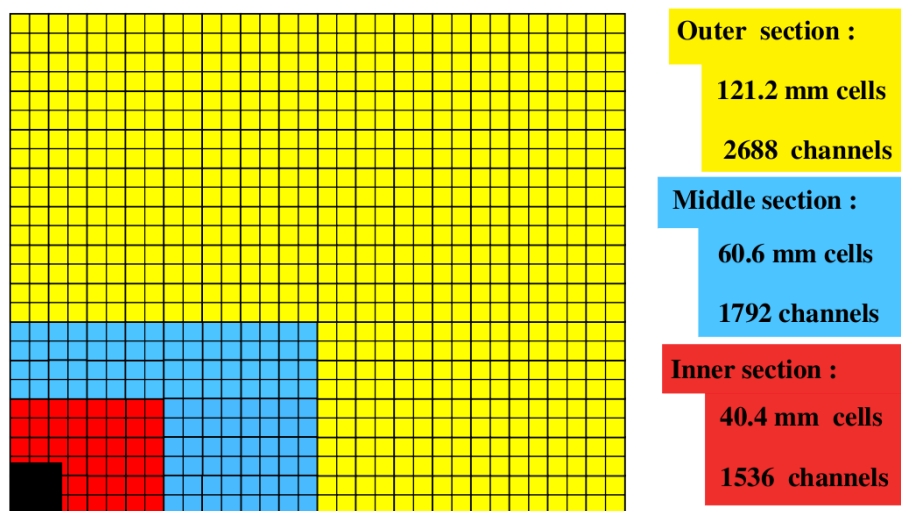


Figure 3.18: The design of the PRS, SPD and ECAL contains three regions of different granularity, as the particle density varies across the detector. The innermost black region is the beampipe. This schematic shows only one quadrant of the sub-detectors, and the lines on the schematic do not correspond to the cell sizes. Source: The LHCb collaboration [80]

higher energies the tracks will be (almost) straight, and the electrons will give deposits in the same cells as the bremsstrahlung photons.

The gain within each ECAL cell is optimised for the low transverse energies studied within B physics, with the calorimeter cells saturating for $E_T > 10$ GeV. This enables a B mass resolution of 65 MeV for the process $B \rightarrow K^*\gamma$. This is however problematic for measuring higher energy electrons if there is significant bremsstrahlung, as the saturation dilutes the ability to accurately determine the energy deposit from the radiation, and correct the electron momentum measured by the tracking systems. This is only an issue, however, when there is a significant amount of high energy bremsstrahlung [6].

3.2.6 Hadronic calorimeter

As for the ECAL, the hadronic calorimeter (HCAL) is mainly used to enable efficient and fast selection of events within the hardware trigger. The HCAL is designed to detect hadronic particles as they pass through the LHCb detector, and is 5.6 hadron interaction lengths deep. This length is determined by the available space within the LHCb cavern; hadron showers may not be fully contained within the calorimeter. The HCAL consists of two regions, with larger cell sizes in the region with lower particle density. This structure

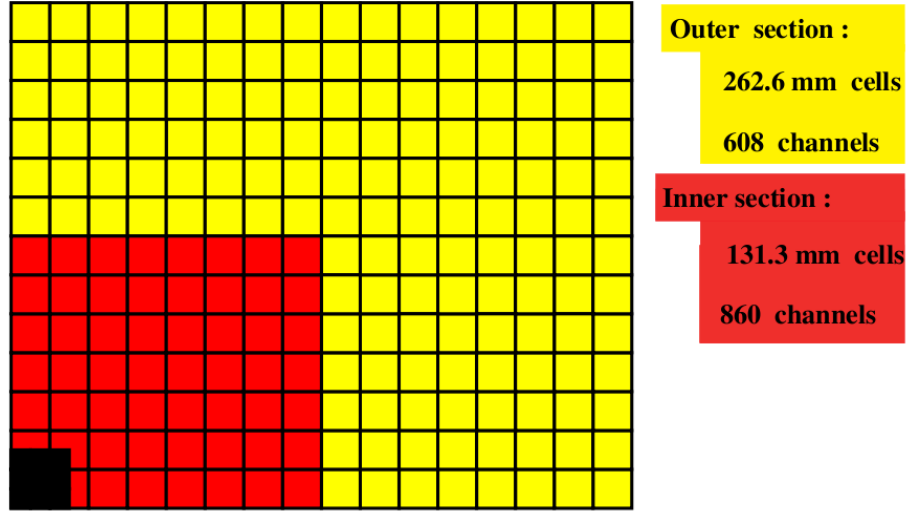


Figure 3.19: One quadrant of the HCAL. Like the ECAL, the HCAL has sections of different granularity to account for the variation in particle density across the detector. The lines on the schematic do not correspond to the cell sizes. Source: The LHCb collaboration [80]

is different to the ECAL structure due to the larger hadronic shower dimensions, and is shown in Figure 3.19 [80].

The HCAL is constructed with scintillating tiles sandwiched between steel absorber plates. The light generated in the tiles is transmitted to photon detectors by WLS fibres. The energy resolution of an HCAL cell is about $\sigma_E/E = 69.5\%/\sqrt{E/\text{GeV}} \oplus 9\%$ [80].

3.2.7 Muon chambers

The muon chambers at the LHCb experiment (shown in Figure 3.20) are designed to allow fast selection of muon candidates but also to allow for efficient muon identification in data analysis [90]. The muon system covers the angular range 20-306 mrad in the horizontal plane, and 16-258 mrad in the vertical plane. This is sufficient to capture $\sim 20\%$ of muons from semileptonic B decays [80].

The muon system consists of five chambers or stations. The first chamber is located before the calorimeter, with the others located downstream of the HCAL. This design allows a fast estimate of the muon transverse momentum in the hardware trigger. Stations M2-M5 are interleaved with 80 cm thick iron absorbers. The minimum momentum for a muon to cross all five stations is, on average, about 6 GeV; the absorbers (including the calorimeters) correspond to 20 interaction lengths for charged hadrons. Stations

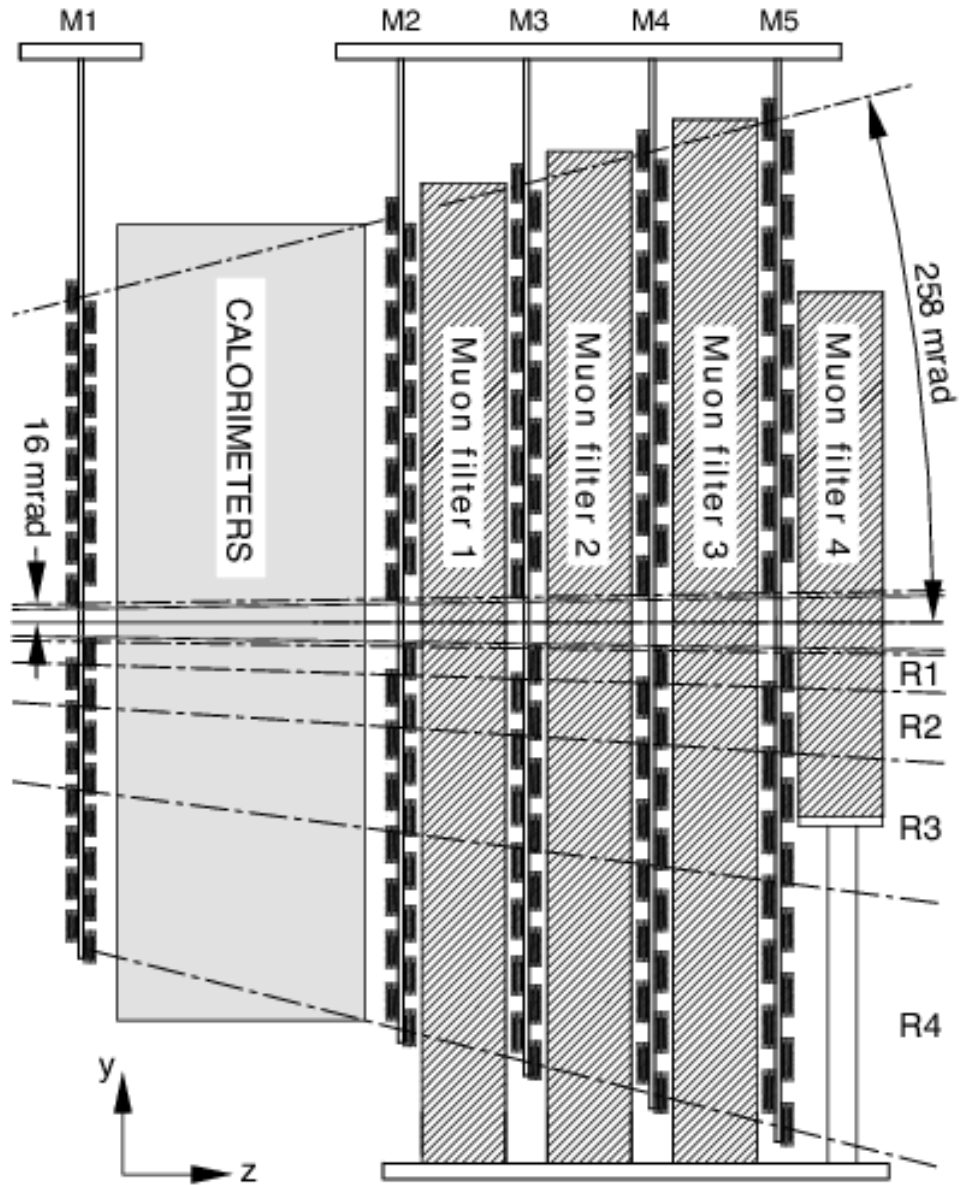


Figure 3.20: The design of the muon chambers. The muon detectors are separated by absorbers. Source: The LHCb collaboration

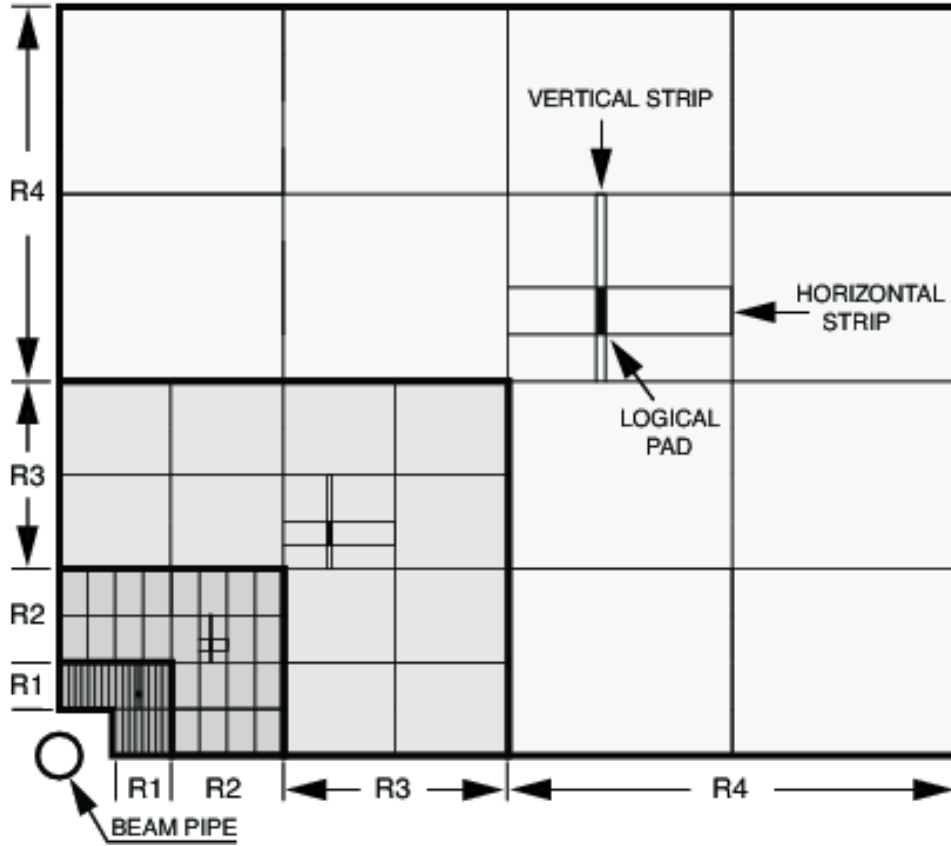


Figure 3.21: The design of each muon chamber (only one quadrant is shown). Source: The LHCb collaboration

M1-3 have high spatial resolution, enabling accurate determination of the transverse momentum. M4 and M5 provide efficient muon identification.

The size of each muon station is scaled to give the same angular coverage; the layout of each station is otherwise largely similar, and is shown in Figure 3.21. Each station is divided into four regions, R1 to R4, with each region at increased distance from the beam axis. The linear dimensions of the regions R1, R2, R3 and R4 and their segmentations scale in the ratio 1 : 2 : 4 : 8. The muon stations use two different technologies; multi-wire proportional chambers (MWPCs) are used for all regions except the inner region of M1. In this region the expected particle rate would exceed safety limits for ageing, so triple-GEM (gas electron multiplier) technology is used instead [80].

The MWPCs contain four 5 mm gas gaps, each containing an Ar/CO₂/CF₄ mixture. A muon crossing a gas gap produces electrons which are accelerated towards 2.8 kV anode wires. It should be noted that in M1 there are only two gas gaps, to reduce the material budget in front of the ECAL. Logical pads are defined by the edges where the

horizontal and vertical strips cross. A combination of wire and pad readout is used. This is necessary, as in R1 and R2, the required resolution would demand unfeasibly small logical pads. The different pads in different layers are combined using the logical ‘OR’ in order to improve trigger efficiency and readout time [90].

Triple-GEM detectors are formed by three gas electron multiplier foils between an anode and cathode plane. The gap between the foils is filled with a different Ar/CO₂/CF₄ mixture. As for the MWPCs, a muon crossing the system will produce an electron, which is attracted across the foils by electric fields. This provides good time and spatial resolution, whilst being less sensitive to radiation damage.

3.2.8 The LHCb trigger

The LHCb detector operates at a much lower luminosity than the LHC design luminosity. This means that the majority of bunch crossings contain only a few proton-proton interactions. This aids triggering and reconstruction as the occupancy is lower than at the general purpose detectors. However, there are still too many interactions to record every event. It is the job of the LHCb trigger to reduce the interaction rate⁴ from ~ 10 MHz to a recordable rate (about 3-5 kHz) [91, 92].

This reduction is done in two stages or levels. Initially, a hardware (Level-0) trigger is used to reduce the rate to around 1 MHz. This is a more manageable rate where a partial reconstruction can be performed. After this reconstruction, the high level trigger is used to confirm Level-0 decisions and then select events using information from a more detailed reconstruction. This reduces the rate to the desired 3 kHz (in data recorded in 2011) or 5 kHz (in data recorded in 2012). Individual physicists at LHCb can then select and study these data [91, 92].

The structure of the trigger is summarised in Figure 3.22 and discussed in more detail below.

The LHCb hardware trigger

The Level-0 hardware trigger attempts to reconstruct the highest E_T hadron, electron and photon clusters in the calorimeters and the two highest p_T muons in each of the

⁴The crossing rate is about 40 MHz, but many bunch crossings do not contain visible proton-proton interactions.

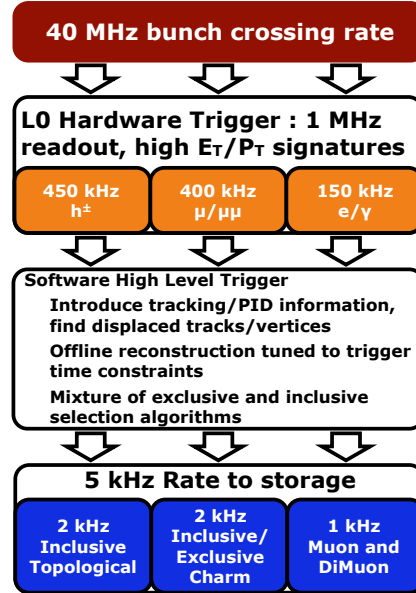


Figure 3.22: The LHCb trigger in 2012. Source: The LHCb collaboration

four muon chambers. There are also dedicated triggers designed to accept events with very low occupancies, which are of interest for the study of central exclusive production (CEP) [91].

The different Level-0 trigger lines use information about the level of pile-up, information from the calorimeters, and information from the muon chambers. The pile-up system aims to distinguish bunch-crossings containing multiple collisions from those containing single collisions. It uses 4 silicon sensors like those used in the VELO to measure the radial position of tracks. It also provides an estimate of the position of primary vertex candidates along the beamline, and the number of backward travelling tracks. The calorimeter trigger system looks for high E_T hits in the ECAL and HCAL, indicating the presence of electrons, photons, neutral pions or hadrons. Clusters are formed by adding together the transverse energy of 2×2 calorimeter cells, and selecting those clusters with the highest transverse energy. Information from the PRS and the SPD enables particle identification. The muon chambers allow stand-alone muon reconstruction, providing a transverse momentum resolution of about 20 %. This is performed by processing strip and pad elements to form towers pointing towards the interaction point. The Level-0 muon trigger looks for tracks within the muon stations, based on seeds within the M3 station. For each logical pad hit in M3, the expected position for hits in M2, M4 and M5 is found by assuming a straight line between the hit and the interaction point. Corresponding

hits are sought in these stations. Because the structure of the muon chambers is scaled to provide the same angular coverage, there is a one-to-one correspondence between pads in each chamber. This makes it simple to implement the muon trigger using only logical decisions as track extrapolation is easier and faster with such a design [91].

The Level-0 decision unit collects information from all the components to form the overall trigger. It combines information from the different systems with simple logic to provide a decision for each bunch crossing within $4\mu\text{s}$. This includes the particle time of flight, and all cable and electronic delays, meaning the trigger itself makes the decision within $2\mu\text{s}$ [91]. The Level-0 decision unit is based on a specially designed electronic ‘TELL1 board’ [93]. To enable the decisions to be made sufficiently quickly the technology and computational approach used within the Level-0 trigger were chosen with great care. The detector data are serialised, optical links are used as transport media, and high density devices are used to make decisions.

There are different possible ways for the collisions within a bunch crossing to pass the trigger, since there are many trigger lines, and if one of these is passed then the entire event will be read out and recorded (the different lines are joined using a logical OR). The lines are largely stable, with the established lines undergoing relatively few changes throughout the data taking period [91]. The collection of Level-0 lines run is limited by the space available to program the Level-0 decision unit. A typical selection of the main lines within a trigger configuration key (TCK) is shown in Table 3.2, along with their approximate rates.

To stop a few events with high occupancy dominating the CPU processing time in the software trigger, ‘global event cuts’ (GEC) are applied in the Level-0 trigger. The dominant GEC is on the number of hits in the SPD. The effect of these requirements depends on the specific physics analysis studied, and will be discussed in the relevant chapters.

The LHCb software trigger

The software (or high level) trigger (HLT) is another important component within the LHCb experiment. We therefore discuss it here.

An event which is accepted by Level-0 is transferred by the on-line system to the Event Filter Farm (EFF), which consists of approximately 1000 multi-core computing

Trigger	Selection requirements	Rate / kHz
L0Electron	$n_{\text{SPD}} < 600, E_{\text{T}}^e > 2.5(3) \text{ GeV}$	160
L0Photon	$n_{\text{SPD}} < 600, E_{\text{T}}^\gamma > 2.5(3) \text{ GeV}$	80
L0Muon	$n_{\text{SPD}} < 600, p_{\text{T}}^\mu > 1.48(1.76) \text{ GeV}$	340
L0Hadron	$n_{\text{SPD}} < 600, E_{\text{T}}^h > 3.5(3.7) \text{ GeV}$	405
L0DiMuon	$n_{\text{SPD}} < 900, p_{\text{T},1}^\mu \times p_{\text{T},2}^\mu > 1.296(1.6) \text{ GeV}^2$	75

Table 3.2: Typical Level-0 trigger lines, their selection requirements in 2011 (with 2012 requirements shown in brackets) and their typical rates in 2011 (the precise rate depends upon the instantaneous luminosity). In addition to the lines shown here, there are minimum bias lines, which do not select events at random, and low multiplicity lines, which specifically select events with a low number of SPD hits (n_{SPD}). There are also lines which do not have any requirements on the number of SPD hits, but these are only run once in every 10^4 collisions. Finally, there are also lines which monitor beam-gas collisions in events where one of the beams contains no protons in that particular beam-beam crossing.

nodes. The HLT is itself based on the same software used throughout the LHCb data processing and is a C++ application (called ‘MOORE’) that runs on every CPU in the EFF. Currently there are 26110 copies of MOORE running in the EFF. The HLT application has access to all data from the detector and takes about 30 ms to run on each event. The HLT is split into two stages: HLT1 and HLT2 [91, 92].

The main function of HLT1 is to confirm decisions made at Level-0 using more complete event information. It reduces the input rate from 1 MHz to about 50 kHz, making use of a partial event reconstruction. The off-line VELO reconstruction software is sufficiently fast to permit the full 3D pattern recognition to be run on all events that enter the HLT. This enables the primary vertices (reconstructed proton-proton interactions) to be found [91, 92]. The majority of HLT1 trigger selections (‘lines’) then select events containing tracks which are significantly displaced from these vertices. This is of particular interest since heavy flavour particles are expected to travel before decaying, creating tracks which do not point to the primary vertices. However, for the studies within this thesis, these trigger lines are not useful, as electroweak bosons decay very rapidly⁵. Therefore we will only discuss the muon and electron triggers here.

The muon triggers in HLT1 rely on finding VELO tracks that are muon candidates. The VELO tracks are reconstructed using a Kalman filter based track fit [94, 95]. For every VELO track a search window is implemented in the M3 station by extrapolating the VELO track in a straight line, with the size of the window defined by the expected

⁵The Z boson has a lifetime of about 10^{-25} seconds.

deflection of a 6 GeV muon in the horizontal plane, and by the expected effect of multiple scattering in the vertical plane. Hits in the window are combined with VELO tracks to form candidate tracks, which are then further combined with hits in M2, M4 and M5. A muon candidate is accepted if it has a hit in M3 matching the VELO track and if the χ^2 fit of the candidate track (containing the VELO track and muons hits) per degree of freedom is less than 25 [91, 92]. Track segments in the OT and IT are also reconstructed. Requiring high momentum tracks reduces the size of the search windows that need to be considered, increasing reconstruction speed. The track momenta are determined using a Kalman filter [94, 95] based track fit. Individual trigger lines within HLT1 then allow for specific selection of events [91, 92]. The typical rates of such lines are shown in Table 3.3.

The electron triggers in HLT1 are based on the Level-0 trigger candidates [91, 92]. The reconstructed Level-0 electrons are re-analysed, and matching tracks are sought in the tracking stations. The search window is determined based on the position and energy of the ECAL cluster. If a successful search occurs, then a track fit is performed incorporating VELO information. The tracks coming from this search are then also reconstructed using the Kalman filtering process [94, 95]. Individual lines using these electron candidates then allow for the selection of events passing specific criteria. The typical rate of the line designed to select events containing high p_T electron candidates is shown in Table 3.3.

Once the HLT1 has reduced the trigger rate to about 50 kHz, the HLT2 is able to select events based on a complete event reconstruction. This reconstruction is very similar to the full ‘off-line’ reconstruction (discussed in the next section). The main differences are that a simplified geometry is used for track fits and that the RICH information is only provided on-demand for a small subset of events [91, 92].

The HLT2 makes use of forward tracking algorithms for all VELO tracks. This technique is also used in the off-line event reconstruction. Other techniques are also only used offline; the tracking efficiency in the trigger is lower than in the off-line reconstruction by about 1 – 2% per track. To save CPU time, reconstructed tracks in the HLT2 must satisfy $p_T > 0.5$ GeV and $p > 5$ GeV [91, 92].

The trigger lines in HLT2 are less generic than those in HLT1, and select events specifically for subsets of physics studies. The selection requirements and typical rates for lines used for the study of electroweak production physics are summarised in Table 3.4.

Trigger	Selection requirements	Fractional Rate (%)
HLT1SingleMuonHighPT	• $p_T > 4.8$ GeV, • $p > 8$ GeV, • $\chi^2_{\text{track}}/\text{dof} < 4$	1.4
HLT1DiMuonHighMass	• $p_T > 0.5$ GeV, • $p > 6$ GeV, • $\chi^2_{\text{track}}/\text{dof} < 4$, • $M(\mu\mu) > 2.7$ GeV	3.1
HLT1SingleElectronNoIP	• $p_T > 10$ GeV, • $\chi^2_{\text{track}}/\text{dof} < 4$	1.4

Table 3.3: Typical HLT1 trigger lines for electroweak production physics studies at LHCb are shown, along with the most important selection requirements and their typical rate (as a percentage of the total HLT1 rate), though the precise rate depends upon the instantaneous luminosity. The numbers shown are for 2012 data.

The LHC machine only delivers stable colliding beams for about 30% of the year [92]. In order to use the idle time well at the EFF, about 20% of Level-0 events are stored locally on hard discs to be processed in the time between the fills. This increases the number of events available for analysis. This technique was introduced in 2012 running, and, along with other improvements, allowed the output of the HLT to be increased from 3 kHz in 2011 running to 5 kHz [92].

The efficiency and performance of the various triggers depend upon the process being studied, and will be discussed in later chapters. However, it is useful to introduce now the notation used within the experiment to label trigger decisions:

- “Decision” refers to whether or not a particular trigger line was passed (‘fired’) by the event in question.
- “TOS (Trigger On Signal)” refers to whether a specific candidate object (or collection of objects) fired a particular trigger. It is the response to the question: “would this candidate fire the trigger if present on its own?”.

Trigger	Selection requirements	Rate / Hz
Hlt2SingleMuonHighPT	• $p_T > 10$ GeV	45
Hlt2SingleMuonLowPT	• $p_T > 5$ GeV • Prescale: 0.002	5
Hlt2DiMuonZ	• $M(\mu\mu) > 40$ GeV	0.05
Hlt2SingleElectronTFVHighPt	• $p_T > 15$ GeV	25
Hlt2SingleElectronTFHighPt	• $p_T > 10$ GeV • Prescale: 0.01	1

Table 3.4: Typical HLT2 trigger lines for electroweak production physics studies at LHCb are shown, along with the most important selection requirements, beyond the standard ones outlined in the text, and their typical rate (the precise rate depends upon the instantaneous luminosity). Some lines are ‘prescaled’ to only be run on a small fraction of events, as their rate would otherwise be too high. The numbers shown are for 2012 data. In addition to the lines shown there are other trigger lines which extend the dimuon invariant mass range down to $M(\mu\mu) > 2.5$ GeV, for the study of the Drell-Yan process at lower invariant masses. Including these lines, and the lines for the study of CEP, the triggers used to study electroweak production processes at LHCb take a rate of between 100 and 150 Hz.

- “TIS (Trigger Independent of Signal)” refers to whether the event would still be triggered if the candidate object(s) were removed from the event. This can be defined with respect to a specific trigger, or to a collection of triggers.

3.2.9 Detector readout and management

The Data Acquisition (DAQ) system [93, 96] is responsible for transporting the data taken in a given bunch crossing to permanent storage. This interfaces with the detector, the EFF and the HLT, before transferring the selected events for permanent storage to CERN. There is a 40 Tb storage system which offers buffer capacity in case the link to CERN is interrupted.

The Timing and Fast Control (TFC) system drives all stages of data readout at the detector. It distributes the beam-synchronous clock, the trigger reset and fast control commands simultaneously across the detector. The Experimental Control System (ECS) is responsible for monitoring the state of the entire detector. It monitors the voltage supply to the different components of the detector, the temperature changes in the detector, gas flows and pressure, and is also responsible for monitoring the trigger, TFC and DAQ systems [80].

3.2.10 Recorded integrated luminosity and running conditions

The LHC and the LHCb detector have been operational and collecting data in 2009, 2010, 2011 and 2012. In 2010, 2011 and 2012 the LHCb experiment collected about $\int \mathcal{L} dt = 37 \text{ pb}^{-1}$, $\int \mathcal{L} dt = 1.1 \text{ fb}^{-1}$ and $\int \mathcal{L} dt = 2.1 \text{ fb}^{-1}$ of data respectively [97]. The accumulated luminosity throughout the calendar years is shown in Figure 3.23. The efficiency of the LHCb detector is shown in Figure 3.24: the detector operated at $\sim 93\%$ efficiency, with the dominant inefficiencies being due to the DAQ system and readout dead time.

The average number of proton-proton interactions in a bunch crossing is labelled ν . This is related to the average number of visible proton-proton interactions in bunch crossings, μ . The two are different because the calculation of μ excludes bunch crossings where the proton-proton interaction is not visible to the detector (typically because the interaction is either elastic or diffractive with no decay products in the LHCb acceptance) [98]. The number of interactions, n , in a given bunch crossing is governed by

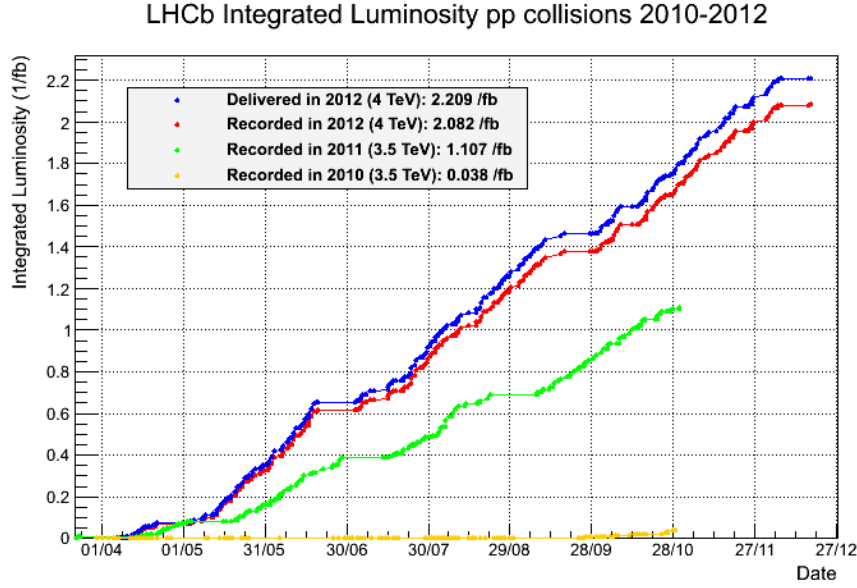


Figure 3.23: The total integrated luminosity delivered to and recorded by the LHCb detector in 2010, 2011 and 2012.

a Poisson distribution, with probability $P(n|\mu)$ given by

$$P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}. \quad (3.1)$$

The average number of observable proton-proton interactions is also measured (the ‘pile-up’). This is different to the average value of μ because bunch crossings where there are no proton-proton interactions are not observed. This means that the average pile-up is the zero suppressed mean of the distribution of the number of visible proton-proton interactions [98]. The average μ and pile-up are shown in Figure 3.25 and Figure 3.26 respectively.

At ATLAS and CMS, a decreasing instantaneous luminosity is seen throughout a fill due to depleting proton beams. However, at LHCb, the luminosity levelling changes to ensure that the instantaneous luminosity, and therefore μ and ν , are roughly flat within each fill.

The average number of visible proton-proton interactions in each bunch crossing, μ , is proportional to the luminosity. Any stable measure of the interaction rate can therefore be used as a relative luminosity monitor. In practice the number of hits in the Pile-Up veto stations in the VELO and the average number of VELO tracks provide the most robust estimates of the relative luminosity. The μ values estimated using these

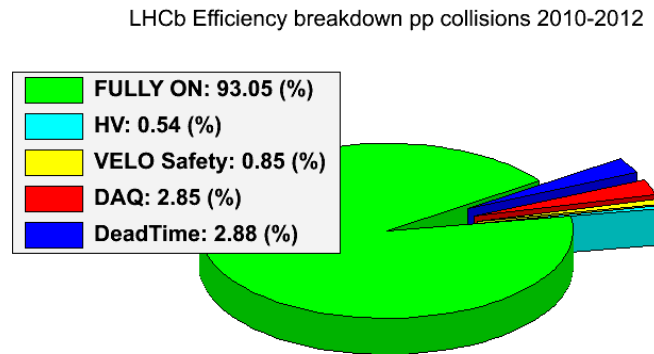


Figure 3.24: The efficiency with which the LHCb detector was able to collect the luminosity provided by the LHC machine, and the cause of inefficiencies in data collection. Source: The LHCb collaboration

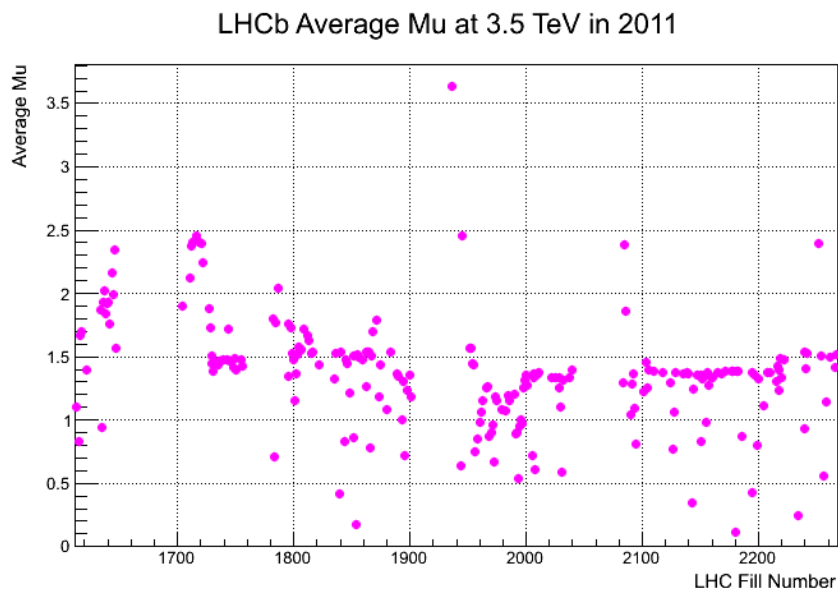


Figure 3.25: The average value of μ during 2011 fills at LHCb. Source: The LHCb collaboration

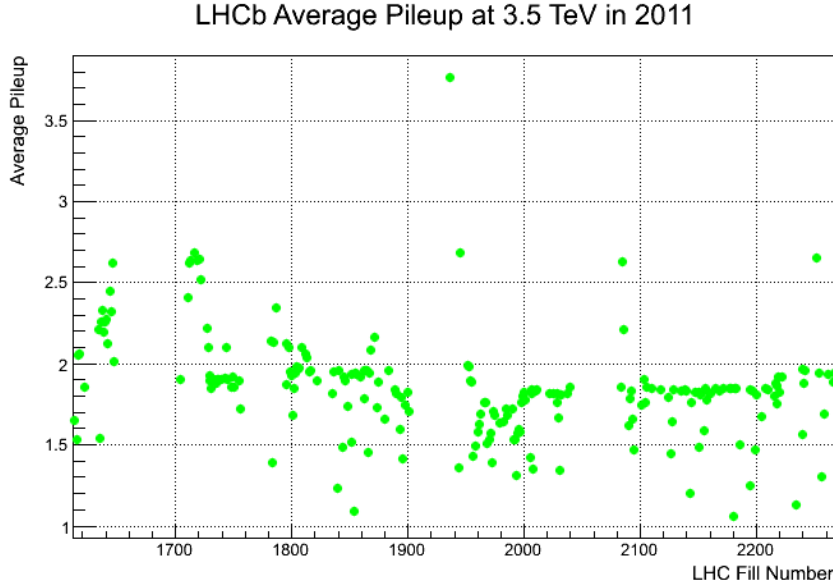


Figure 3.26: The average pile-up in bunch crossings at LHCb in 2011 fills. Source: The LHCb collaboration

two subsystems are shown in 2010 data in Figure 3.27. The agreement is stable at the $\sim 1\%$ level, with the offset from unity due to the difference acceptances of the different sub-systems [97].

The absolute luminosity is determined for short dedicated periods of data-taking. This can then be extrapolated to find the absolute luminosity at any period using the monitors discussed above. In order to calibrate this luminosity scale, the measurements of the absolute luminosity seek to determine the beam densities as a function of position. Knowledge of this beam profile is needed to find the absolute luminosity since the average instantaneous luminosity is given by⁶

$$\mathcal{L} = 2N_1N_2f \int \rho_1(x, y, z, t)\rho_2(x, y, z, t)dx dy dz dt, \quad (3.2)$$

where there are $N_{1,2}$ protons in each bunch in beams 1 and 2, f is the beam-crossing frequency, and the density of each bunch is given by $\rho_{1,2}$ [97]. In order to calculate the absolute luminosity the number of protons in each bunch therefore also needs to be known. This is found by measuring the induced current in beam current transformers placed close to each beam [97].

⁶This equation only holds in the relativistic limit and when the beam crossing angle is zero.

Measurements of the particle densities in each beam are performed using two methods: Van der Meer scans and beam-gas imaging [97]. The Van der Meer scan method is based on separating the two beams and measuring the change in interaction rate as a function of their displacement. The beam-gas imaging method relies on the interactions of each beam with the residual gas (mainly hydrogen, carbon and oxygen) in the LHC machine. This residual gas is uniformly located throughout the beampipe (this has been confirmed at the level of 0.05%). The beam-gas interactions give rise to tracks in the LHCb detector which allow the vertices of the beam-gas interactions to be reconstructed. The location and density of these vertices enable the beam angles, profiles and relative positions to be found. These two procedures to measure the beam densities have different advantages. The Van der Meer scan has a high rate of particle interactions with respect to beam-gas imaging. However, beam-gas imaging provides a non-invasive approach, where the beam positions are stable and so effects associated with the displacement of the beams are small. The two approaches have uncorrelated systematic uncertainties [97].

The measurement of the number of particles in each beam provides the largest systematic uncertainty in the luminosity determination. The overall luminosity uncertainty is currently 3.5 % in data recorded in 2010 and 2011 [97], and 1.2 % in data recorded in 2012⁷ [99, 100].

3.3 Event Reconstruction and the LHCb software environment

In this section we discuss the reconstruction of events at LHCb using the information from the detector subsystems discussed previously. We also briefly discuss the rest of the LHCb computing infrastructure.

3.3.1 Track reconstruction at LHCb

Track reconstruction is performed using seed hits in the VELO or the T-stations, adding in information from the other tracking stations with forward/backward projection and seeking hits in various search windows. Track segments in the different sections of detector are then combined if a good match is found. All potential tracks are fitted with

⁷This uncertainty was about 5% until recently, and a paper is in preparation from the LHCb collaboration documenting this new result [99].

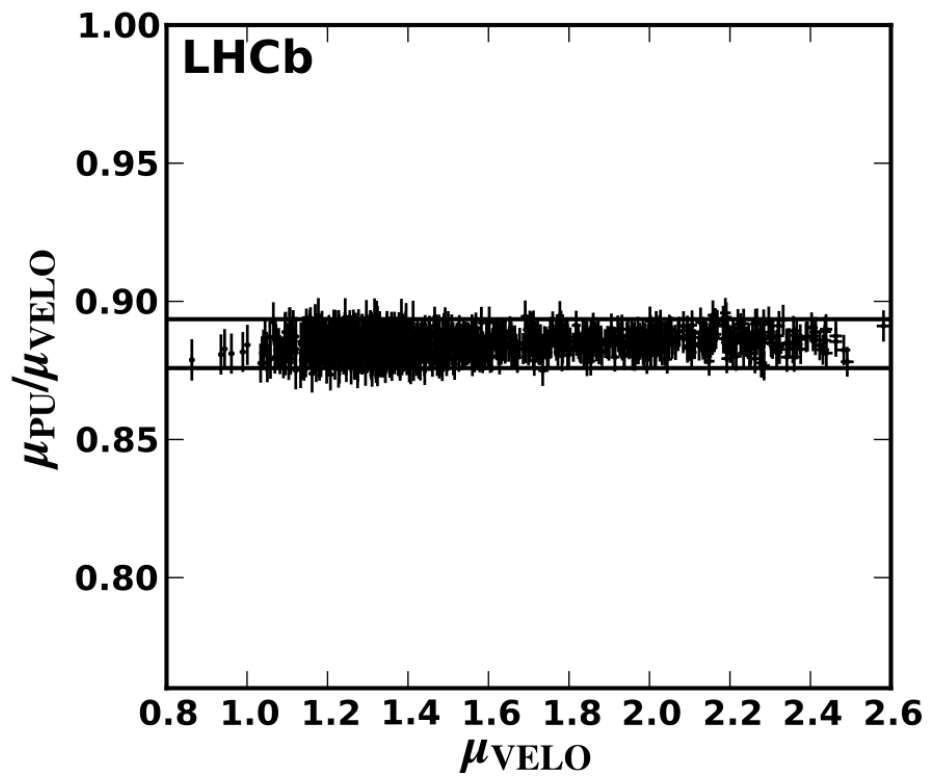


Figure 3.27: Estimates of μ from different detector sub-systems in 2010 data, which enables the relative luminosity to be found. From Ref. [97].

a Kalman filter [94, 95], extracting the track information and determining the probability that the hits do correspond to a particle traversing the detector. This fit accounts for multiple scattering and energy loss within the detector. Finally, the overall set of tracks found is considered. Some tracks within this set may be double-counted ‘clones’. A clone killing algorithm, based on the Kullback-Liebler distance, is run [101, 102]. This compares the tracks, and if any two tracks share many hits or if the tracks parameters contain sufficiently similar information then the track with the most hits is kept. The track momentum resolution at LHCb varies for tracks from 0.4% for tracks with momentum of 5 GeV to 0.6% at 100 GeV. In order to select high quality tracks the standard LHCb reconstruction algorithms require a ‘good’ χ^2/dof associated with the track fit: for data collected in 2011(2012) tracks are only recorded if they have $\chi^2/\text{dof} < 5(3)$ [103]. The detector alignment was better understood in the 2012 data-taking period than the 2011 data-taking period, so the χ^2 values associated with track fits in 2012 data tend to be smaller than those in 2011 data, accounting for the different requirements in the different periods [103]. The tracking efficiency for these requirements is typically high ($> 90\%$) [104]. Additional track selection requirements can be used to provide a cleaner selection of candidate events.

LHCb considers different classes of tracks. These are:

- long tracks, which correspond to tracks which contain information from both the VELO and the T1-3 stations.
- downstream tracks, which correspond to tracks with information in the TT and T1-3 stations, but no information in the VELO. These tend to correspond to tracks formed from the decay of long-lived (neutral) particles (for example, the K_S^0).
- upstream tracks, which correspond to tracks with information only in the VELO and the TT. These tracks tend to correspond to low momentum particles which are swept out of the LHCb acceptance by the magnet.
- VELO tracks, which correspond to tracks which only contain VELO information.
- T tracks, which correspond to tracks which only contain information from the T1-3 stations.

3.3.2 Particle Identification at LHCb

This thesis contains studies of muon decays of vector bosons and is therefore reliant on the ability of the LHCb to reconstruct muons. Within the offline analysis of data at

LHCb there is a standard Boolean variable which can be used to select a set of tracks containing a high purity of muons with low contamination. This variable is labelled `ISMUON`. This Boolean variable is true if a track lies sufficiently close to hits in the muon chambers at LHCb (a search window is created within each muon station based on the track kinematics), and is otherwise false. If a particle has momentum above 10 GeV, then hits must be found in M2, M3, M4 and M5 in order to satisfy the `ISMUON` criterion. The performance of the muon systems (both within triggering and for particle identification) will be discussed later when discussing the efficiency on the analyses considered.

In addition, it is also possible to calculate the relative likelihood that a track corresponds to a particular particle species. This is performed using information from the all the particle identification sub-detectors - the muon chambers, the calorimeters and RICH 1 and RICH 2 all contribute [88, 90, 105]. The set of likelihoods for different hypotheses of the particle species then form the basis for particle identification requirements at LHCb [106].

Alternatively, a neural network based approach can be employed, using information from all LHCb detector subsystems to provide the probability that a track corresponds to a particular particle species [106, 107].

3.3.3 The LHCb software environment

The computing environment at LHCb uses a software framework and architecture called GAUDI [108]. This framework extends throughout the LHCb experiment. The software trigger is implemented in a package, MOORE (discussed earlier) [91]. This stage is followed by the full event reconstruction, managed by BRUNEL [109]. BRUNEL implements both the track reconstruction algorithms and the particle identification based on the RICH, muon and calorimeter sub-detectors. Detailed physics studies and analyses of detector performance take place in a package named DAVINCI. Dedicated algorithms can be created and used here to search for particular types of events, or to create event wide information (for example jet reconstruction). Much of this computing at LHCb takes place on the worldwide computing grid. This is managed using DIRAC [110], with the interface with the LHCb software being controlled by GANGA [111]. There are also many other, smaller, packages. Two such examples are EOLAS [112] and PANORAMIX [113], which both enable events at LHCb to be visualised.

3.3.4 Simulating proton-proton interactions at LHCb

Simulation of events at LHCb occurs in three stages. These are discussed briefly below.

Event simulation is performed in the GAUSS [114, 115] package. The simulation used for the hard process is PYTHIA 6 [12]⁸, interfaced with PHOTOS [74] for the inclusion of QED FSR. The decay of final state hadrons is managed by EVTGEN [116, 117]. The EVTGEN package was specifically built by the BaBar collaboration [118] to handle heavy flavour decays.

The detector geometry and material content are approximated within GEANT4 [77, 78, 114]. This detector simulation is then implemented in GAUSS [114, 115]. The GAUSS framework also handles the particles produced in the primary physics process [115], and the production of secondary particles in interactions with the detector.

Finally, the detector digitisation is simulated within BOOLE [119]. This package models the response of each sub-detector to the particles passing through it, and simulates the readout of the detector (and the Level-0 trigger). Effects such as noise and spillover (where information from previous bunch crossings remain within the detector electronics) are simulated at this stage. After the simulation within BOOLE the simulated data are treated in the same manner as the data recorded by the LHCb detector.

3.4 Offline Data Storage

All events that are selected by the software trigger are stored permanently at CERN on tape. However, for ongoing usage the LHCb data are also stored on disc at CERN and at other Tier One sites across the globe [120]. This allows easy access to the data.

There is not enough disc space to store every event readout by the trigger, so pre-selections are run in advance of the data being uploaded to the Tier One storage sites. This process is known as ‘stripping’ the data. Stripping lines select individual events that will be of interest for studies of physics processes and detector performance. The requirements placed by these lines must be loose enough so that all desired events are selected (as re-strips of the data are rare) but tight enough to fit in the available disc space. Lines which have a large overlap of events are collected together to form streams.

⁸Simulation using PYTHIA 8 is now starting. However, differences between the two programs are typically small.

Line	Selection requirements	Retention Rate (%)
WMuLine	Selects events containing a single muon with $p_T^\mu > 15$ GeV	0.257
Z02MuMuLine	Selects events containing two oppositely charged muons with $p_T^\mu > 15$ GeV and a combined invariant mass greater than 40 GeV	0.001
MuMuSSLIne4	Selects events containing two muons of the same charge with a combined invariant mass greater than 20 GeV	0.002

Table 3.5: Three typical stripping lines in the EW stream, and the percentage of triggered events that they accept.

Each stream is considered in turn, and all the events selected by the lines in a given stream will be written to the same files. By placing lines with a large overlap of events in the same stream the number of files containing the same event is reduced. This enables disc space to be used more effectively. All events considered in this thesis are selected by the electroweak (EW) stream, which is used by analyses within the LHCb QCD, Electroweak and Exotica (QEE) working group. The overall EW stream contains nearly 60 stripping lines, and retains about 2.1 % of all events selected by the trigger. Three typical stripping lines are shown in Table 3.5.

Chapter 4

Jet Reconstruction at LHCb

“It is sometimes said that scientists are unromantic, that their passion to figure out robs the world of beauty and mystery... It does no harm to the romance of the sunset to know a little bit about it.”

— Carl Sagan

In this chapter we discuss jet reconstruction at LHCb. We start by considering the particle flow algorithm we use to select inputs for jet reconstruction, and then discuss the jet reconstruction process. We also detail the selection requirements we make to remove poorly reconstructed jets. We discuss the jet energy scale, and the overall performance of jet reconstruction at LHCb. As is usual for studies of jet reconstruction at hadron colliders, this discussion is based on simulated data. Measurements of jets at LHCb are therefore strongly reliant on simulation. However, when we take a quantity from simulation, we cross-check it against data to set a systematic uncertainty. This will be discussed in subsequent chapters, as this is specific to the analysis discussed later. The work described in this chapter is based on the work of the author and others in Refs. [4, 121–124]. The author’s work focused mainly on creating the initial framework for jet reconstruction at LHCb (which was then further developed by others), determining the performance of jet reconstruction in simulation, and comparing this performance with that achieved in data.

Jet physics in the forward region probes a different kinematic region to that at the general purpose detectors. A jet with total momentum of 400 GeV in the forward region will have p_T of around 100 GeV for $\eta \sim 2$, and about 40 GeV for $\eta \sim 3$. A hard particle produced with p_T of 400 GeV and with $\eta \sim 2$ would have total momentum in excess of

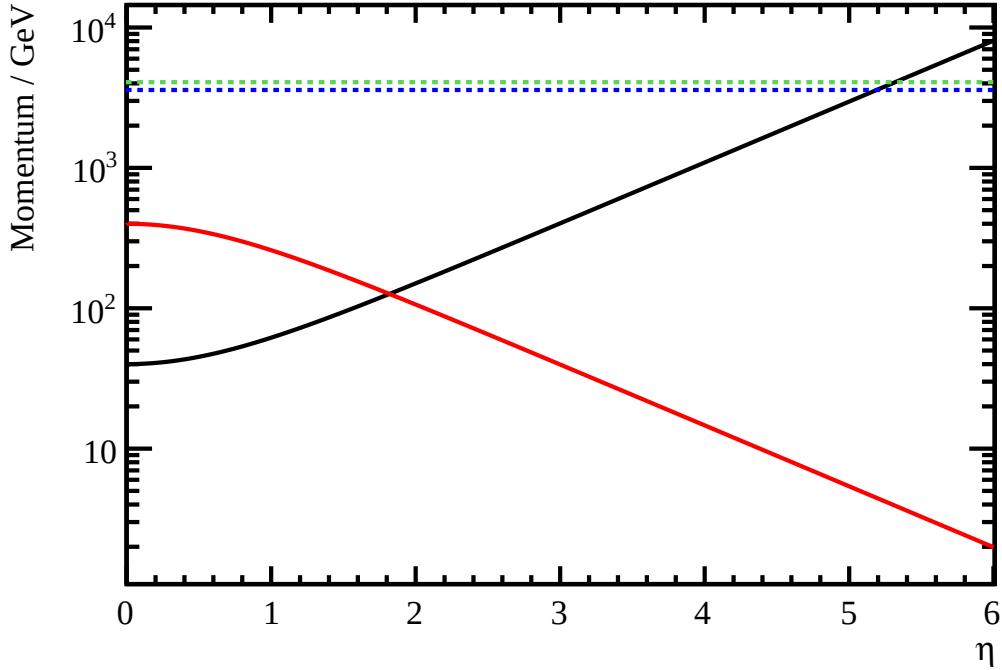


Figure 4.1: The relationships between the pseudorapidity, momentum and transverse momentum of jets. The black curve shows the total momentum of a jet with p_T equal to 40 GeV. The red curve shows the p_T of a jet with a total momentum equal to 400 GeV. Also shown as dashed lines are the LHC kinematic limits: jets cannot have total momentum greater than 3.5 TeV (blue) or 4 TeV (green) in 2011 and 2012 running respectively.

1 TeV. These relations are shown in more detail in Figure 4.1. Consequently LHCb is sensitive to jets in a different region to that of ATLAS and CMS: the majority of jets at LHCb have comparatively low transverse momenta; relatively few events contain jets with p_T above 100 GeV. This chapter therefore concentrates on jets with p_T in the region between roughly 10 GeV and 100 GeV.

4.1 The particle flow algorithm

Jets are reconstructed with the standard particle flow algorithm used by LHCb [4, 121, 125]. This algorithm is designed to use information from throughout the LHCb detector to provide a list of input particles that are subsequently reconstructed into jets. Jet reconstruction algorithms at some other detectors simply use information from calorimeters, running the jet clustering algorithms on a list of calorimeter-based inputs. Such an approach is not feasible at LHCb as the calorimeter cells are coarse and saturate at

high energies. However, the tracking systems at LHCb (discussed in Chapter 3) enable accurate energy resolution of charged particles. The particle flow approach uses different detector subsystems to measure different particles, improving the jet energy resolution from that which would be achieved from a ‘simple’ calorimeter based approach. The particle flow algorithm is tuned to optimise the jet energy resolution, using information from tracks where possible, and only uses calorimeter information where no track information is available (for example, for neutron and photon contributions).

An advantage that the LHCb experiment has over the other experiments at the LHC is that collisions at LHCb take place in a low pile-up environment: collisions at LHCb typically contain only one or two reconstructed primary vertices. Consequently the effects of particles produced in other proton-proton interactions are lessened. In addition, because of the excellent impact parameter resolution in the VELO, tracks which originate in the VELO can have a primary vertex requirement placed on them, further reducing the effect of pile-up interactions. However, there is no way to associate calorimeter clusters accurately with an individual vertex, so all calorimeter clusters are considered.

The particle flow algorithm starts by selecting all long¹ tracks from the primary vertex of interest. All downstream tracks are also selected. The particle identification subsystems are used to determine the most likely species of each track, and the corresponding particle mass is used to set the particle four-momentum. No additional track quality requirements are placed on the tracks beyond the standard LHCb track quality requirements discussed in Section 3.3.1. Upstream tracks are also considered for their primary vertex information; the momentum resolution of these tracks is very poor, but the momentum of these tracks is also typically comparatively small, as they are bent out of the detector coverage by the magnetic field. These particles are therefore simply assigned zero momentum within the particle flow algorithm. As the k_T algorithms used to reconstruct jets are infra-red safe, the presence of zero momentum particles in the list of inputs does not change the number of jets or the jet energy [52]. It does however increase the number of particles in the jet associated with the primary vertex. This is a useful quantity that will be discussed later, as it is a variable that can be used to reject ‘fake’ jets. If a group of tracks are consistent with a neutral resonance (a V0 - typically a K_s or Λ - or a B or a D meson) then the tracks are removed from the list of inputs (if they were initially included) and instead the resonance is used. The advantage of this approach is that the resonant particle will tend to ‘point’ back to the proton-proton interaction whereas the tracks may have a

¹The LHCb nomenclature for track types was discussed in Section 3.3.1.

large impact parameter with respect to the vertex. Where it is possible to reconstruct multiple V0s from the same tracks, the V0 with the best vertex χ^2 is used.

In addition to track-based inputs, the particle flow algorithm considers reconstructed neutral particles as inputs for jet reconstruction. Isolated clusters in the electromagnetic and hadronic calorimeters are associated with neutral electromagnetic and hadronic particles respectively. The clusters are considered to be isolated if they are sufficiently far from tracks. This is defined in terms of the χ^2 separation of the cluster from nearby tracks. For clusters in the ECAL, if the χ^2 separation from every track is greater than 25 then the cluster is considered isolated. For clusters in the HCAL, then the required separation for a cluster to be considered to be isolated is chosen to vary with the cluster energy. For cluster transverse energies below 5 GeV the separation must be greater than 25, whereas for transverse energies above 5 GeV then the cluster separation must be greater than 16. HCAL clusters are only considered if they have transverse energy greater than 2 GeV.

However, a jet contains many particles produced at small angles with respect to each other. Simply taking isolated neutral particles will underestimate the number of neutral particles in a jet, as some will fall close to tracks. The main advantage of the particle flow approach is that it enables the recovery of this neutral content. For each track in the event, the expected calorimeter deposits in both the ECAL and HCAL are found [126] as a function of the track momentum and the best hypothesis of the track species. In other words, the algorithm expects electrons to leave different deposits in the calorimeters to charged hadrons. This procedure is discussed in more detail in Ref. [126], but we provide an overview here. The expected energy deposits are measured by selecting isolated tracks which are well associated with calorimeter clusters, an example of which is shown in Figure 4.2. The expected response is found separately for the different regions within the electromagnetic calorimeter system, given the different sizes of the calorimeter cells. Small differences in the calorimeter response are seen between data and simulation, so different expected energies are calculated for data and simulation. No significant differences are seen between the calorimeter response for collisions at $\sqrt{s} = 7$ TeV and collisions at $\sqrt{s} = 8$ TeV. If a track leaves deposits in both the ECAL and the HCAL then the deposits are combined to form an overall calorimeter energy deposit, as the energy deposited in the HCAL and the ECAL tends to be anti-correlated.

Within the particle flow algorithm the total expected deposit from charged particles (tracks) in each calorimeter cell is then compared to the measured non-isolated deposits found in the calorimeters. If a cluster of cells shows an excess over the expected energy

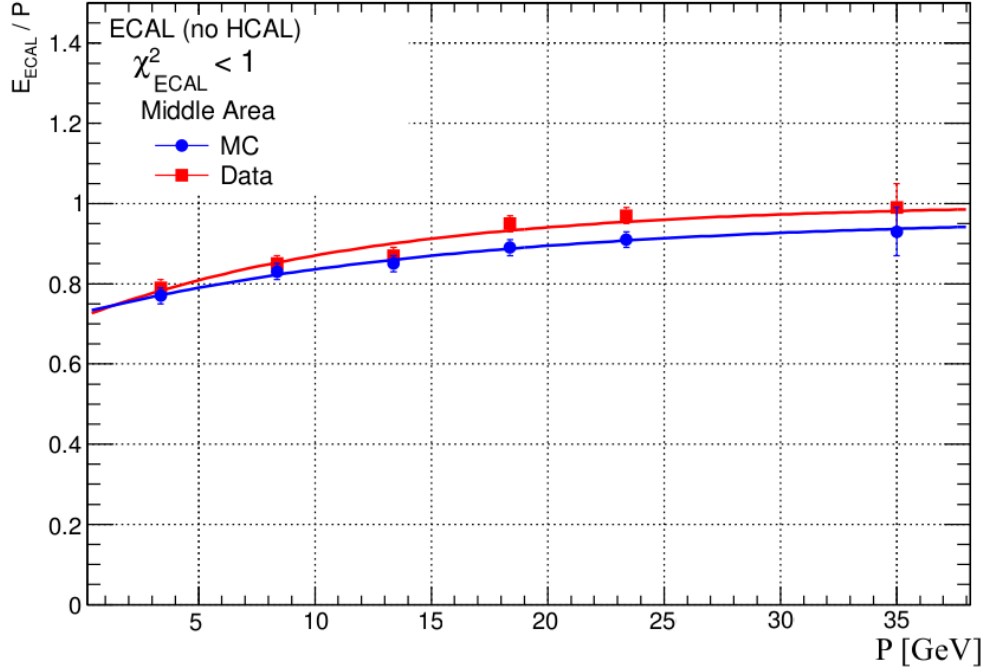


Figure 4.2: Expected energy deposits in the ECAL for electrons for data and simulation where there is little associated HCAL activity. From Ref. [126].

deposit then the excess is hypothesised to be associated with a neutral particle. To be considered the cluster must have excess energy greater than 1 GeV [127]². Finally we note that as calorimeter cells saturate, it becomes harder to recover neutral energy deposits [127].

The particle flow algorithm allows for the exclusion of certain particles. The measurement of jets produced in association with Z bosons is discussed later in this thesis. For this case, and other similar cases, it is important to neglect the tracks associated with the decay particles of the Z boson in the jet reconstruction. This is handled easily by the algorithm, and the candidate tracks corresponding to the Z boson decay are not included in the list of particles to be clustered into jets.

The clustering of the list of input particles is handled using the FastJet [53, 128] interface. Jets are reconstructed using the anti- k_T algorithm [52] with $R = 0.5$. This value of R gives reasonably small differences between the hadron level and the parton level [54], as discussed in Chapter 2, and the jet energy resolution at LHCb shows little

²Different requirements were tested here, and the 1 GeV requirement provided the best improvement in the jet energy resolution.

variation with values of R in the range 0.4 to 0.7 [127]. The chosen value of R is the same as the value used by the CMS experiment, allowing easier comparison with CMS results where applicable. The LHCb software has the ability to reconstruct jets with p_T above 5 GeV, but in this thesis we only consider jets with $p_T > 7.5$ GeV and $2.0 < \eta < 4.5$.

Truth jet reconstruction

It is useful at this stage to introduce the reconstruction of simulated jets at the hadron level using truth information. These jets correspond to the true jets in the simulated event - effectively, the jets which would be reconstructed by a perfect detector. These jets are found by selecting all true stable³ final state particles after the parton shower and hadronisation, and using these as the inputs to be clustered into jets. The same clustering algorithm that is run on the reconstructed particles is run on these inputs.

Jet composition

The fraction of the jet transverse momentum carried by charged particles was determined at LEP [129]: charged particles are expected to constitute roughly 60% of a jet's p_T , with neutral particles contributing the remaining p_T (about 30% of the total p_T is typically associated with neutral electromagnetic particles). Neutrinos contribute less than 1% of a jet's p_T (for light-parton-initiated jets⁴). This behaviour is seen clearly in the LHCb simulation: the fraction of the jet p_T carried by different types of particles is shown at hadron level for light parton initiated jets in Figure 4.3. This can be compared to the fraction carried by the different particle flow inputs in the reconstructed jets in Figure 4.4. Tracks correspond to roughly 60-68% of the reconstructed jet p_T . Of this contribution, resonant neutral particles which decay to tracks correspond to about 1%. There are a few trends in Figure 4.4 that are worth discussing in detail, as they reveal the performance of the particle flow approach:

- The fraction of the reconstructed jet's p_T made up from isolated neutral particles decreases with p_T . At higher p_T jets become more collimated and tend to contain more particles, so fewer neutral particles and clusters are isolated. The fraction of the jet p_T from particles found using the neutral recovery techniques sees a corresponding increase.

³We define stable particles to be those with a mean lifetime greater than 10^{-8} s.

⁴The neutrino content of jets which are initiated by heavy flavour particles tends to be higher.

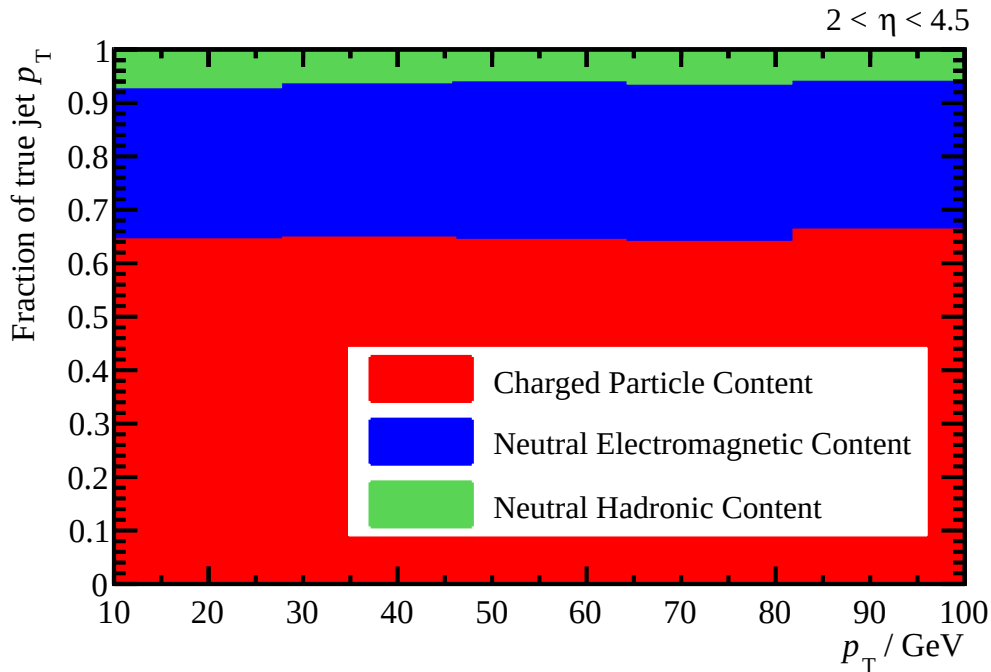


Figure 4.3: The mean fraction of the jet's p_T carried by different types of particles within the jet, at truth level.

- The fraction of the reconstructed jet p_T carried by charged particles at the lowest jet p_T decreases slightly with increasing p_T . This is due to the jet selection requirements (discussed later) biasing this variable.
- Above 30 GeV the fraction of the reconstructed p_T contained in tracks then increases. This is due to some of the jet neutral energy being lost through the saturation of the calorimeter clusters. This neutral energy is not recoverable, and consequently tracks make up a greater proportion of the jet p_T .
- At low p_T , tracks make up about 62% of the reconstructed jet transverse momentum, though simulation predicts that tracks correspond to 65% of the jet p_T at truth level. This discrepancy is due to track detection inefficiencies; track reconstruction at LHCb is about 90-95% efficient. If the track is not reconstructed, then the energy associated with the track will still be measured in the calorimeters and assigned to the jet.

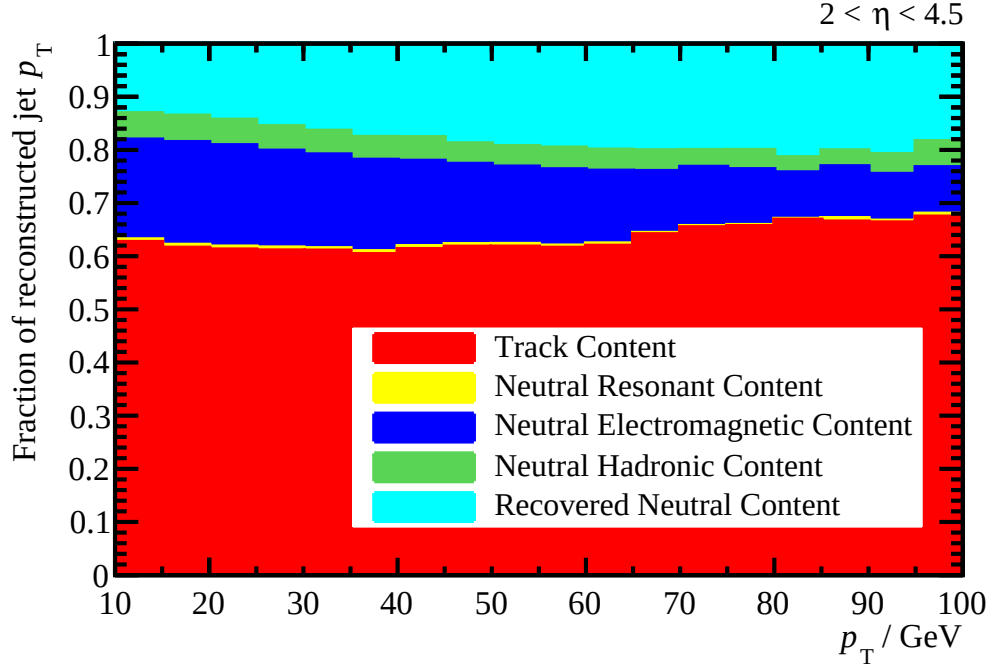


Figure 4.4: The mean fraction of the jet’s p_T carried by different types of particles within the jet in simulation at reconstructed level.

4.2 Jet identification selection

The jet reconstruction efficiency - the probability that a jet is reconstructed - is typically $\mathcal{O}(99\%)$, since most jets contain a large number of particles. However, within the sample of reconstructed jets there will be ‘bad jets’ - ‘fake’ jets made from neutral particles which have been produced in pile-up proton-proton interactions, or made from poorly measured particles. High quality jets are selected using selection requirements which reject the majority of ‘bad jets’. Three variables are used for these requirements⁵:

- n_{Pointing} : the number of particles in the jet which point to the primary vertex (PV) of the jet,
- mtf: the fraction of the jet’s p_T carried by the hardest charged particle in the jet,
- mpt: the p_T carried by the hardest charged particle in the jet.

The distributions of these variables for ‘good’ light quark/gluon initiated jets and fake jets are shown in Figure 4.5. Fake jets tend to only have few particles associated with the primary vertex. A requirement that jets contain at least two particles which point to a

⁵Other variables were studied, but the variables listed here were found to give the best performance [130].

primary vertex removes about 40 % of fake jets whilst rejecting few good jets. In addition fake jets tend to have higher fractions of the jet energy in one, often misreconstructed, particle: this particle can be a charged particle, leading to a large value of mtf and mpt, or a calorimeter cluster, leading to a low mtf and low mpt.

The fake jets dominated by a single calorimeter cluster predominantly have low p_T [130]. An absolute minimum momentum requirement placed on the hardest track in the jet, using the mpt variable, is therefore better at removing these fakes and retaining ‘good’ jets than a relative requirement (using the mtf variable) [130]. However, when the fake jet contains one high momentum track there is no obvious absolute momentum requirement to be made. In this case a fractional requirement (mtf) is chosen.

Based on these studies, and optimisation of the fake/good jet separation [130], the following criteria are therefore used to select candidate jets:

- $n_{\text{Pointing}} \geq 2$: There must be at least two particles in the jet pointing back to the PV,
- $\text{mtf} < 0.75$: No one charged particle carries more than 75% of the jet p_T ,
- $\text{mpt} > 1.8 \text{ GeV}$: The jet must contain a charged particle with $p_T > 1.8 \text{ GeV}$.

These selection requirements were tuned on simulation, with the cut values chosen to remove the majority of bad jets whilst retaining good jets. The fake rate is higher for low p_T jets, and whilst the precise rate depends on the analysis, the fake rate for the Z+jet process is typically around 4% for low p_T jets, decreasing to about 1% for higher p_T jets (with $p_T > 20 \text{ GeV}$). A small number of true jets are rejected by the identification requirements, so that the jet selection efficiency is not 100%. This efficiency is typically between 75% and 96% (with the highest efficiencies achieved at higher p_T), but clearly depends on the process studied. Discussion of the precise values and uncertainties on these values will therefore take place when discussing the specific Z+jet analysis.

4.3 Jet energy scale

Not all the energy associated with a jet in the LHCb acceptance can be reconstructed, owing to calorimeter saturation and other effects. In addition, in events with pile-up interactions calorimeter clusters and downstream tracks which are not associated with the proton-proton interaction of interest will be considered as particle flow inputs and

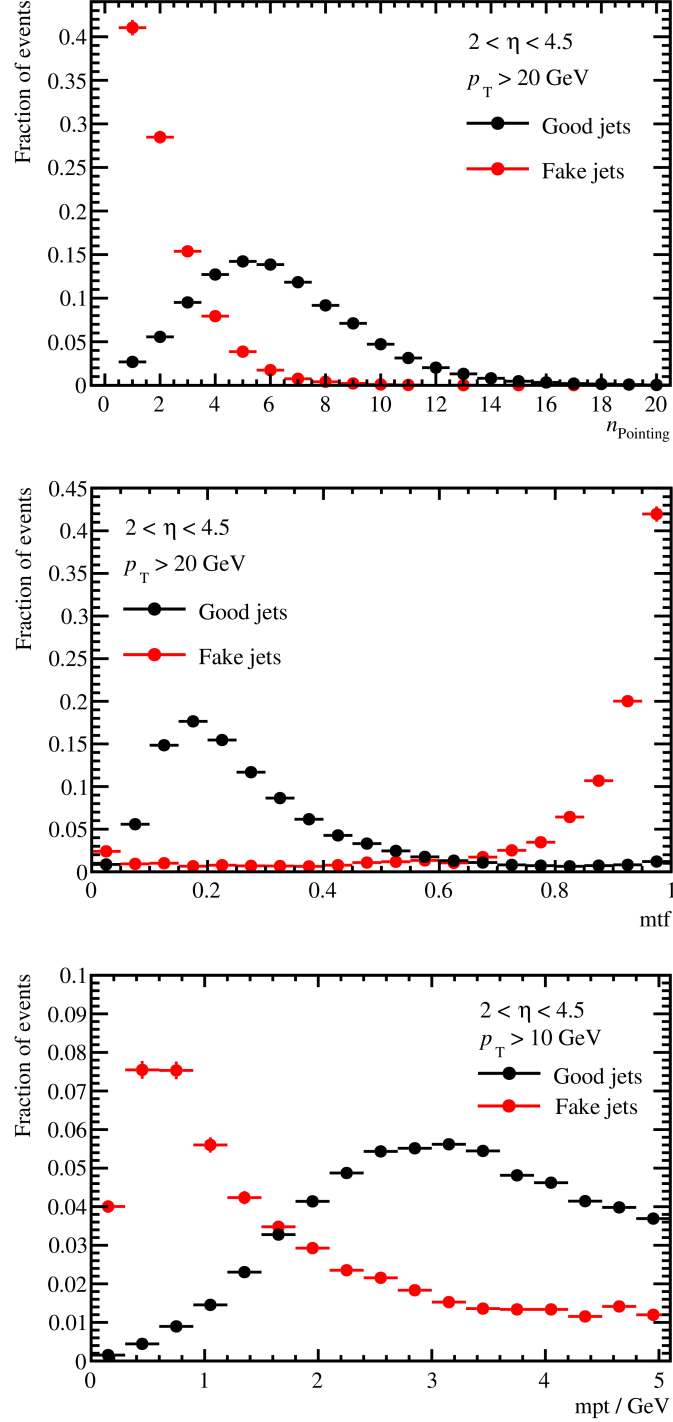


Figure 4.5: The distributions of fake jets (red) and ‘good’ jets (black) in the n_{Pointing} (top), mtf (middle) and mpt (bottom) distributions. The mtf and n_{Pointing} distributions are shown for $p_T > 20$ GeV. The mpt distribution is shown for $p_T > 10$ GeV and is truncated to only show the region with $0 < \text{mpt} < 5$ GeV, as the fake jets removed by the selection requirement on this variable tend to have low p_T . The uncertainties shown are statistical. None of the jet identification selection requirements are placed on these distributions.

clustered into jets incorrectly. We must therefore correct the energy of the clustered jets to obtain the best estimate of the energy of the jet at the hadron level⁶.

We use the standard correction provided by the “LHCb Tools for Jet Reconstruction Working Group” for light quark- or gluon-initiated jets. The correction factor, k , is applied such that the reconstructed 4-momentum $p^\alpha(\text{reco}) = k \cdot p^\alpha(\text{clustered})$, with the value of k defined so that the reconstructed p_T gives the best estimate of the true p_T [122]. The correction factor, k , is seen to depend on three variables: the jet p_T , the jet η , and the number of primary vertices in the interaction [122]. Other variables, like the fraction of the jet’s p_T carried by charged particles, were also considered⁷, however their effect on the correction factor is small ($< 1\%$).

The best estimate of the ratio $\frac{p_T(\text{clustered})}{p_T(\text{true})}$ is found as a function of the variables discussed above, using multivariate regression techniques implemented in the TMVA package [122, 131]. Specifically, neural networks are used to determine this estimate, using the variables listed above as inputs [132]. The inverse of the best estimate of the ratio sets the correction factor. This correction factor is then extracted in bins of the relevant variables (bins of width 1 GeV are used for reconstructed p_T , bins of width 0.1 are used for the reconstructed pseudorapidity, and three bins in the number of primary vertices are used, corresponding to one, two, and three and more primary vertices). Such an approach provides greater statistical precision than directly extracting the correction factor only using events which are in the relevant bins [122].

The correction factor is shown as a function of the jet p_T , η and number of primary vertices in Figure 4.6. The correction factor increases with p_T . This is due to the effect seen in Figure 4.4: the reconstructed charged particle fraction is seen to increase with p_T . The correction factor accounts for the unreconstructed neutral energy. The correction factor is also seen to increase dramatically at low and high η . In these regions the neutral content of jets is underestimated, as this region is close to the edge of the detector geometric acceptance, where the calorimeter efficiency to detect neutral particles is reduced. The correction also changes in the rest of the detector, as the calorimeter cell size varies across the detector. The correction factor also decreases as the number of primary vertices increases. This is because energy from neutral particles produced in

⁶For this thesis we call the raw, unclustered jet p_T the ‘uncorrected p_T ’. The corrected p_T is labelled as the jet p_T . This corresponds to the ‘reconstructed jet’. For simulation, the truth information is labelled as being ‘truth or true level’ (depending on the context).

⁷Jets predominantly consisting of charged particles are expected to be better reconstructed than jets predominantly reconstructed from calorimeter clusters, which are subject to saturation at high energies. However, this has little effect on the correction factor.

other proton-proton collisions will be incorrectly included within jets in events where there are multiple proton-proton interactions. However, this effect is small ($< 0.5\%$). Typically, the factor takes values within 10% of unity, suggesting that the reconstruction is performing well and that the chosen LHCb acceptance ($2 < \eta < 4.5$) is reasonable.

Uncertainties on the jet energy scale

Different sources of uncertainty on the jet energy scale were considered [122]. We summarise these below:

- In order to determine the statistical uncertainty on the evaluation of the jet energy scale the correction was evaluated on different sub-samples. The mean correction in these sub-samples sets the overall correction, whilst the uncertainty on the mean defines the statistical uncertainty on the correction. This uncertainty is typically less than 1% [122].
- The matching of true jets and reconstructed jets was found to contribute a negligible uncertainty: the matching requirements were varied and the change in the jet energy scale was found to be below the level of the statistical uncertainty [122].
- The systematic uncertainty associated with the multivariate method used to find the correction factor was determined by effectively finding the correction factor using an alternative approach. A Gaussian was fitted to the peak of the $(p_{T\text{reco}} - p_{T\text{true}})/p_{T\text{true}}$ distribution in bins of the jet p_T and the number of primary vertices. The difference between the mean of this Gaussian and zero sets the uncertainty on the energy scale from the method. This uncertainty tends to be largest at low p_T [122].
- A closure test is performed by determining the correction factor on events which have already had their p_T corrected. The difference between the new correction factor and unity is assigned as a systematic uncertainty. This uncertainty tends to be largest in regions with a small sample size [122].
- The systematic uncertainty associated with the flavour content of the jet is evaluated by determining the correction factor separately for jets formed by the emission of gluons and jets formed by the emission of light quarks. The difference in response is typically about 1-2%, and is reasonably constant [122].

The correction factor is derived using simulation. Therefore, in addition to these method-based uncertainties, there is an uncertainty associated with the degree to which modelled detector response to jets in simulation is accurate. One commonly used

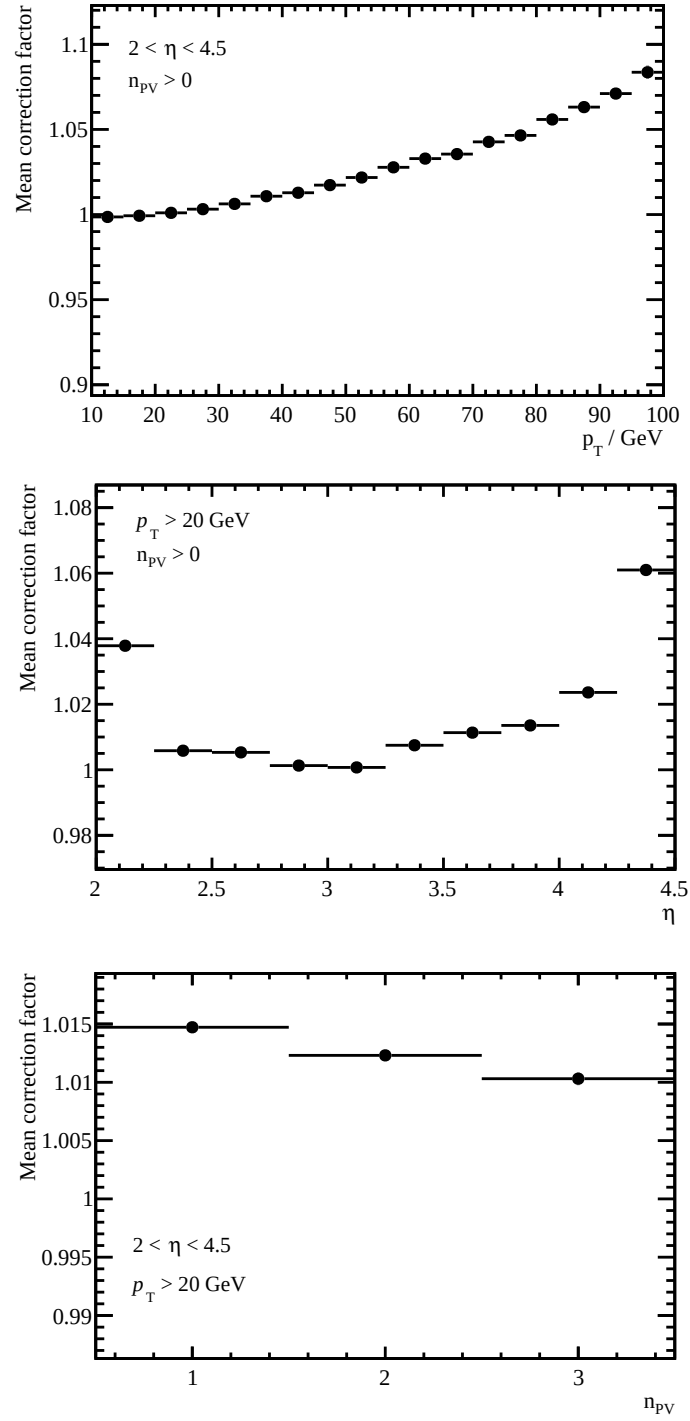


Figure 4.6: The jet p_T correction factor as a function of the jet p_T (top), η (middle) and the number of primary vertices (n_{PV} , bottom).

approach to determine this uncertainty is to extract the correction factor from data by considering events where a photon and a jet are emitted back to back and are expected to have balanced transverse momenta (the ATLAS collaboration used this method and documented it in Ref. [133]). However, such an approach is not feasible at LHCb, due to the poor energy resolution of the ECAL when measuring high energy photons. Instead, back-to-back Z+jet events are used to determine how well the simulation models the detector response to jets. Since the main analysis this thesis discusses is the production of Z+jet events, we will cover this systematic uncertainty later, when we discuss the degree to which Z+jet simulation provides an accurate model for events in data, and the degree to which simulation based assumptions affect the Z+jet analysis.

The method-based uncertainties correspond to an overall uncertainty in the jet p_T that varies between the different bins used to determine the correction factor. This uncertainty is 2% at best, and is only larger than 5% in low statistics regions. Typical contributions to the energy scale uncertainty are shown in Figure 4.7.

4.4 Jet reconstruction performance

We consider here the performance of jet reconstruction at LHCb in the standard LHCb simulation. We will discuss the resolution with which we reconstruct jets, both in terms of their angular variables, and their transverse momentum. The simulation sample used for these studies is a PYTHIA 6 sample of Z+jet events, but other samples give compatible results [127].

In order to consider these resolutions we must match jets at truth level to the corresponding reconstructed jets (if they exist). We consider here two different methods of matching jets between the truth level and the reconstructed level.

1. We match truth jets to the nearest reconstructed jet in $\eta - \phi$ space. We will refer to this as “ R matching”.
2. We match truth jets to the hardest p_T reconstructed jet within a radius of 0.5 in $\eta - \phi$ space. We will refer to this as “ $R - p_T$ matching”.

In the majority of cases the two methods provide the same matching between truth and reconstructed level jets. This is because no two jets (at the same level - either truth or reconstructed) can be closer than the radius parameter used in the anti- k_T algorithm (which in this case is 0.5), as discussed in Section 2.2.3.

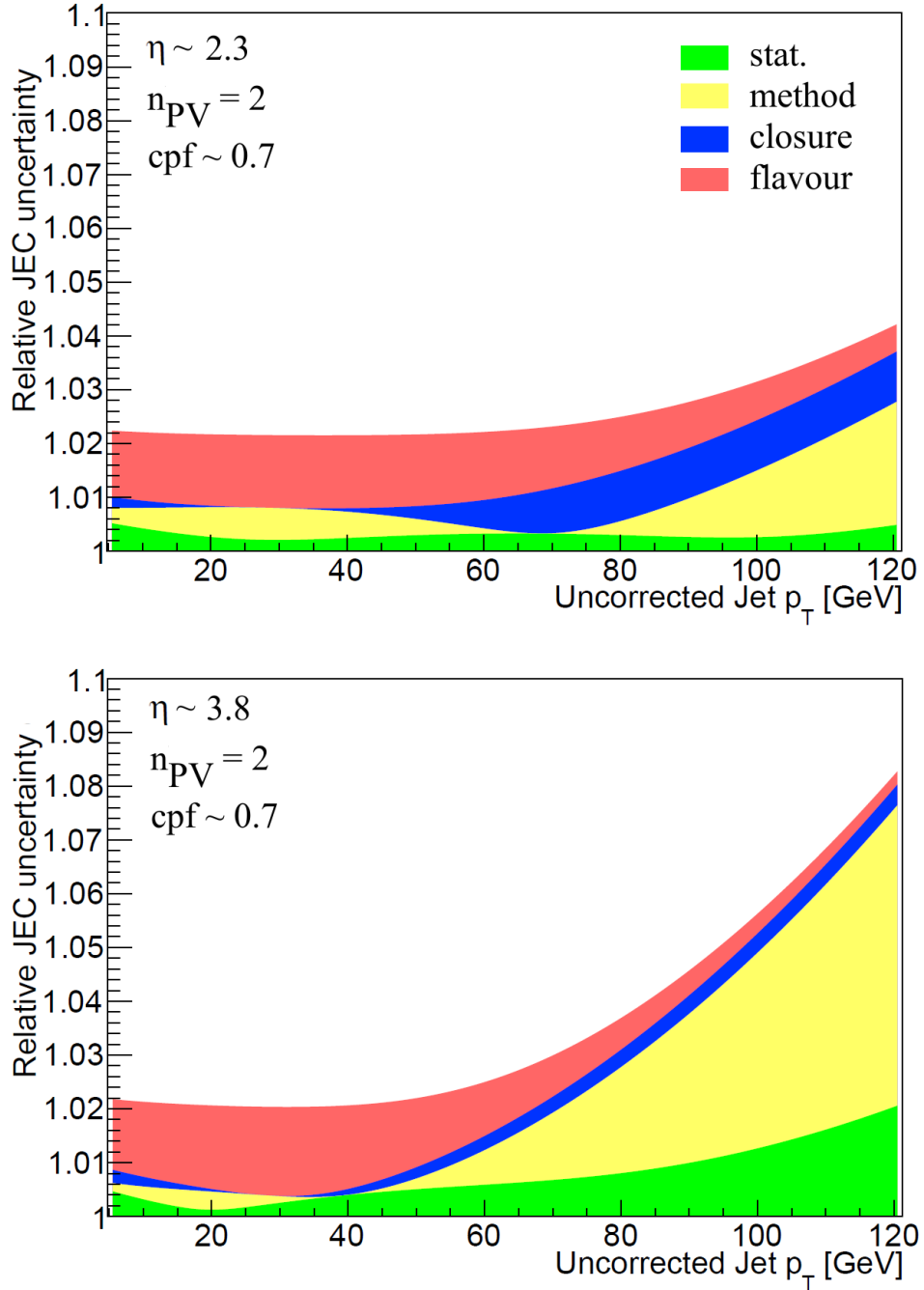


Figure 4.7: Typical values of the uncertainty in the jet energy correction factor, for two different pseudorapidities, for events containing two primary vertices and where the fraction of the jet p_T corresponding to charged particles (cpf) is $\sim 70\%$. Adapted from Ref. [122]

Jet direction resolution

For the determination of the jet direction resolution, the separation of the truth jets from the matched reconstructed jets in $\eta - \phi$ space is studied. This performance is shown in Figure 4.8, along with the cumulative distribution. We define the resolution, $\sigma_{68\%}(\Delta R)$ as the distance in $\eta - \phi$ space such that the probability of getting a value of $\Delta R(\text{Reco}, \text{Truth})$ less than $\sigma_{68\%}(\Delta R)$ is 68%. The distribution of this variable as a function of the jet p_T is shown in Figure 4.9. The resolution improves with increasing jet p_T , as expected. As the jet p_T increases the jet will tend to include more particles. With more particles in the jet, the resolution of the jet direction improves. The resolution shows little dependence on the jet pseudorapidity, even close to the edge of the LHCb acceptance. This means that the effect of jets outside the defined acceptance being reconstructed inside the same acceptance is small.

Transverse momentum resolution

We also measure the accuracy with which LHCb can reconstruct the transverse momentum of jets. This is a key number which will determine the feasibility of various analyses, and influence the choice of bin widths when measuring jet p_T distributions.

The distribution of the ratio of the reconstructed jet p_T to the truth jet p_T is shown in Figure 4.10, for all events in the LHCb acceptance with $p_T > 20$ GeV and for the subset of events which only contain one primary vertex. This subset of events is expected to have the best p_T resolution. Again, the two matching methods provide indistinguishable results.

The ratio shows a clear peak at 1.0, confirming that the jet p_T reconstruction at LHCb provides a good measurement of the jet p_T at hadron level. However, the distribution is clearly non-Gaussian, since the different regions in the distribution of the ratio correspond to different effects. The region below the peak tends to correspond to a particle in the jet not being reconstructed, whereas the region around the peak (and above it) is largely affected by the resolution with which the particles in the jet are reconstructed.

In order to quantify the resolution we take two approaches:

1. We define the quantity $\sigma_{68\%min}$ to be the smallest width in $p_{T\text{reco}}/p_{T\text{true}}$ that contains 68% of the events. We take half of this quantity, $\frac{1}{2}\sigma_{68\%min}$ as a measure of the resolution.

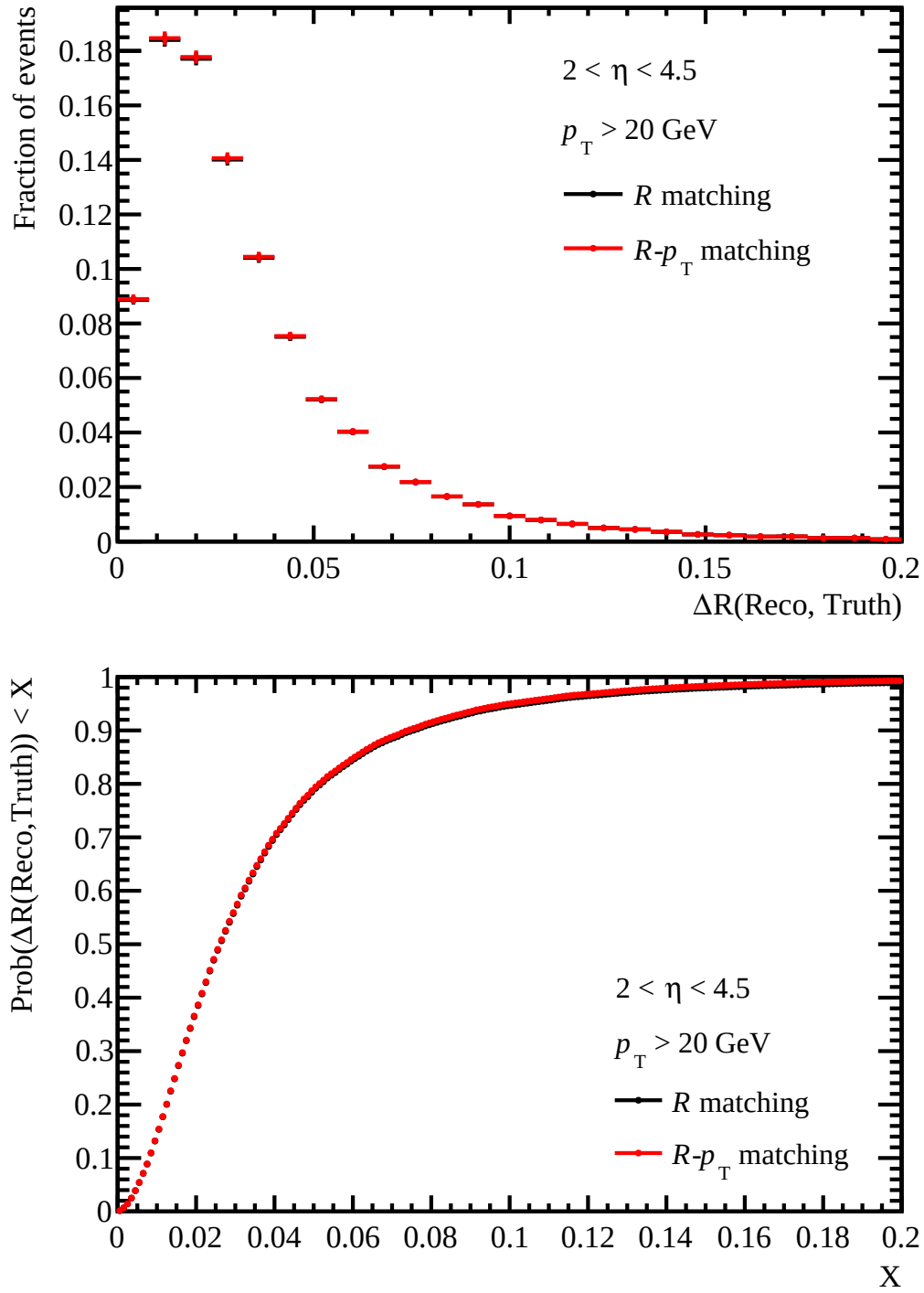


Figure 4.8: The distribution of $\Delta R(\text{truth}, \text{reco})$ (top) and shown cumulatively (bottom). These plots are for jet $p_T > 20$ GeV. The two matching methods give indistinguishable results.

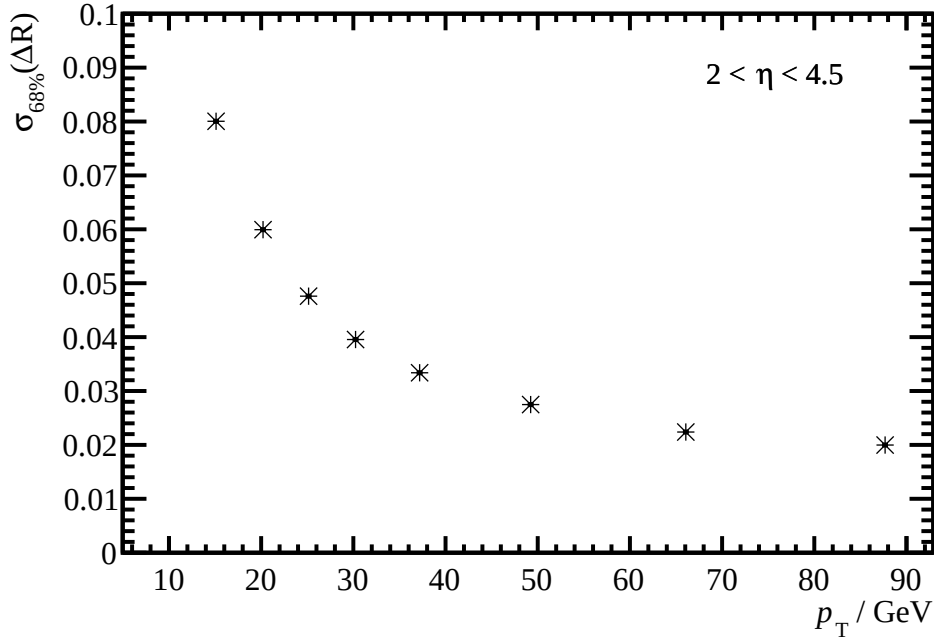


Figure 4.9: The jet direction resolution at LHCb, as a function of the jet p_T . The statistical uncertainties are too small to be visible on this plot.

2. We fit a Gaussian to the peak of the distribution (within the region $0.8 < p_{T\text{reco}}/p_{T\text{true}} < 1.2$), and take the standard deviation of this Gaussian fit to define the resolution.

These two measures of the resolution would be equivalent for a Gaussian-distributed ratio. However, since the distribution is non-Gaussian, the second method returns smaller values for the resolution. A third method was considered: finding half the width between the points in the $p_{T\text{reco}}/p_{T\text{true}}$ distribution where the number of events is $e^{-1/2}$ times the number found at the peak. However, this method gives a larger statistical uncertainty than the two methods used, and tends to give results which are compatible with the Gaussian fit.

The transverse momentum resolution is shown as a function of the number of primary vertices and the jet p_T in Figure 4.11. The jet p_T resolution is typically between 10 and 15%, depending upon the definition used, and is found to have little dependence on the jet η . For the majority of jets with $p_T > 20$ GeV the jet p_T resolution is between 11 and 13%. The values found are consistent with preliminary studies of the p_T balance in di-jet events in data [127].

The resolution is best in events with only one primary vertex, as expected. In events containing more primary vertices, particles produced in other collisions may be incorrectly assigned to the jet, degrading the resolution. However, since the particle flow algorithm uses track-based information to reconstruct the majority of the jet p_T , and tracks are assigned to primary vertices, the effect of the presence of additional primary vertices is small (about 1%), as shown in Figure 4.11. At low p_T the jet transverse momentum resolution improves as the jet p_T increases, since there are more particles in jets as the p_T increases, so the effect of a mismeasurement of any one particle decreases. However, an optimal resolution is achieved when the jet p_T is in the range 30 – 60 GeV. Above this p_T , the resolution worsens for two reasons. First, the need for neutral recovery increases with jet p_T . This process introduces a large uncertainty into the energy of each individual recovered calorimeter cluster, and the jet p_T resolution correspondingly worsens. Secondly, the calorimeter cells start to saturate at higher values of the jet p_T , so less of the jet content is measured, with the jet energy correction becoming larger to correct for this saturation effect. With less of the neutral content of the jet reconstructed and an ‘average’ correction becoming necessary, the jet p_T resolution suffers. Comparison of jet reconstruction at LHCb with and without the neutral recovery shows that the neutral recovery process improves the jet p_T resolution by about 1-2%⁸.

⁸It is important to note that this means that the overall resolution has improved from, e.g. about 13% to, e.g. about 11% (and not by 2% on the 13% figure, which would be negligible).

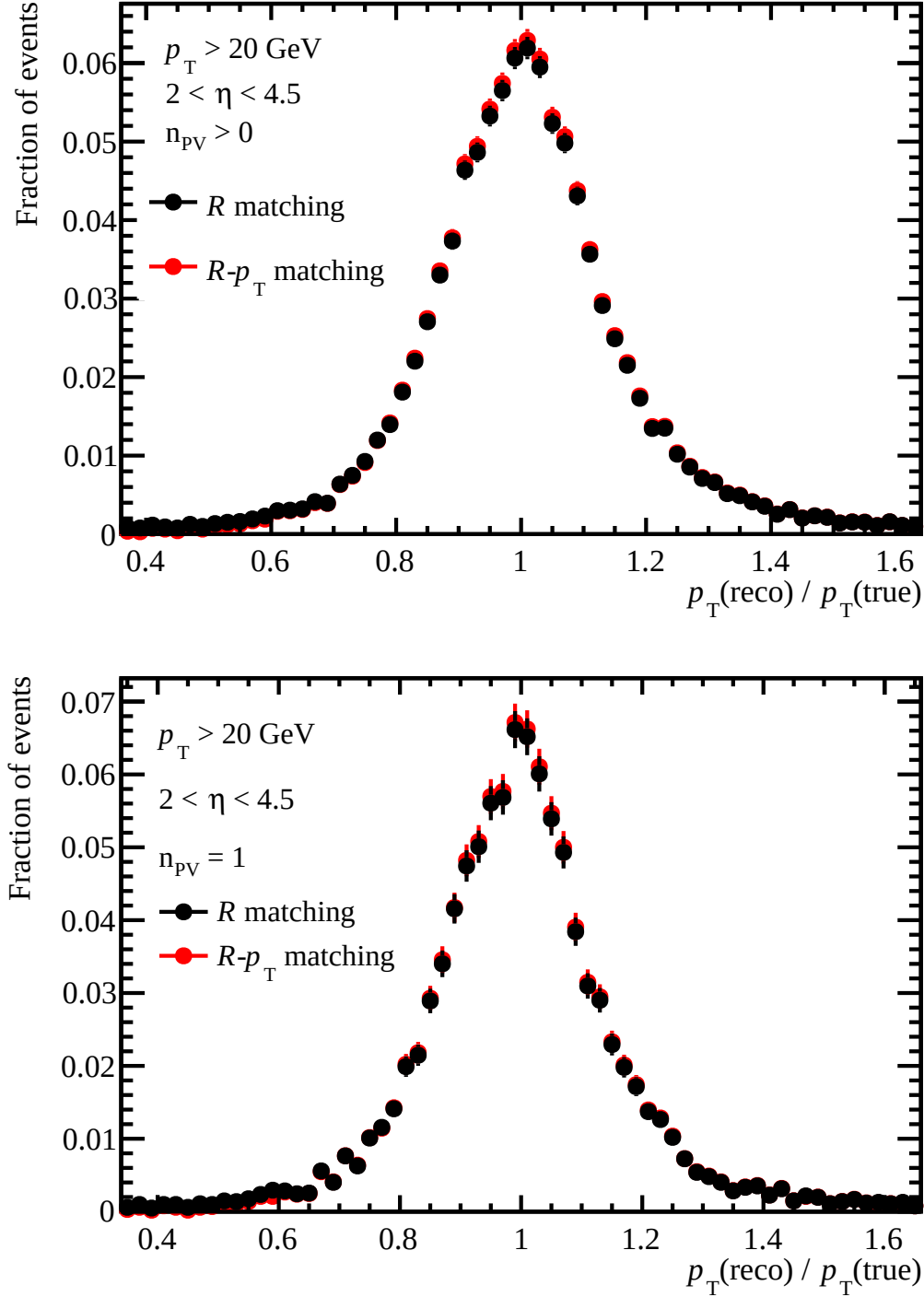


Figure 4.10: The ability of LHCb to reconstruct the jet p_{T} . The top plot shows all events with $p_{\text{T}} > 20 \text{ GeV}$, whereas the bottom plot shows the subset of events containing one primary vertex above the same p_{T} threshold. These are the cleanest events in which the best resolution is achieved. The uncertainties shown are statistical.

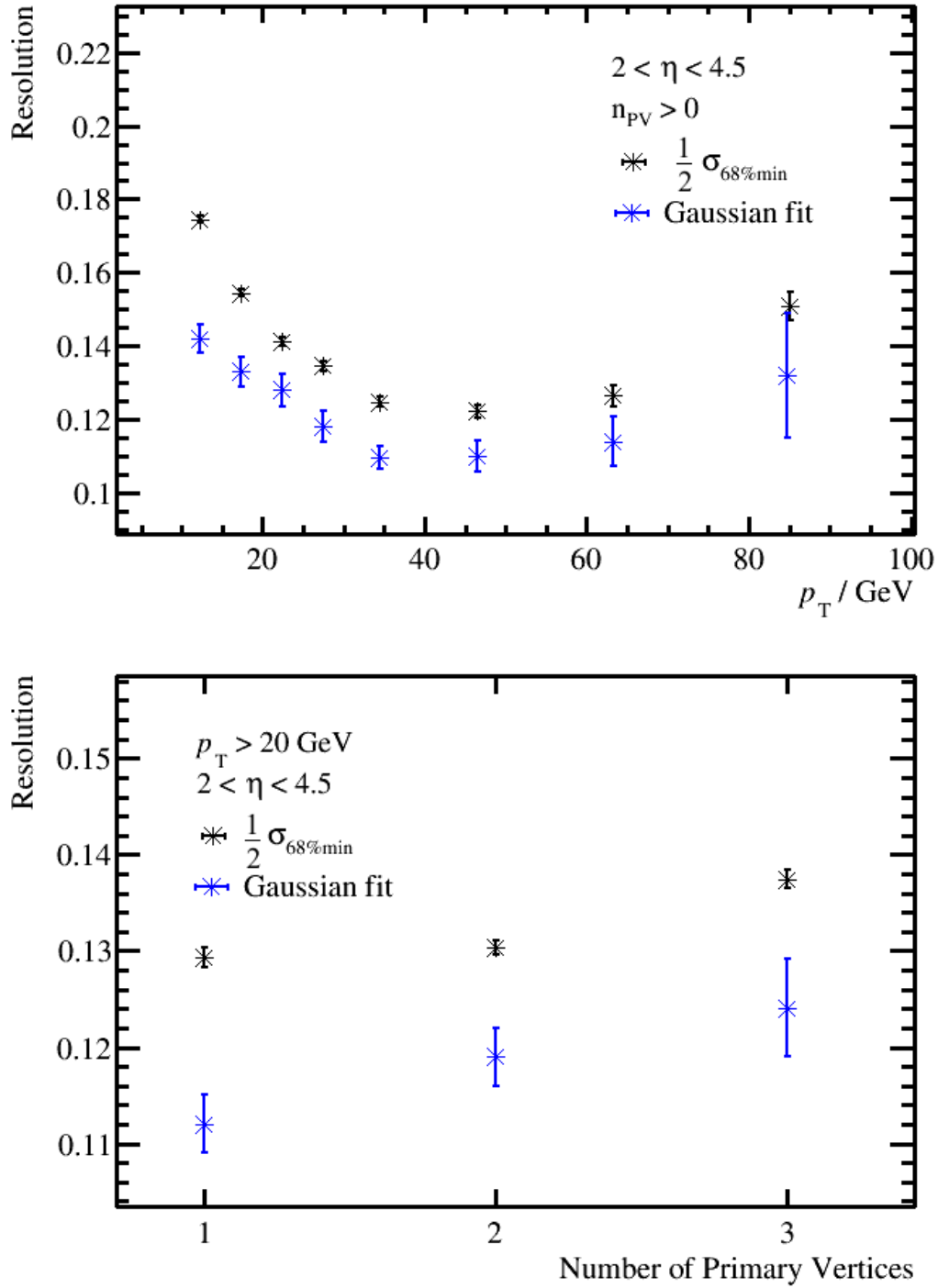


Figure 4.11: The resolution of the jet p_T at LHCb, measured using two different definitions, as a function of the jet p_T (top) and the number of primary vertices (bottom) for jets with $p_T > 20$ GeV. The uncertainties shown are statistical.

4.5 Summary of jet reconstruction at LHCb

We have briefly discussed jet reconstruction at LHCb. The LHCb particle flow approach uses information from throughout the LHCb detector to reconstruct jets, with the best jet transverse momentum resolution typically between 11% and 13%, depending upon the jet p_T , and the number of proton-proton interactions in the event. This performance is comparable with that published⁹ by the general purpose detectors when reconstructing jets with $p_T < 100$ GeV: the ATLAS experiment has an energy resolution between 8 and 15% [134] (depending upon rapidity and p_T), whilst the CMS experiment reconstruct jets with a p_T resolution of about 10% [135].

We now turn our attention to the analysis of events detected at LHCb, making use of the work presented in this chapter.

⁹Discussions with members of the ATLAS and CMS collaborations suggest that their jet p_T resolution has improved since these publications were produced.

Chapter 5

Production of Z bosons in association with jets at LHCb

“The beginning in every task is the chief thing.”

— Plato

The analysis discussed in this thesis considers events containing a Z boson and at least one jet, where the Z boson directly decays to two muons, which we refer to as ‘Z+jet’ events. Z boson decays to two muons tend to be the cleanest Z boson decays to measure in the LHCb detector environment, so this decay channel has advantages over other potential measurements which probe similar underlying physical processes (for example, events where the Z boson decays directly to two electrons). We measure jet production above two transverse momentum thresholds: $p_T > 10$ GeV and $p_T > 20$ GeV, allowing us to consider the dependence of different effects on the jet transverse momentum. Predictions based on pQCD above the 20 GeV threshold are expected to give reasonable agreement with measurements, whereas the region above 10 GeV is more sensitive to soft QCD effects [136].

This chapter starts with an overview of measurements of massive electroweak vector boson production at the LHCb experiment, summarising why such studies are interesting and reviewing existing work. We then cover the selection of Z+jet events, and compare the kinematic distributions of selected events in data and simulation. We consider the purity and efficiencies associated with this selection. These studies are of importance for measurements of the Z+jet production cross-section, discussed in Chapter 6.

5.1 Overview of electroweak production measurements in the forward region

The collisions producing massive electroweak vector bosons at LHCb involve one parton with high Bjorken- x and one parton with low Bjorken- x , as shown in Figure 5.1. This is because the collisions must be boosted into the forward region to be in the LHCb acceptance: the rapidity of a directly produced particle of mass M is related to the x values of the colliding partons (labelled 1 and 2) by $x_{1,2} \simeq \frac{M}{\sqrt{s}} e^{\pm y}$, where $\sqrt{s}$ is the centre-of-mass energy associated with the proton-proton collision. Collisions in the forward region therefore have the ability to probe values of x which are small, and values of x which are close to unity.

The parton distribution functions (PDFs) at high x are already well constrained by previous results from HERA [66, 137] and other experiments. However the low x region has not been directly probed at high Q^2 by previous measurements [66]. In order to calculate predictions for electroweak boson production in the forward region the known parton distributions have to be evolved over many orders of magnitude using the DGLAP equations. These predictions therefore have a large theoretical uncertainty associated with the PDF values used, as shown in Figure 5.2. This means that measurements of electroweak boson production at LHCb can be used to probe PDFs at low x and high Q^2 , testing their behaviour and the DGLAP equations that govern their evolution in this regime. This opportunity is uniquely available at the LHC to LHCb, as the coverage of ATLAS and CMS do not allow such low values of x to be probed. Measurements of electroweak boson production at LHCb are therefore directly complementary to those at the general purpose detectors, and play a key role in the study of PDFs at the LHC.

LHCb has presented measurements of inclusive¹ W and Z boson² production in the muon decay channels using data collected in 2010 [5] (summarised in Figure 5.3) and inclusive Z production in the electron [6] and tau decay channels [138] using data collected in 2011. All these measurements correspond to proton-proton collisions with a centre-of-mass energy of $\sqrt{s} = 7$ TeV. These measurements have all shown good agreement with Standard Model predictions based on a variety of PDF parametrisations (with the

¹We define ‘inclusive’ to mean events with $\geq n$ jets. Inclusive Z production refers to events with ≥ 0 jets. Inclusive Z+jet production refers to events with at least 1 jet. Exclusive Z+1 jet production corresponds to events containing exactly one jet.

²Throughout this and subsequent chapters we use Z to refer to both the Z boson and the virtual photon (γ^*) contributions.

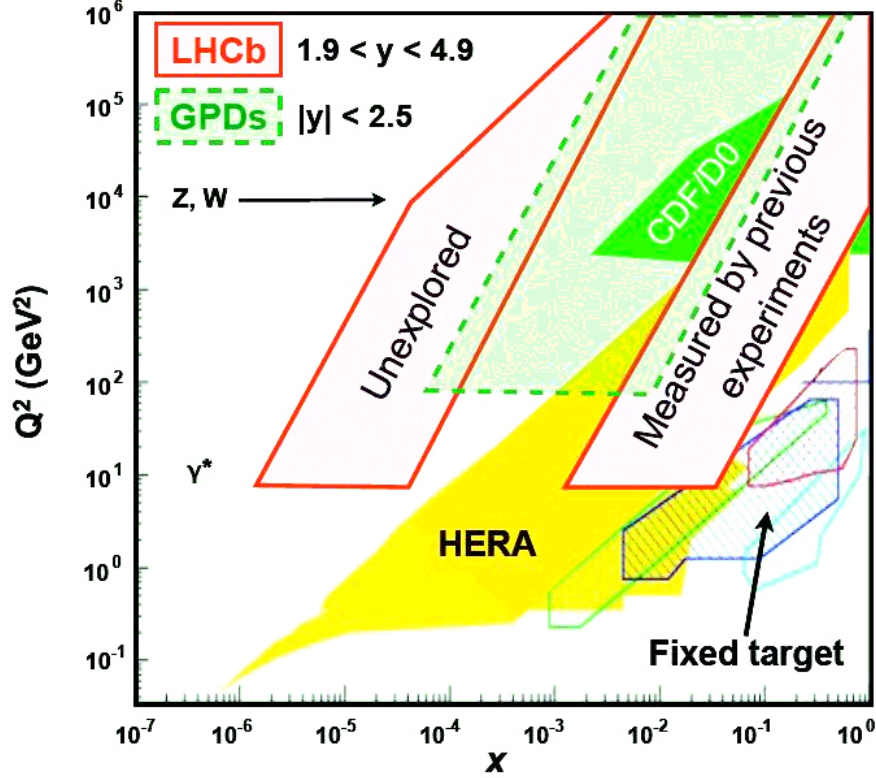


Figure 5.1: The values of x and Q^2 probed by different experiments. Measurements of electroweak boson production at LHCb are sensitive to PDFs at small values of x that previous experiments have not probed. The PDFs used to calculate theoretical predictions of processes at LHCb therefore have to be evolved over many orders of magnitude from the values previously determined. Adapted from Ref. [66] by J. Anderson.

most precise measurements probing the Standard Model at the level of 1-2%). These measurements also agree with measurements made at ATLAS and CMS [139].

Forward Z+jet production

This thesis concentrates on measurements of Z production in association with jets, continuing the studies at LHCb outlined above. The leading order Feynman diagrams for the Z+jet process are shown in Fig. 5.4 (for comparison, the leading order Feynman diagram for inclusive Z production is shown in Fig. 5.5). Theoretical predictions evaluated by the author at $\mathcal{O}(\alpha_s^2)$ using POWHEG [8–11] show that roughly two thirds of Z+jet events in the LHCb acceptance contain gluons in the initial state of the hard process, with this fraction increasing with the jet p_T . Z+jet production is therefore sensitive to the gluon PDF (more information on the sensitivity of vector boson production at the

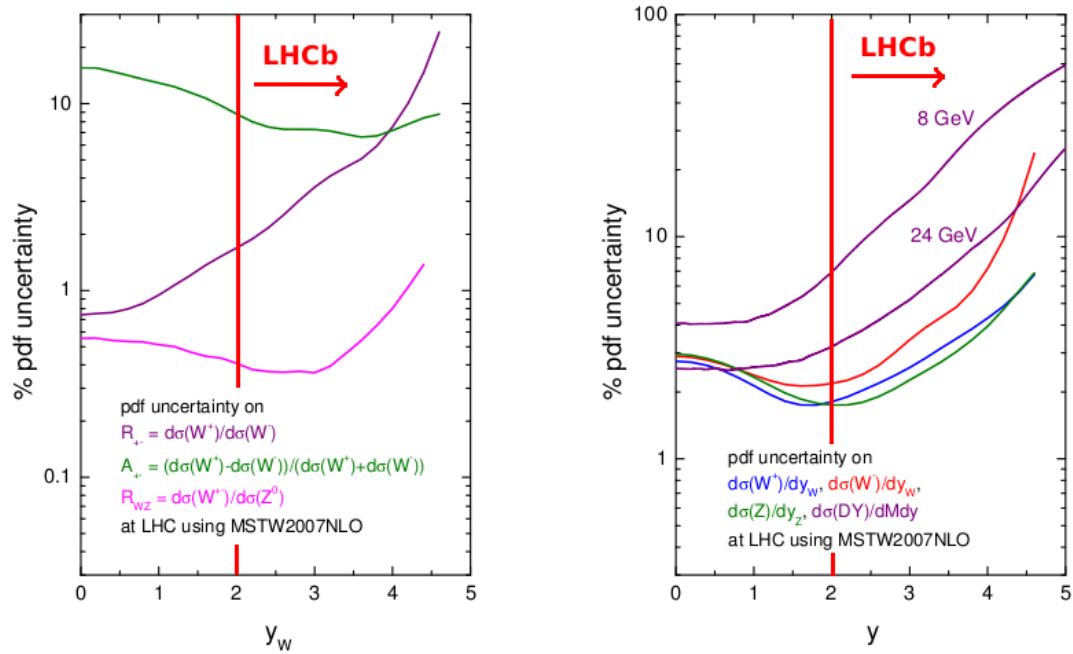


Figure 5.2: PDF uncertainties on standard model measurements of electroweak vector boson production as a function of rapidity. The left hand plot shows W boson production uncertainties and the right hand plot shows uncertainties associated with the Drell Yan process. In the forward region the PDF uncertainty grows, as the PDFs are less well constrained by previous measurements, and have to be evolved over many orders of magnitude from these previous constraints. This means that measurements of vector boson production at LHCb have the ability to probe the reliability of this PDF evolution. Adapted from Ref. [66] by J. Anderson and the author.

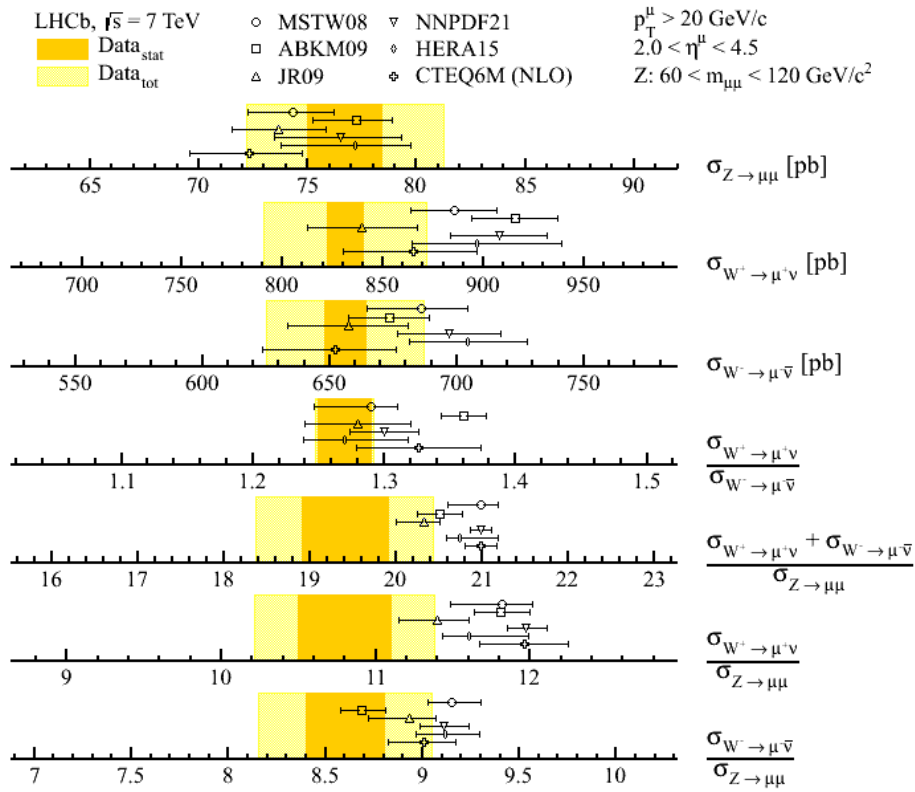


Figure 5.3: Summary of previous results in electroweak vector boson production with muon final states at LHCb. The data are shown as the gold band, with the inner band showing the statistical uncertainty and the outer band showing the systematic uncertainty. The points correspond to theoretical predictions calculated using PDF parametrisations from different PDF fitting groups, with the uncertainties corresponding to the combination of the scale and PDF uncertainties in quadrature. From Ref. [5].

LHC to the gluon PDF is given in Ref. [140]). LHCb occupies a region of phase space where both the Z boson and at least one jet are produced in the forward region. This acceptance is different to that of the general purpose detectors, making measurements at LHCb complementary to those at ATLAS and CMS. Measurements of Z+jet production at LHCb provide an interesting test of both PDFs and the DGLAP equations at low- x and high Q^2 , with particular sensitivity to the gluon PDF. Such studies of the proton's inner structure are important in their own right, but measurements which are sensitive to the gluon PDF at low x are also of interest for the detailed way they probe the PDF evolution. The gluon saturation problem states that the gluon content of the proton is expected to saturate with decreasing x [141] (at very low x). Such saturation is yet to be observed.

As mentioned above, theoretical predictions of the Standard Model cross-section for the Z+jet process have been calculated at $\mathcal{O}(\alpha_s^2)$ [11, 67, 68, 71, 73, 142–144]. Similar analyses at ATLAS [145] and CMS [146–148] have shown reasonable agreement between data and such predictions. In addition to the PDF sensitivity, studies of Z+jet production in the forward region provide complementary tests of pQCD (and the Standard Model) to those at the general purpose detectors. The importance of higher orders in pQCD, and the predictive ability of Monte Carlo generators, can also be examined by comparing the measured data to predictions calculated at different orders in the perturbative expansion of the strong coupling.

Measurements of Z+jet production in the forward region provide important tests of the DGLAP equations, with particular sensitivity to the gluon PDF, and also probe pQCD and the Standard Model. We therefore turn our attention to the data collected by the LHCb experiment, and the experimental considerations we must make to perform a measurement of this process.

5.2 Event selection

In this section we discuss the event selection for Z+jet candidate events, starting with the data samples considered. We also discuss the kinematic distributions of the selected events, comparing their distributions in data and simulation.

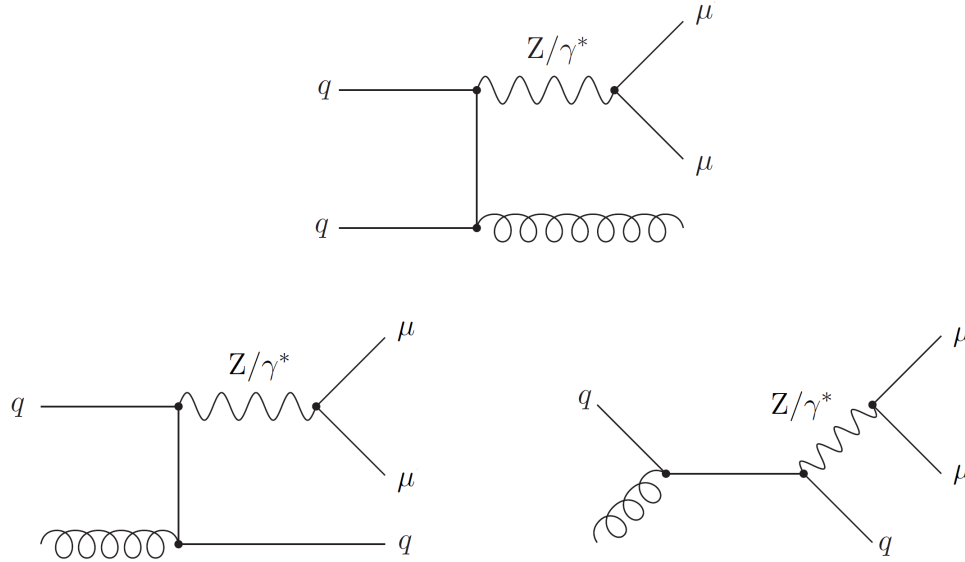


Figure 5.4: The lowest order Feynman Diagrams for Z+jet production at a hadron collider (with the boson subsequently decaying to 2 muons). Charges are not displayed, as all particles in each diagram can be swapped for their corresponding anti-particles and the diagram would still be valid.

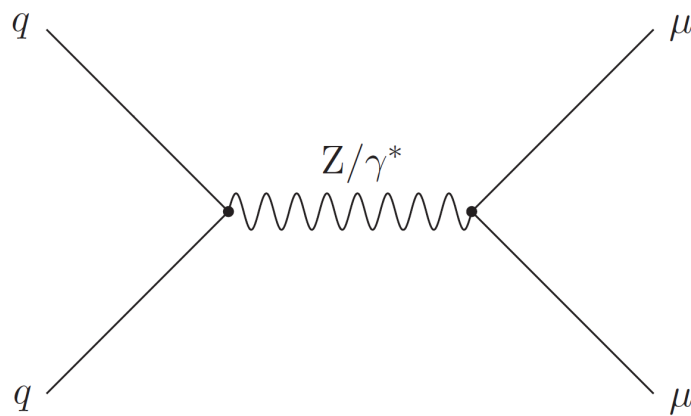


Figure 5.5: The leading order Feynman Diagram for Z production (with the boson subsequently decaying to 2 muons). Charges are not displayed, as all particles in the diagram can be swapped for their corresponding anti-particles and the diagram would still be valid.

5.2.1 Event samples

Data

The analysis presented in this thesis is based on data collected in 2011 with the LHCb detector, corresponding to an integrated luminosity of $\int \mathcal{L} \cdot dt = 1.01 \pm 0.04 \text{ fb}^{-1}$. This is the entire dataset collected in 2011 by LHCb, apart from a small sample (corresponding to an integrated luminosity of about 0.02 fb^{-1}) triggered using a different global event cut (discussed in Section 3.2.8) on the number of SPD hits from the rest of the dataset. As this sample has a different Z boson detection efficiency, and contains few events, it is not considered here.

The analysis is performed using Reco³ 12 and Stripping⁴ 17. This version of the reconstruction software has been extensively studied by the LHCb collaboration.

Monte Carlo simulation

The analysis of Z+jet production also relies on Monte Carlo (MC) simulation which has been interfaced with the detector simulation, as we use simulated events to determine the detector response (though systematic uncertainties are determined from comparisons between data and simulation). The LHCb event simulation environment is based on PYTHIA [12] using the CTEQ6ll [149] parametrisation of PDFs, and was discussed in more detail in Section 3.3.4. The simulation studied here is the MC11⁵ sample of the Z+jet process: the simulation has been analysed using the same reconstruction and stripping selection that is applied to the data collected at the detector. This simulation sample corresponds to a leading order description of the hard process, with the parton shower simulating the effects of further quark or gluon emission. This is expected to provide an accurate description of the content of (and detector response to) light quark jets [150]. The contribution of heavy flavour jets to the Z+jet process is negligible, as it requires heavy flavour in the initial state, and the initial state is dominated by light partons. We have roughly nine times as many Z+jet events in simulation as in data. We also use MC11 simulation of the inclusive Z process (where all jets are produced through a Parton Shower) as a cross-check. This shall be discussed in more detail later.

³This refers to the version of the LHCb particle reconstruction software used.

⁴This refers to the version of the stripping used.

⁵The standard simulation used in LHCb for data taken in 2011.

5.2.2 Selection requirements

Z boson event selection

We use the standard LHCb Z boson selection for candidate $Z \rightarrow \mu^+\mu^-$ events, which was used in the analysis of inclusive Z production at LHCb [5] (we will refer to this as the ‘nominal selection’). We select events containing two tracks satisfying the ISMUON flag (discussed in Section 3.3.2). We then require that these tracks are consistent with a common production vertex, and satisfy the following criteria:

- $p_T^\mu > 20\text{GeV}$,
- $2.0 < \eta^\mu < 4.5$,
- and we require that the invariant mass of the dimuon candidates, $M^{\mu^+\mu^-}$ must lie in the range $60 < M^{\mu^+\mu^-} < 120\text{ GeV}$.

These kinematic requirements also form the fiducial acceptance for the cross-section analysis (discussed later). The invariant mass range was chosen to minimise differences between results quoted at the Born and ‘bare’ levels [151]. In addition, to select well reconstructed events we require that:

- the relative uncertainty on the momentum of each muon is less than 10%: $\sigma_p/|p| < 0.1$. This requirement rejects few events (only 0.08% of events fail this requirement). The requirement is made to ensure that the muon charge type (positive or negative) is known to a high degree of precision. Events with very poor momentum resolution will have a poorly measured track curvature, and so the charge of the particle producing the track will not be determined with sufficient accuracy.
- the muons come from well measured tracks: the χ^2 probability associated with the track fit, $\text{Prob}(\chi^2, \text{dof}) > 0.001$.
- at least one of the muons is TOS (‘trigger on signal’, defined in Section 3.2.8) all the way through the trigger, on the following lines: L0Muon, Hlt1SingleMuonHighPT, Hlt2SingleMuonHighPT. This trigger route selects events containing at least one muon with $p_T > 10\text{GeV}$. The trigger route makes no other kinematic requirements. We refer to this as the single muon trigger route.

After applying these requirements we select 53182 candidate $Z \rightarrow \mu^+\mu^-$ events in the 2011 dataset.

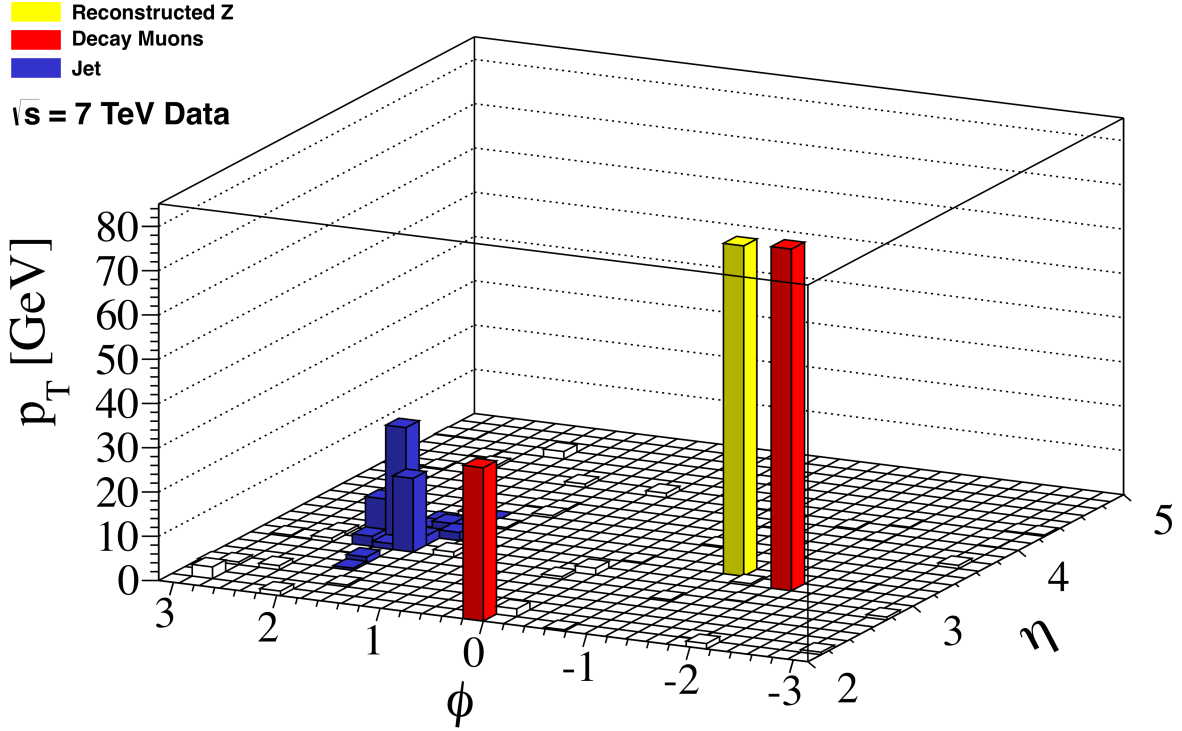


Figure 5.6: A Z+jet event measured at LHCb. The Z boson and the jet are produced back-to-back in ϕ , and are measured to have similar momenta: $p_T^Z = 75$ GeV; $p_T^{\text{jet}} = 65$ GeV. In order to show the angles in the event, the pseudorapidity, rather than the rapidity, is plotted. Consequently the Z boson is further forward than both the decay muons. All the activity in the event is shown in this plot. The event contains only one primary vertex, and little other activity is measured (all reconstructed particles in the LHCb acceptance are shown in this plot). However, it is a fairly typical event Z+jet event at LHCb, as the beam crossings at LHCb provide a low pile-up environment.

Jet selection

As well as the jet identification requirements placed on reconstructed jets (see Section 4.2), we also place kinematic requirements, restricting ourselves to jets within the LHCb acceptance, requiring $2.0 < \eta^{\text{jet}} < 4.5$. We also restrict ourselves to jets which are well separated from the decay muons of the Z boson. We require, for a jet to be considered, $\Delta R(\text{jet}, \mu) > 0.4$. This reduces contributions to jets from clones of, and from final state radiation (FSR) of, muons produced in the Z boson decay.

After applying these requirements, in addition to those on the decay muons, we select 4118(10576) candidate Z+jet events in the 2011 dataset for $p_T^{\text{jet}} > 20(10)$ GeV. A typical event is shown in Figure 5.6.

5.2.3 Kinematic distributions of selected events

The kinematic distributions of selected events in data and simulation are shown in Figs. 5.7, 5.8, 5.9, 5.10, 5.11 and 5.12, for the jet p_T above the 10 and 20 GeV thresholds. No efficiency corrections are applied to these distributions. Broadly, we see agreement between data and simulation in most distributions at better than the 10 % level.

The agreement between data and simulation in the leading jet p_T distribution, shown in Fig. 5.7, and the leading jet η distribution, shown in Fig. 5.8, is particularly impressive. At high p_T , the relative number of events in data is slightly higher than the number of events in simulation. This is expected, as it is at high p_T that the effect of higher orders in pQCD are most important (and simulations at higher orders in pQCD will tend to include a larger number of jets with high p_T). The leading jet η distribution is shown in Figure 5.8. The distribution favours jet production at lower pseudorapidity. This is because high p_T jets are easier to produce at lower pseudorapidities, as shown in Figure 4.1. This effect becomes more pronounced when the jet p_T threshold is increased.

There is some disagreement between data and simulation in the Z boson p_T distribution at low p_T , shown in Fig. 5.9. Similar discrepancies have been seen in measurements at other LHC detectors [152]. The vector boson p_T is known to be hard to simulate accurately at low p_T , due to the effects of multiple gluon emission and soft QCD [136]. However, in the high p_T tail, above the peak in the distribution, data and simulation show good agreement. The position of the peak is seen to change with different p_T requirements being made on the jet. This is also well understood: with higher p_T jets in the event, the Z boson will also tend to have a higher p_T in order to give zero total p_T in the event (as the Z and jet tend to be produced back-to-back in ϕ).

There is also a small discrepancy between data and simulation in the Z boson rapidity distribution, shown in Fig. 5.10. This discrepancy is less than 10% (typically about 5%), and is probably (at least, partly) due to the sensitivity of the Z boson rapidity distribution to the choice of PDF. Indeed, more recent PDFs parametrisations than the standard LHCb choice show similar changes in forward Z production. The peak in the distribution is due to the requirement that both decay muons lie in the LHCb acceptance ($2.0 < \eta^\mu < 4.5$). If there were no kinematic requirements placed on the decay muons then the distribution of $|y|$ would be observed to fall monotonically. In addition, increasing the jet p_T requirement from 10 GeV to 20 GeV slightly increases the fraction of events where the Z boson is produced at low rapidity. This is for the same reasons as those discussed above for the jet η distribution: more energetic events tend to occur in

the central (low rapidity) regions. However, this effect on the Z rapidity distribution is very small.

The difference in rapidity⁶ between the Z and the leading jet, $\Delta y(\text{Z}, \text{jet}) \equiv y^{\text{Z}} - y^{\text{jet}}$, is shown in Figure 5.11. This distribution shows good agreement between data and simulation. The shape of the distribution is driven by the kinematic acceptance of the detector: in selecting events with $2 < y < 4.5$ the difference in rapidity between the Z and the jet is constrained to lie within the region between -2.5 and $+2.5$. The peak lies above zero, since jet production is most favoured at low pseudorapidities, whereas we tend to measure Z bosons which are produced in the centre of our acceptance, given the requirements that both decay muons must lie in the LHCb acceptance. As the jet p_{T} threshold increases, the peak of the Δy distribution becomes sharper.

The simulation overestimates the number of events where the Z boson and jet are produced back-to-back, as seen in Fig. 5.12. The simulation relies on a parton shower to produce jets around the hardest parton in the event. The parton shower is only accurate in the collinear approximation, as discussed in Section 2.2.2 - it underestimates jets which are produced at large separations from the hardest jet. The hardest jet will tend to be produced back-to-back with the Z boson, but may not be in our acceptance. So, in the simulation, jets which are not back to back with the Z boson will tend to be produced in the parton shower, and the number of these is likely to be underestimated. As the jet p_{T} threshold increases, the fraction of events where the Z and the jet are produced back-to-back increases. This is due to the fraction of events containing an additional jet decreasing as the leading jet p_{T} increases. With fewer additional jets in the event, the Z boson and the jet momenta must balance each other more precisely in the transverse plane: the Z boson and the jet become more back-to-back in ϕ .

The jet multiplicity shows the largest discrepancy between data and simulation. If we consider the ratio of the number of reconstructed inclusive Z+2 jet events to the number of reconstructed inclusive Z+jet events, $N_{\text{Z}+2 \text{ jet}}/N_{\text{Z}+\text{jet}}$, for $p_{\text{T}}^{\text{jet}} > 10 \text{ GeV}$, we see that there is a large difference between data (0.157 ± 0.004) and simulation (0.098 ± 0.001) ⁷. The simulation we run is for the Z+jet process, and only includes additional jets beyond the jet generated in the hard process when they are generated by the parton shower. However, parton showers underestimate the radiation of partons at large angles from the

⁶We consider for this distribution the true rapidity of the jet. Since the invariant mass of a light parton induced jet is typically much smaller than the total momentum of the jet, the difference between the jet pseudorapidity and the jet rapidity is negligible.

⁷The uncertainties quoted here are statistical.

primary emission, and consequently the number of additional jets (above a significant p_T threshold, like the 10 GeV threshold applied here) is also underestimated. The underestimate in this region of phase space is not surprising: the simulation is only expected to be accurate for the leading jet (for which we see excellent agreement between data and simulation, in Fig. 5.7).

We see then that there are three areas where the simulation fails to accurately reproduce the data: at low p_T^Z , at small $\Delta\phi$, and in events which contain more than one jet. However, these situations are largely equivalent. If the Z boson has a small p_T in an event containing a jet, then another jet must be produced in the event to balance the total transverse momentum⁸. In turn, the production of the additional jet means that the angular separation of the Z boson and the leading jet (in our acceptance) is reduced. This region is dominated by higher orders in pQCD, and so it is not surprising that the leading order simulation fails to reproduce the data.

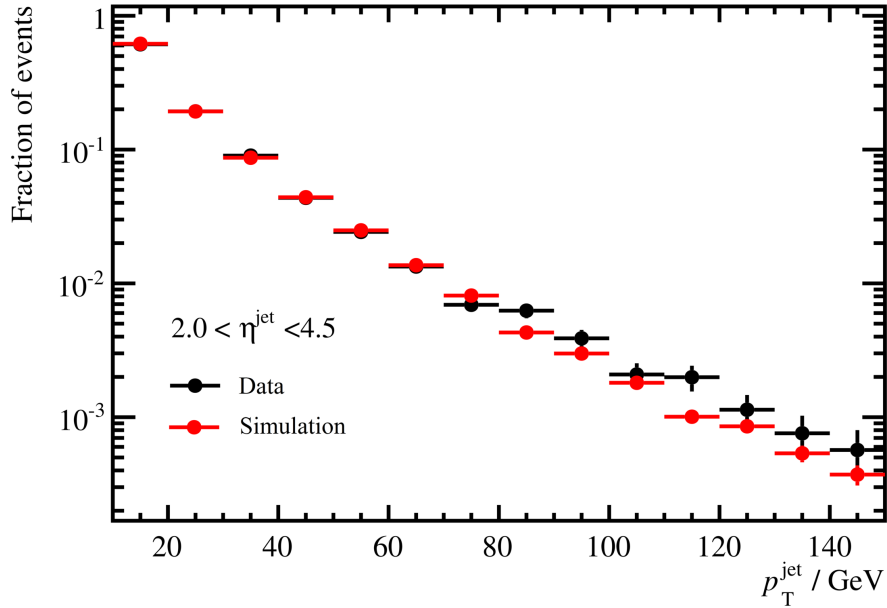


Figure 5.7: The leading jet transverse momentum distribution in data (black) and simulation (red), for Z+jet events. No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

⁸We note that this additional jet may not be in the LHCb acceptance.

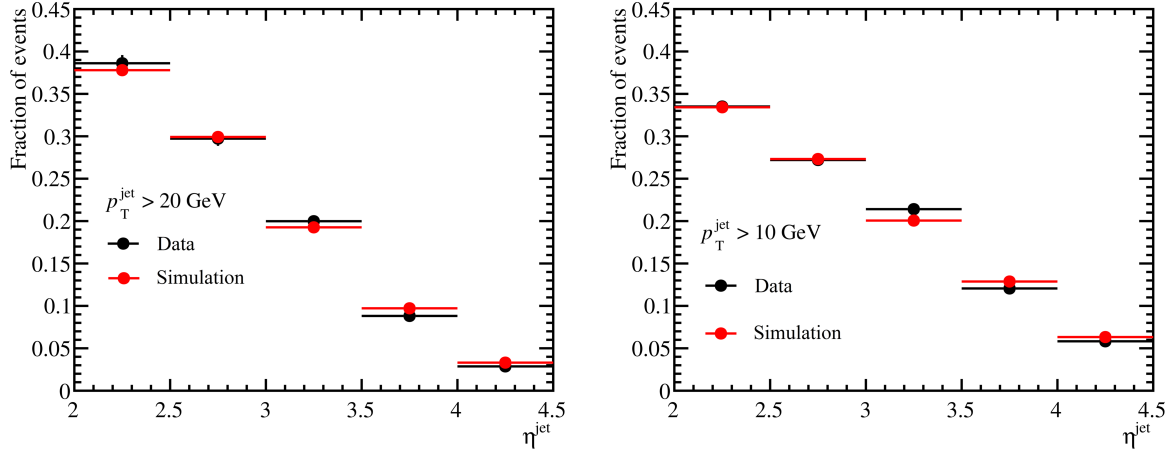


Figure 5.8: The leading jet pseudorapidity distribution in data (black) and simulation (red), for Z+jet events, for $p_T^{\text{jet}} > 20 \text{ GeV}$ (left) and $p_T^{\text{jet}} > 10 \text{ GeV}$ (right). No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

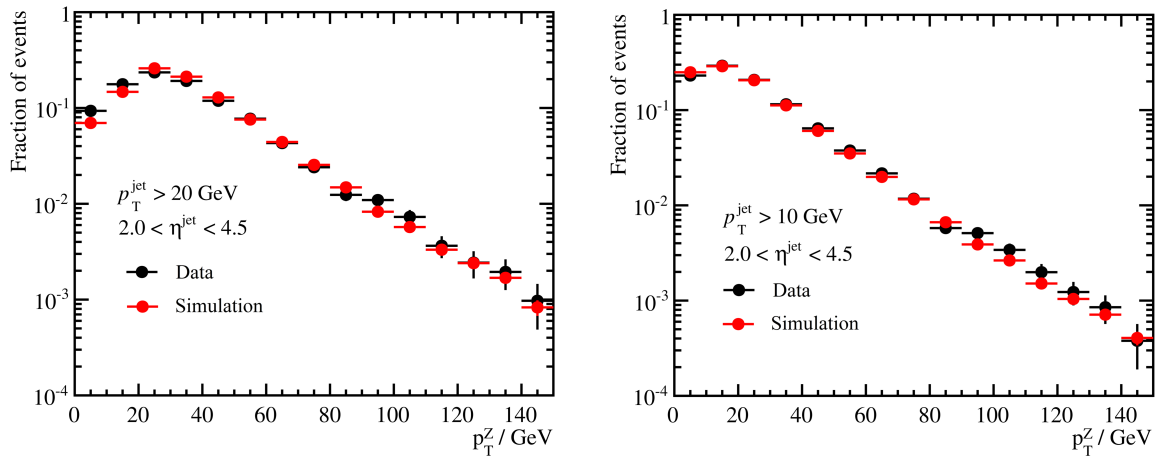


Figure 5.9: The Z transverse momentum distribution in data (black) and simulation (red), for Z+jet events, for $p_T^{\text{jet}} > 20 \text{ GeV}$ (left) and $p_T^{\text{jet}} > 10 \text{ GeV}$ (right). No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

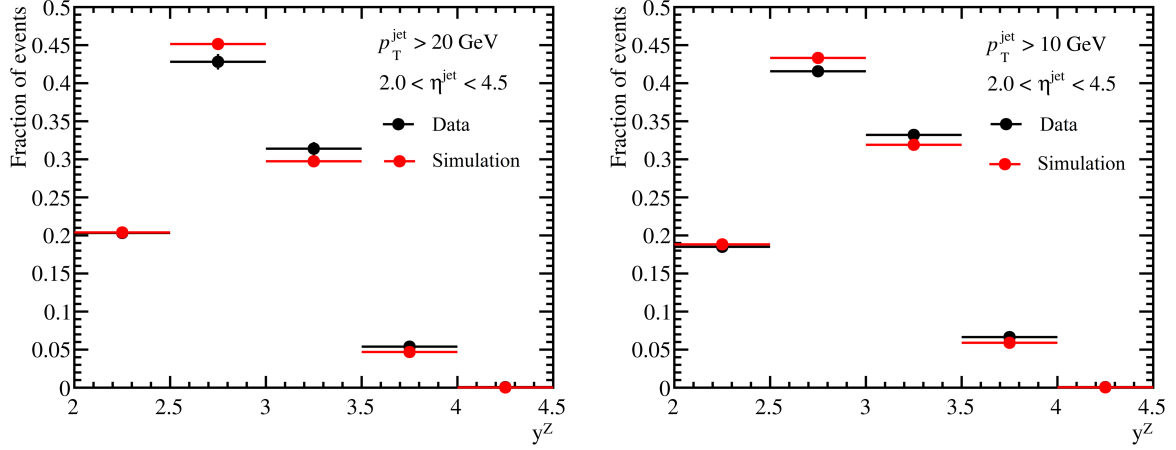


Figure 5.10: The Z rapidity distribution in data (black) and simulation (red), for Z+jet events, for $p_T^{\text{jet}} > 20$ GeV (left) and $p_T^{\text{jet}} > 10$ GeV (right). No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

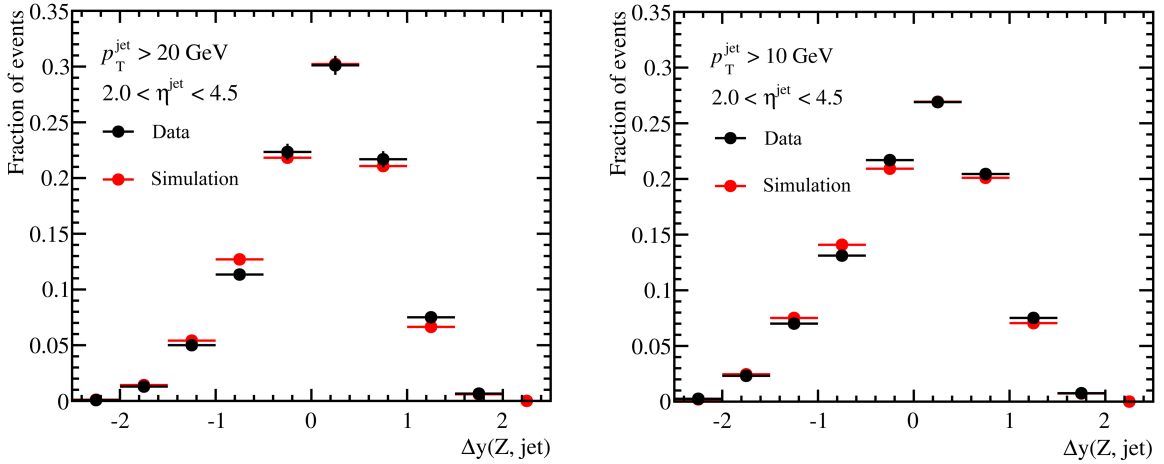


Figure 5.11: The difference in rapidity between the Z boson and the leading jet in the event, in data (black) and simulation (red), for Z+jet events, for $p_T^{\text{jet}} > 20$ GeV (left) and $p_T^{\text{jet}} > 10$ GeV (right). No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

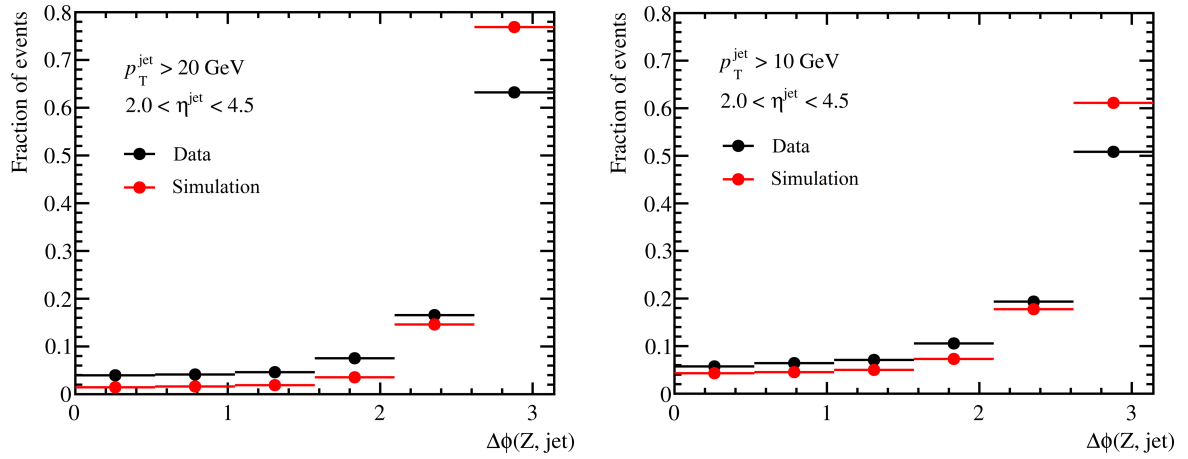


Figure 5.12: The modulus of the difference in azimuthal angle between the Z boson and the leading jet in the event, in data (black) and simulation (red), for Z+ jet events, for $p_T^{\text{jet}} > 20 \text{ GeV}$ (left) and $p_T^{\text{jet}} > 10 \text{ GeV}$ (right). No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

5.3 Background

Other processes at LHCb can give rise to particles that mimic the Z+jet events. We consider the background contribution from different sources:

1. We consider the combinatorial background, from random combinations of muons in the event that happen to be produced with the correct invariant masses. These muons could come from semileptonic B decays, W boson decays, or could be mesons that have either decayed in flight or incorrectly pass the ISMUON flag. A jet could also be produced in the same event from unrelated QCD processes (and multiple parton interactions), leading to the event passing all our selection criteria. We estimate this background by considering the number of same sign events ($\mu^+\mu^+$ or $\mu^-\mu^-$) which otherwise pass our selection criteria, using the same dataset collected in 2011. Assuming no charge bias, this should equal the number of opposite sign events which would pass our selection from the same background combinatorial processes. Using this method we estimate this background to be 16 ± 4 events, for the 10 GeV jet p_T threshold, and 5 ± 2 for the 20 GeV threshold, where the uncertainties are statistical. The number of $\mu^+\mu^+$ and $\mu^-\mu^-$ events are consistent. The number of same sign events has been demonstrated to give a good estimate of the combinatorial background in other studies of electroweak boson production [6], where alternative methods of predicting the combinatorial background were shown to agree at the level of $\sim 5\%$. In our case this uncertainty is much smaller than the statistical uncertainty in the background.
2. $Z \rightarrow \tau\tau$ decays, where both taus subsequently decay to muons, and a jet is produced in association with the Z boson form another potential background source. We rely on Monte Carlo simulation to estimate the number of these events which would pass our selection. We estimate this background to be 12 ± 3 for the 10 GeV jet p_T threshold, and 7 ± 3 for the 20 GeV threshold. The uncertainties are statistical.
3. We consider $t\bar{t}$ production, where each top quark decays to a b quark and a W boson. The b quarks will be associated with jets. If the W bosons each decay to a muon and a neutrino, the overall final state could mimic the Z+jet final state. NLO simulation suggests that we would detect 14 events in 1 fb^{-1} of data [153] with both muons passing our η and p_T selection, neglecting reconstruction efficiencies. Half of these events will pass our dimuon invariant mass selection [154]. Assuming that we have an efficiency to reconstruct both muons of 70% (see Section 5.4), this means that our expected background level is 5 ± 2 . The background level does not change

between the 10 GeV and 20 GeV thresholds, as the b -quark-induced jets in these events are expected to have $p_T \sim m_{\text{top}}/2$, which is much larger than the threshold. The uncertainty is statistical.

4. The diboson production process with the largest cross-section in our fiducial acceptance, W boson pair production, is around 20% of the $t\bar{t}$ production cross-section [5] in our acceptance. The requirement that a jet is produced in our acceptance in association with the W boson pair further reduces this cross-section. It is therefore considered a negligible background source in comparison to the other possible background sources considered. Other diboson production processes are even rarer, and so also contribute a negligible amount to the sample studied.
5. Heavy flavour decays of Z bosons, where the heavy flavour decay products decay to muons, provide a negligible contribution, as the muons rarely have sufficient energy to pass the invariant mass and p_T requirements.

We therefore estimate the total background to be 33 ± 6 for the 10 GeV jet p_T threshold and 17 ± 4 for the 20 GeV threshold. This corresponds to sample purities, ρ , of 0.9969 ± 0.0005 , and 0.9959 ± 0.0010 respectively. The results for the two thresholds are consistent - there is no evidence for the background level varying significantly with jet transverse momentum. These purities are also consistent with the purity of 0.9971 ± 0.0003 measured in the preliminary measurement of inclusive Z production in data collected in 2011 [155]. Previous studies [5, 155] with higher statistics have also shown no evidence for any variation of the background level with different kinematic distributions of the Z boson. We therefore assume that the purity is flat. The effect of this assumption is minimal as the total background level is small.

We have not considered here the effect of events which contain a (correctly) reconstructed dimuon decay of a Z boson, but where a jet is incorrectly reconstructed (no jet is present at truth level, but a jet is reconstructed in the event). These correspond to ‘fake jets’, and will be discussed separately later.

Cross-check of the background levels

The study described above showed that the total background level in selected Z +jet events (at both the 10 GeV and 20 GeV jet p_T thresholds) is consistent with the background level in the inclusive Z events.

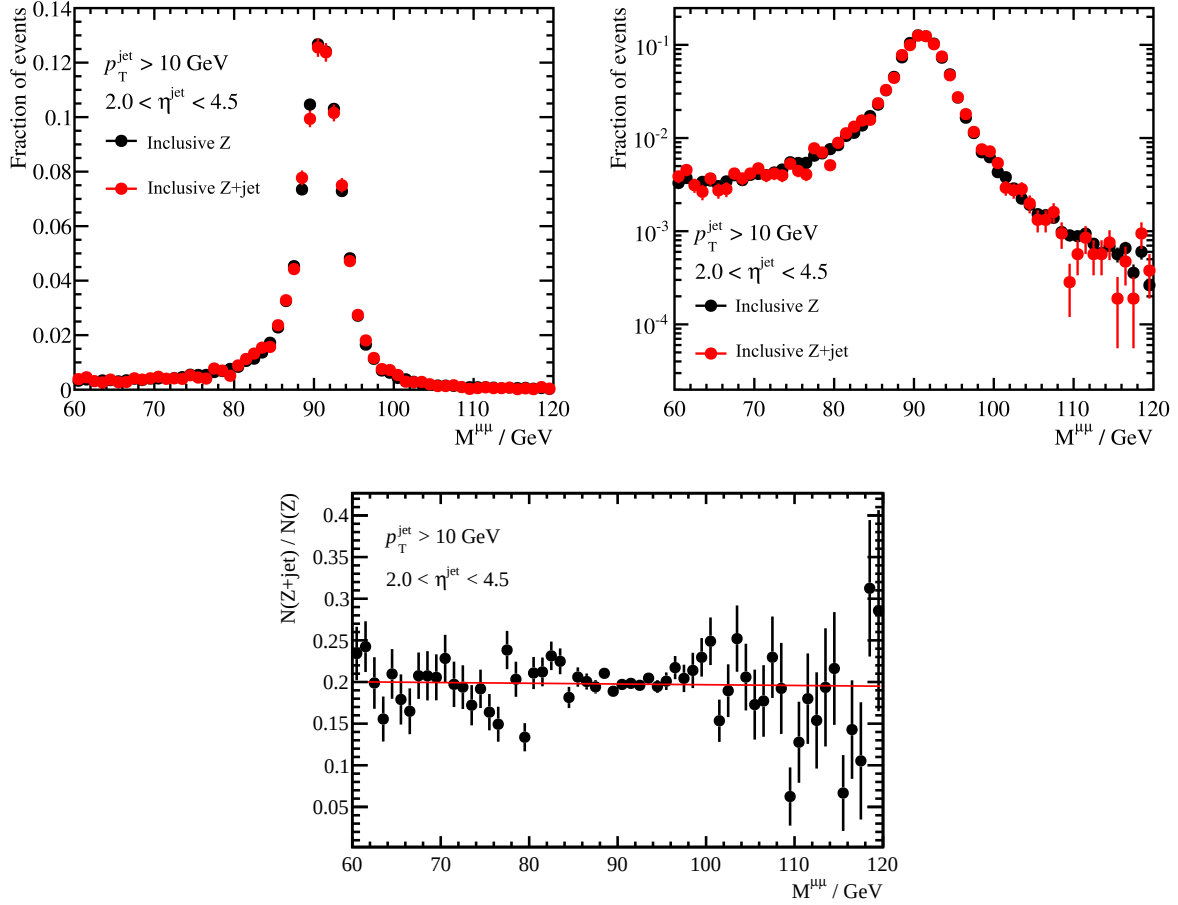


Figure 5.13: The top plots show the invariant mass distributions of dimuon candidates for inclusive Z (black) and inclusive Z+jet (red) events, where both distributions are normalised to have unit integral, for jet $p_T > 10$ GeV. The top-left plot has a linear scale, whilst the top-right plot has a log scale for the y-axis. The bottom plot shows the ratio of the number of Z+jet events to the number of inclusive Z events as a function of the invariant mass of the dimuon pair, for jet $p_T > 10$ GeV. For all plots, the uncertainties shown are statistical. The distribution in the bottom plot is fit with a straight line, $p_0 + p_1 \cdot M^{\mu^+\mu^-}$, with the best fit determined using minimisation of the χ^2 . We find $p_0 = 0.20 \pm 0.02$ and $p_1 = 0.00003 \pm 0.00023 \text{ GeV}^{-1}$, consistent with a flat line.

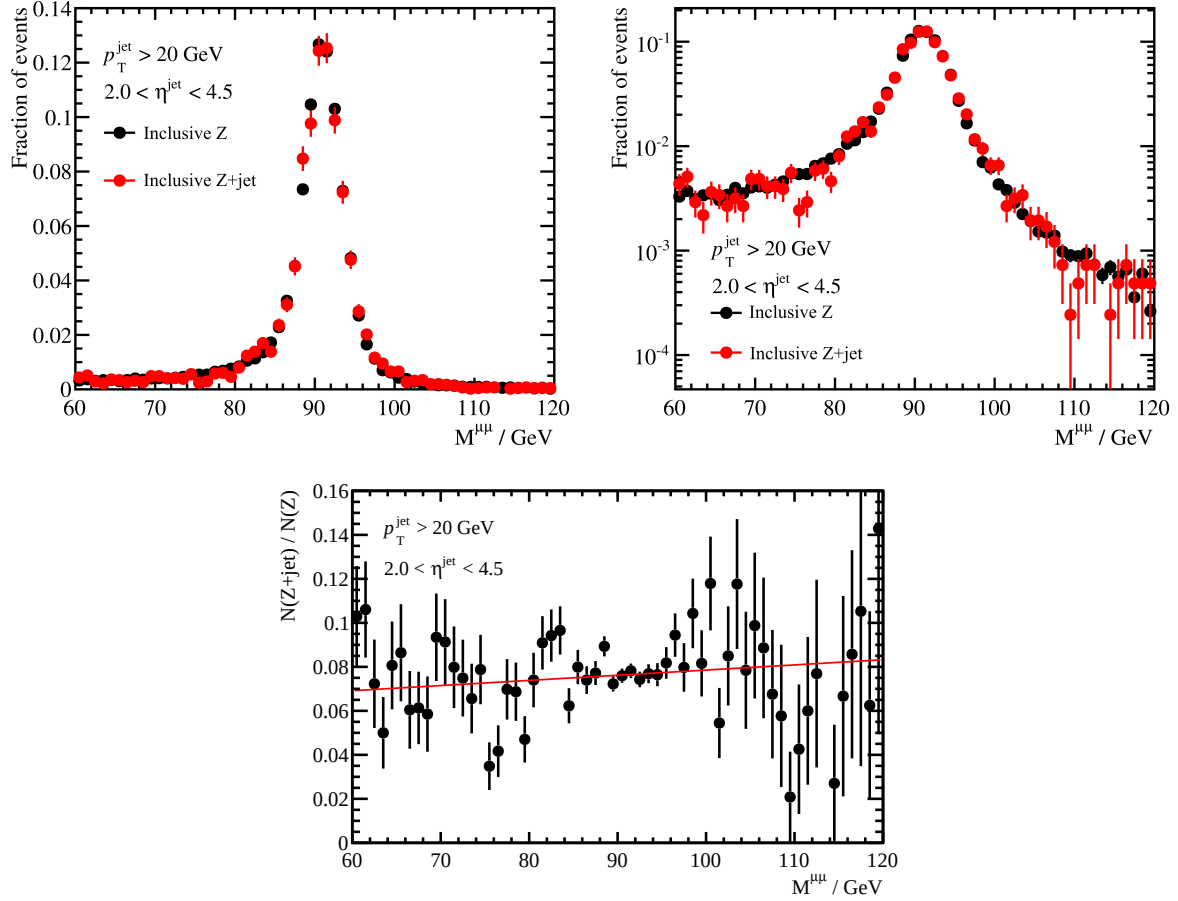


Figure 5.14: The top plots show the invariant mass distributions of dimuon candidates for inclusive Z (black) and inclusive Z+jet (red) events, where both distributions are normalised to have unit integral, for jet $p_T > 20$ GeV. The top-left plot has a linear scale, whilst the top-right plot has a log scale for the y-axis. The bottom plot shows the ratio of the number of Z+jet events to the number of inclusive Z events as a function of the invariant mass of the dimuon pair, for this plot, the jet $p_T > 20$ GeV. For all plots, the uncertainties shown are statistical. The distribution in the bottom plot is fit with a straight line, $p_0 + p_1 \cdot M^{\mu^+\mu^-}$, with the best fit determined using minimisation of the χ^2 . We find $p_0 = 0.07 \pm 0.01$ and $p_1 = 0.0002 \pm 0.0002 \text{ GeV}^{-1}$, consistent with a flat line.

We check this by considering the invariant mass distributions of the dimuon pairs in both inclusive Z and Z+jet events. Signal events will form the characteristic Z peak with a low mass tail associated with the virtual photon. Background events should not form such a peak; instead they are expected to be preferentially located at low dimuon invariant masses. Therefore, if the background levels were higher(lower) in the Z+jet events than in inclusive Z events, we would see proportionally more(fewer) Z+jet events with a low dimuon invariant mass.

The invariant mass distributions for inclusive Z and Z+jet events are shown on the same plot in Figure 5.13 (for the 10 GeV jet p_T threshold). The invariant mass distributions agree well between the Z+jet events and the inclusive Z events. The ratio of the number of Z+jet events to the number of Z events as a function of the dimuon invariant mass⁹, also shown in Fig. 5.13 (for the 10 GeV threshold), is consistent with a flat line. This is consistent with our hypothesis that the background levels are the same in the Z boson and Z+jet samples, assuming that any enhancement or deficit of the background is not symmetric about the centre of the range considered (90 GeV). The same is true of the 20 GeV threshold, shown in Figure 5.14.

5.4 Z boson detection and reconstruction efficiencies

Following Ref. [5], the total Z boson detection efficiency can be factorised into four separate components as $\varepsilon_Z = \varepsilon_{\text{ID}} \varepsilon_{\text{Track}} \varepsilon_{\text{Trigger}} \varepsilon_{\text{GEC}}$, where the ε_X factors correspond to the efficiencies associated with the muon identification, the muon track reconstruction, the trigger requirements and the global event cuts respectively.

We calculate each of these efficiencies separately using data-driven methods. These methods have already been used in analyses of inclusive electroweak boson production at LHCb [5, 6, 155]. The validity of splitting the efficiencies into separate, independent contributions (with negligible correlations between the efficiencies) has been demonstrated by other studies [156].

The muon track reconstruction, identification and trigger selection efficiencies are found as a function of the kinematic parameters of the muon. These can then be combined on an event by event basis, and combined with the global event cut efficiency, to determine an overall Z boson detection efficiency, for the kinematics of an individual

⁹We note when taking the ratio that the Z+jet events are also present in the inclusive Z sample.

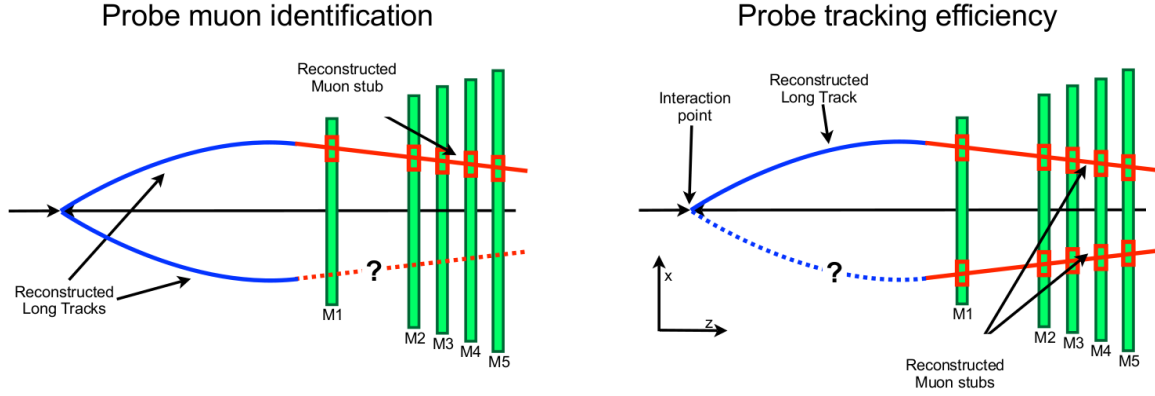


Figure 5.15: The tag-and-probe method of measuring efficiencies for both the muon identification and muon track reconstruction efficiencies. The left-hand schematic shows the method used to measure the muon identification efficiency, whereas the right-hand schematic shows the method used to measure the track reconstruction efficiency. From R. McNulty.

event. We will therefore discuss these efficiencies in turn. Before this, however, we briefly describe the use of tag-and-probe methods to measure efficiencies, as such an approach is used to determine the muon identification, track reconstruction and trigger efficiencies.

A schematic showing the tag-and-probe method is shown in Figure 5.15. The method relies on selecting pure events based on one particle in the event. If this is sufficient to know that there is another particle in the event, then whether or not there is a reconstructed track associated with that particle, or if the particle is correctly identified, provides knowledge of the efficiency associated with track reconstruction or particle identification respectively.

5.4.1 Muon identification

The muon identification efficiency is defined to be the probability that a true muon is reconstructed as a muon, passing the ISMUON requirement. We therefore seek to measure how often true muons pass the ISMUON requirement.

We do this using a data-driven tag-and-probe method (as we do not know the reliability of the simulation). We select events consistent with a $Z \rightarrow \mu^+ \mu^-$ decay, containing two long tracks which have a combined invariant mass consistent with a Z boson. Tight selection requirements are then placed on one of the candidates (the tag), but the other candidate (the probe) is not required to pass any particle identification requirement. A schematic of this is shown in Figure 5.15. We then find the ratio of the number of probe

tracks that pass the ISMUON requirement to the total number of probe tracks considered. This provides a measurement of the muon identification efficiency. This efficiency is studied as a function of relevant variables (like the muon pseudorapidity, or the jet multiplicity). Different muon charges and different magnet polarities are also considered separately.

The requirements used to select the tag-and-probe events are:

- Requirements on both tag and probe:
 - $86 < M(\text{tag, probe}) < 96$ GeV — the invariant mass of the tag and probe particles must lie within 5 GeV of the nominal Z boson mass,
 - The χ^2 probability associated with the track fit must satisfy $\text{Prob}(\chi^2, \text{dof}) > 0.001$; the particles must be associated with high quality tracks,
 - $\sigma_p/|p| < 0.1$,
 - $p_T > 20$ GeV,
 - $2.0 < \eta < 4.5$,
 - $|\Delta\phi(\mu^+, \mu^-)| > 0.1$ - this requirement is very loose, and is simply present to ensure that one particle is not a poorly measured clone of the other.
- Requirements on the tag:
 - $p_T > 30$ GeV,
 - Passes the ISMUON requirement.
 - $\text{ProbNNmu} > 0.68$ - this number corresponds to the neural network based particle identification mentioned in Section 3.3.2,
 - TOS on the single muon trigger lines,

There is no evidence for any dependence of the efficiency on magnet polarity, or for differing μ^+ or μ^- efficiencies. There is also no evidence for any dependence of the muon identification efficiency on the jet multiplicity (see Figure 5.16), but we do see a dependency on the muon pseudorapidity (shown in Figure 5.17). The dip in efficiency at high pseudorapidity corresponds to the edge of the detector where the muon identification performs less well. Previous studies have shown little dependence upon the muon p_T [5, 155], and no evidence is seen here. Indeed, the efficiencies found here agree well with those reported by others [157].

This selection is designed to ensure a high purity $Z \rightarrow \mu^+ \mu^-$ sample, without biasing the muon identification efficiency associated with the probe. The invariant mass distribution of the tag and probe pairs is shown in Figure 5.18. We check for any effects of impurities

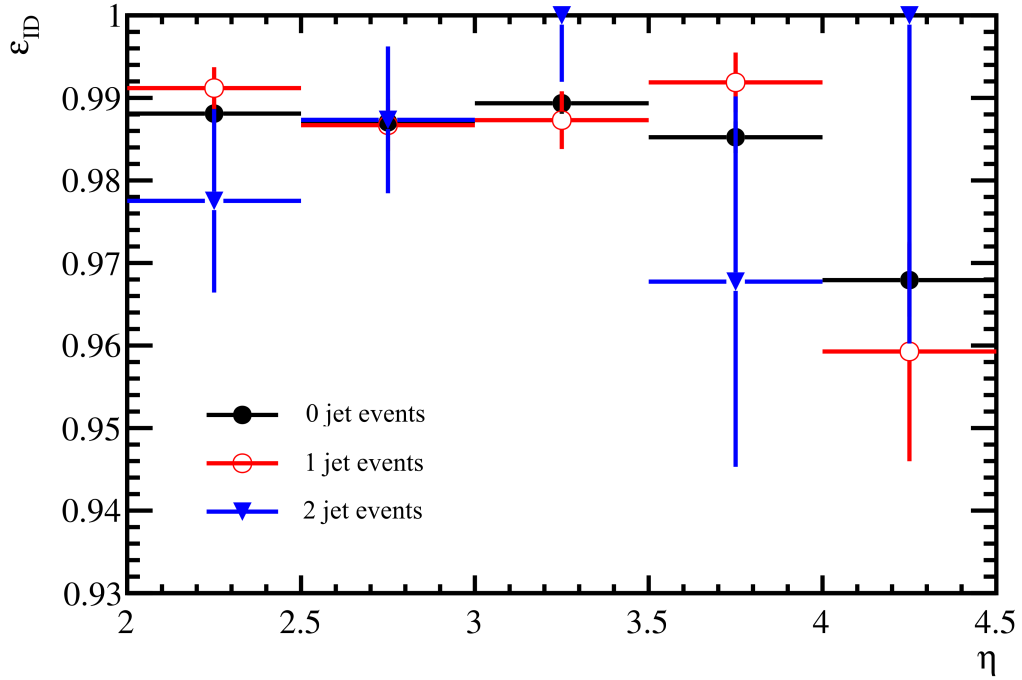


Figure 5.16: The muon identification efficiency, binned as a function of the (exclusive) jet multiplicity and the muon pseudorapidity. We see no obvious dependence on jet multiplicity. The uncertainties shown are statistical. Where the efficiency is close to unity, the uncertainty is assigned using the Feldman-Cousins approach [158].

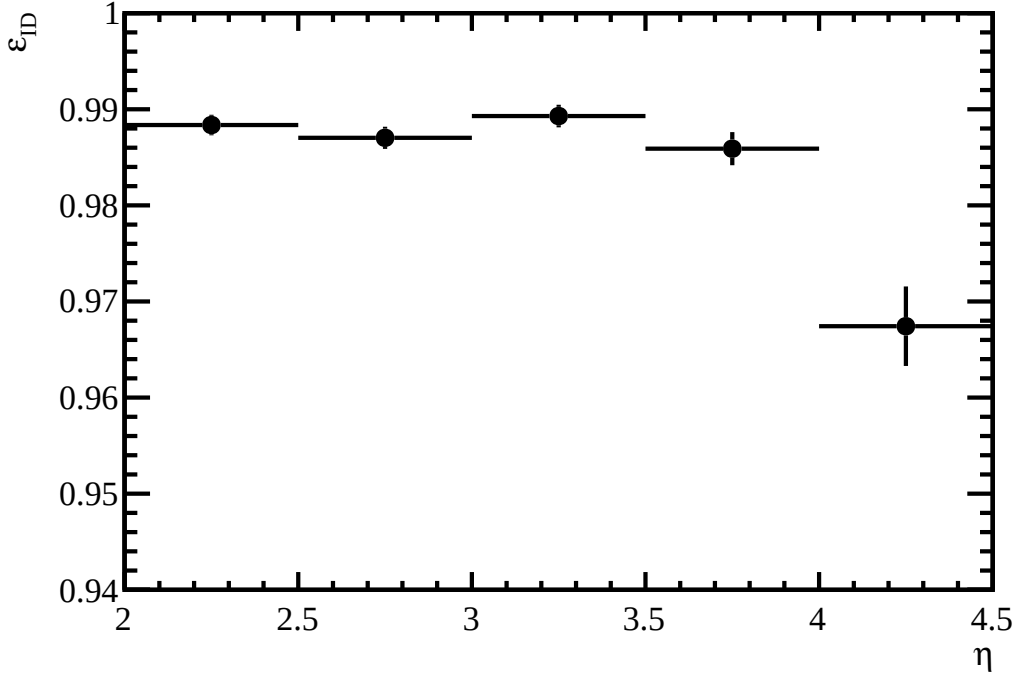


Figure 5.17: The muon identification efficiency, binned as a function of the muon pseudorapidity. The uncertainties shown are statistical, calculated using the binomial approach.

in the efficiency determination by repeating the calculation using a wider mass window, but tightening the other selection requirements so that:

- $70 < M(\text{tag, probe}) < 110$ GeV,
- the HCAL energy deposit associated with the tracks (E_{HCAL}) is small: $E_{\text{HCAL}}/|p| < 0.1$,
- the tag track is isolated - the total transverse momentum associated with tracks in a cone of size 0.7 in $\eta - \phi$ space centred on the tag is less than 10 GeV.

The effect on the efficiency from changing these selection requirements is about 0.5%, and is assigned as a systematic uncertainty. The effect of this on the Z+jet production cross-section shall be discussed later.

Validation of method

We show in Figure 5.19 the muon identification efficiency as a function of the azimuthal angle of the muon in the detector, ϕ . There is a small drop in efficiency noticeable at $\phi = \pm \pi/2$, associated with the change in ‘sides’ of the detector. Whilst the effect is small ($< 1\%$), it could lead to a bias in the tag-and-probe method through correlations in

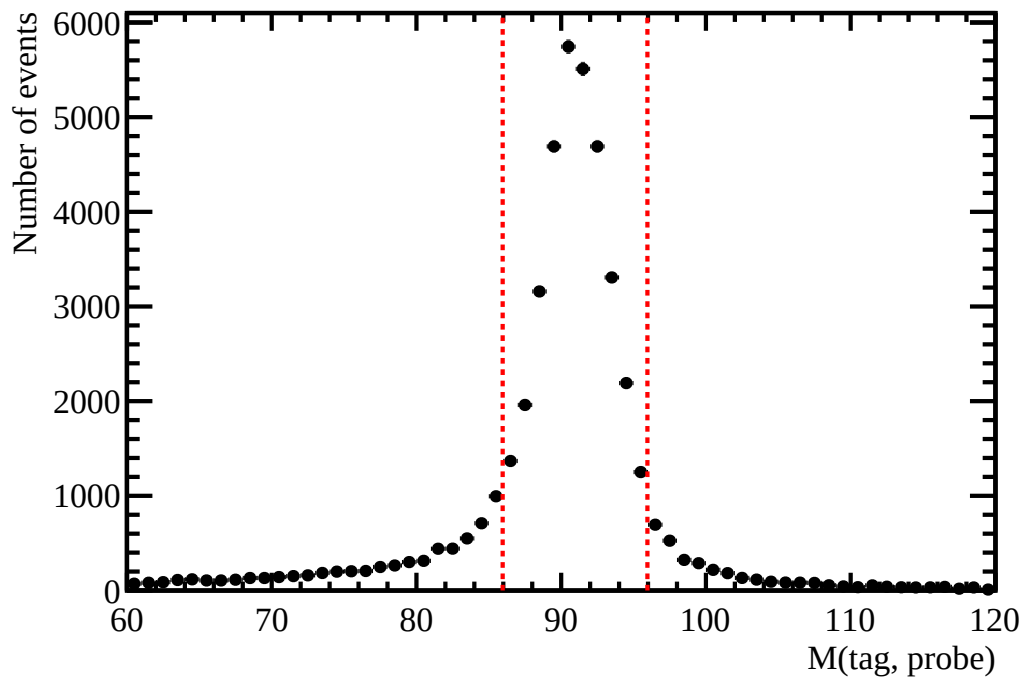


Figure 5.18: The invariant mass distribution of the tag and probe pairs considered when determining the muon identification efficiency is shown in black. The red lines show the invariant mass selection requirements placed on the tag-and-probe pairs. The uncertainties shown are statistical.

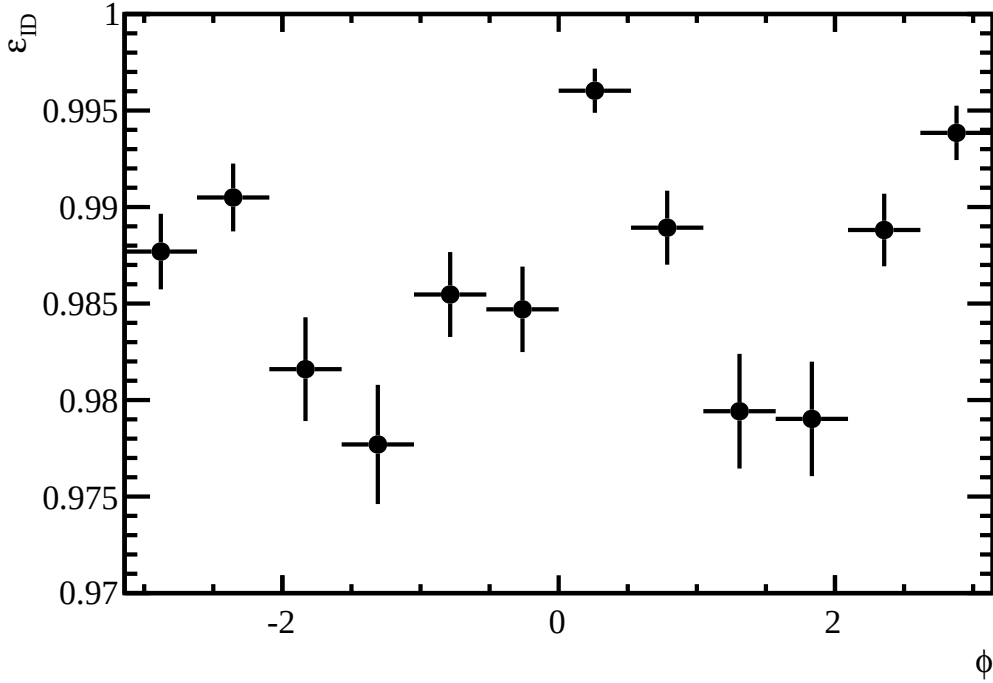


Figure 5.19: The muon identification efficiency, binned in the muon ϕ . The uncertainties shown are statistical.

ϕ between the tag and probe muons [159] when the efficiency is not binned in ϕ . However, the expected effect of such correlations is small. Indeed, biases could be introduced through any potential correlations between the tag and probe particles. For example, the kinematic requirements placed on the tag muon could bias the kinematic distributions of the probe muon, also potentially biasing the measured efficiency.

We therefore investigate any possible bias in the method by performing the same tag-and-probe analysis to determine the muon identification efficiency on simulated data. This efficiency is then compared with the muon identification efficiency found using truth level information. These two efficiency determinations were made using statistically independent Monte Carlo (MC) samples to enable ease of interpretation. The results are shown in Figure 5.20. Any differences between the two methods in the main region of the detector are not statistically significant. In the region $4.25 < \eta < 4.5$, there is possibly a larger discrepancy, but when we apply the tag-and-probe method on data in this region, the statistical uncertainty is much larger than the possible discrepancy. We therefore conclude that there is no evidence for any significant bias in our determination of the muon identification efficiency using the tag-and-probe method.

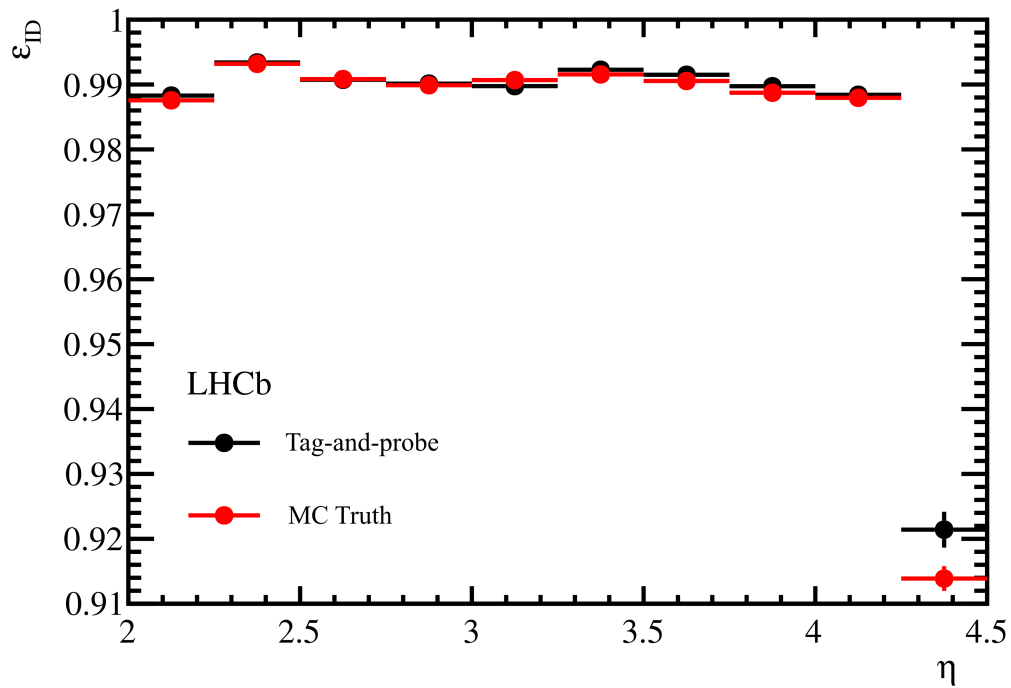


Figure 5.20: The muon identification efficiency measured on statistically independent MC samples using the tag-and-probe method, and directly from MC truth information. The bottom plot is the same as the top plot, but zoomed in on the region of interest. The uncertainties shown are statistical.

5.4.2 Muon track reconstruction efficiency

We apply a similar tag-and-probe analysis to that described above to determine the muon track reconstruction efficiency. This efficiency is the probability that a muon in the detector acceptance gives a reconstructed track that passes the track quality requirements. The tag is a long track which passes the ISMUON requirement, whilst the probe is composed of a well reconstructed track in the muon stations that is linked to hits in the TT. We consider how often the muon-TT probe is matched to a long track that passes the track requirements we consider in the nominal selection: $\text{Prob}(\chi^2, \text{dof}) > 0.001$, $\sigma_p/|p| < 0.1$. The relative number of times that a long track which passes the selection requirements is matched to a muon-TT probe track sets the muon track reconstruction efficiency. The matching is performed by requiring that the long track and the muon-TT track share detector hits and are within a radius of 0.2 of each other in $\eta - \phi$ space [160] and share hits in the muon stations. This method is the standard LHCb approach to determine the track reconstruction efficiency for muons, and more detail is provided in Ref. [160].

The selection used to measure the efficiency is:

- Requirements placed on both tag and probe particles:
 - $80 < M(\text{tag}, \text{probe}) < 100 \text{ GeV}$,
 - $2.0 < \eta < 4.5$,
 - $p_T > 20 \text{ GeV}$,
 - $|\Delta\phi(\mu^+\mu^-)| > 0.1$,
- Requirements placed on the tag particle:
 - passes the ISMUON requirement.
 - $p_T > 30 \text{ GeV}$,
 - TOS on the muon trigger lines,
 - The χ^2 probability associated with the track fit must satisfy $\text{Prob}(\chi^2, \text{dof}) > 0.001$,
 - $\sigma_p/|p| < 0.1$.

The invariant mass distribution of tag and probe pairs is shown in Fig. 5.21. The distribution does not show a sharp peak at the Z boson invariant mass as the momentum resolution of the muon-TT probe track is not as good as for long tracks. Since there is little advantage from a tight selection requirement, the invariant mass requirement is fairly loose.

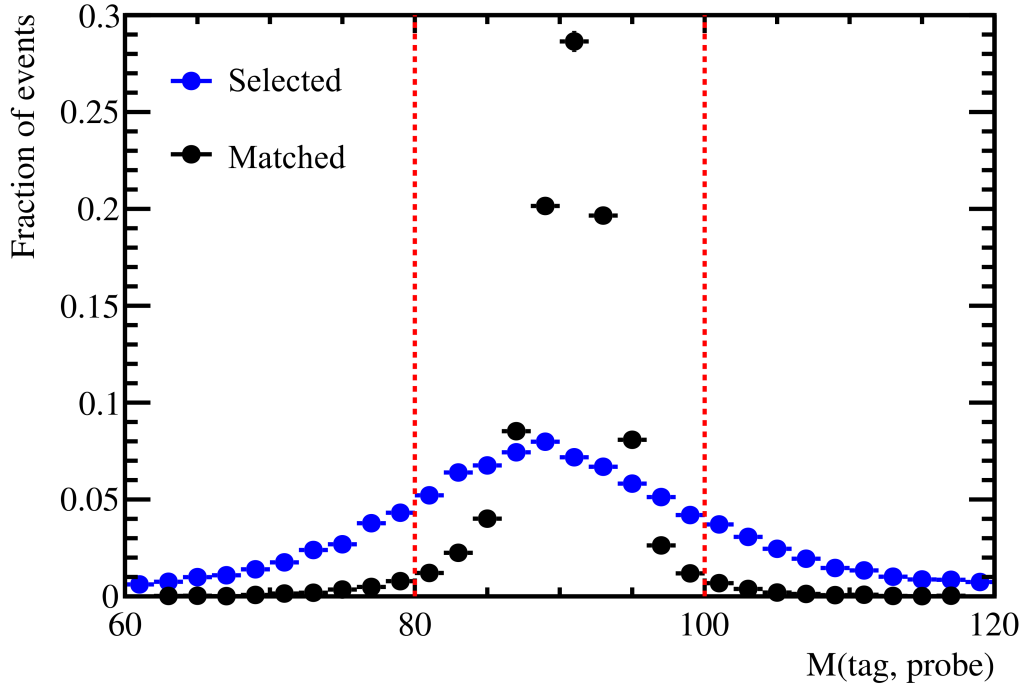


Figure 5.21: The invariant mass distribution of the tag and probe pairs considered when determining the track efficiency is shown in blue. Where there is a matched track, the invariant mass distribution of the matched track and the tag muon is shown in black. The dashed red lines show the values we select between for the tag and probe pairs. The uncertainties shown are statistical.

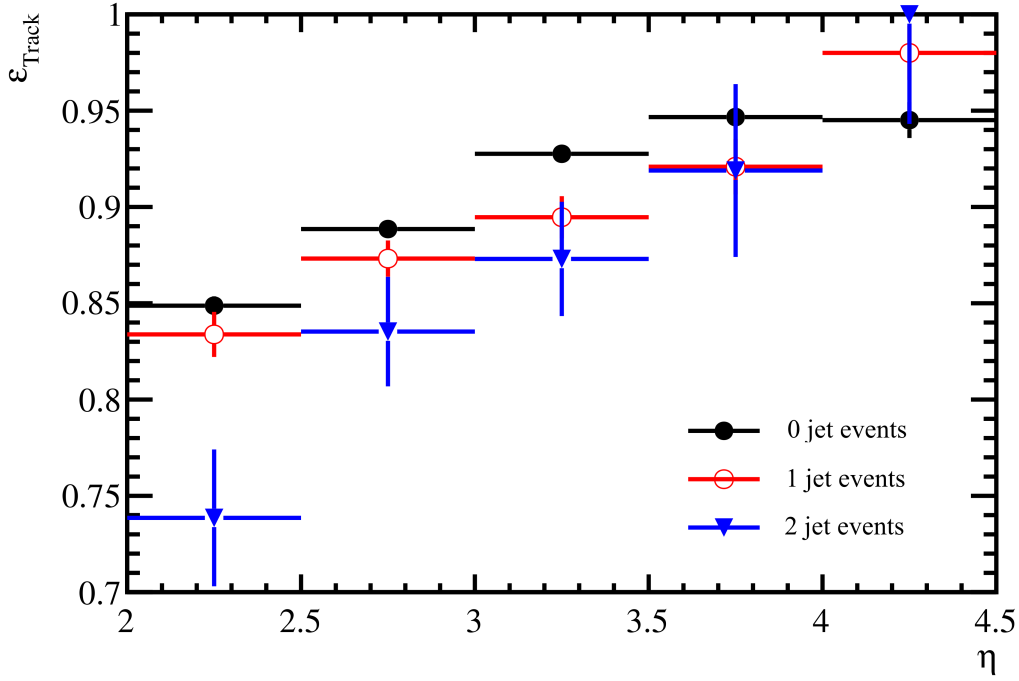


Figure 5.22: The muon track reconstruction efficiency, binned in the (exclusive) jet multiplicity and muon η . The uncertainties shown are statistical.

The measured track reconstruction efficiency is binned in jet multiplicity and muon pseudorapidity (shown in Fig. 5.22). Integrating over the binning in jet multiplicity, the efficiencies agree well with those reported by others [161]. The decrease in the track reconstruction efficiency at low muon pseudorapidity is due to the alignment being less well understood in this region. Consequently the χ^2 values associated with tracks in this region tend to be larger, and fewer tracks pass the selection requirements. The presence of additional hits in tracking subdetectors is also known to reduce the track reconstruction efficiency [160]. This is observed: the presence of a jet in the event is seen to decrease the track reconstruction efficiency by about 2%.

No evidence is found for a change in the track reconstruction efficiency associated with the different magnet polarities, or for different efficiencies associated with positive and negatively charged muons. As with the muon identification efficiency, the track reconstruction efficiency determination is repeated using with a wider mass window ($70 < M(\text{tag, probe}) < 110$ GeV) in the selection requirements, to test the effect of possible background contamination. Any possible effect is below the 0.6% level, and the difference is assigned as a systematic uncertainty.

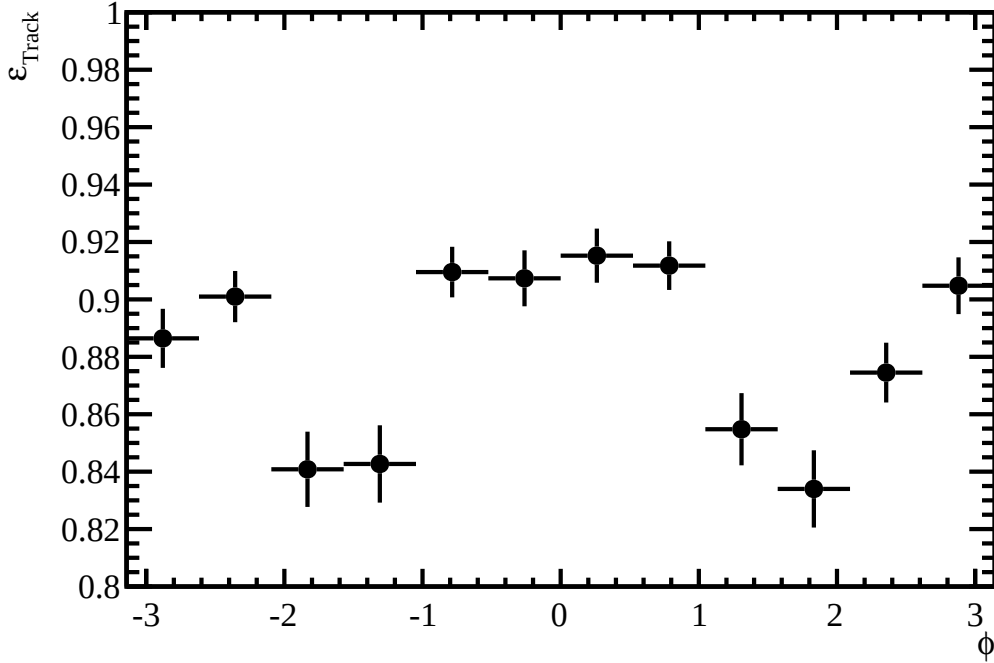


Figure 5.23: The muon track reconstruction efficiency, binned in the muon ϕ . The uncertainties shown are statistical.

We show in Fig 5.23 the muon track reconstruction efficiency as a function of the azimuthal angle of the muon in the detector, ϕ . As with the muon identification efficiency, there is a small drop in the track reconstruction efficiency, about 5%, noticeable at $\phi = \pm\pi/2$. This is associated with the change in ‘sides’ of the detector.

Validation of method

As with the identification efficiency, we check the possibility that there is a bias in the tag-and-probe method by applying it to simulation, and comparing the measured efficiencies to those found on a statistically independent sample from MC truth information. In calculating this efficiency we neglect the effect of the track quality requirements, as these will be considered separately below. The track reconstruction efficiency, found using the two independent methods, is shown as a function of the muon pseudorapidity distribution in Fig. 5.24. We see that the tag-and-probe method yields similar efficiencies to those found using truth information. However, there is a clear disagreement between the efficiencies at a level $> 1\sigma$. The mean absolute offset between the two methods is 0.3%. We therefore conclude that the tag-and-probe method we use to determine

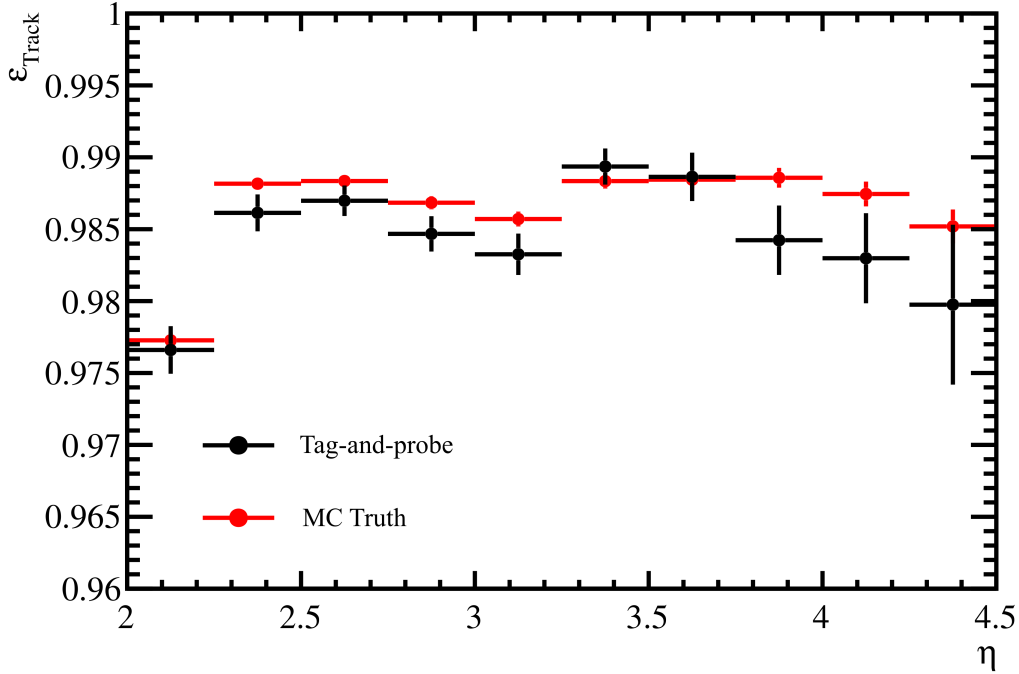


Figure 5.24: The muon track reconstruction efficiency, measured on simulation, using the tag-and-probe method, and from MC truth information. The uncertainties shown are statistical.

the track reconstruction efficiency is accurate to the level of 0.3% (per track). This is consistent with other studies [156], and provides a contribution to the systematic uncertainty associated with the track reconstruction efficiency.

In addition, we need to consider whether the method used to determine the track reconstruction efficiency introduces a bias into the distributions used to select high quality tracks. If the χ^2 probability distribution¹⁰ is different between the matched tracks and those selected in the standard analysis using the nominal selection, then the fraction failing the selection requirements will be incorrectly measured: a bias will be introduced into the efficiency measurement. The probability χ^2 distribution is therefore compared between all tracks in the nominal selection and the tracks which are considered when measuring the track reconstruction efficiency. No significant differences are seen between the distributions: the selection of events used to measure the efficiency does not introduce a bias. The two distributions are shown, overlaid, in Figure 5.25.

¹⁰We neglect the $\sigma_p/|p|$ distribution since this has a negligible impact on the efficiency, only rejecting 0.08% of events.

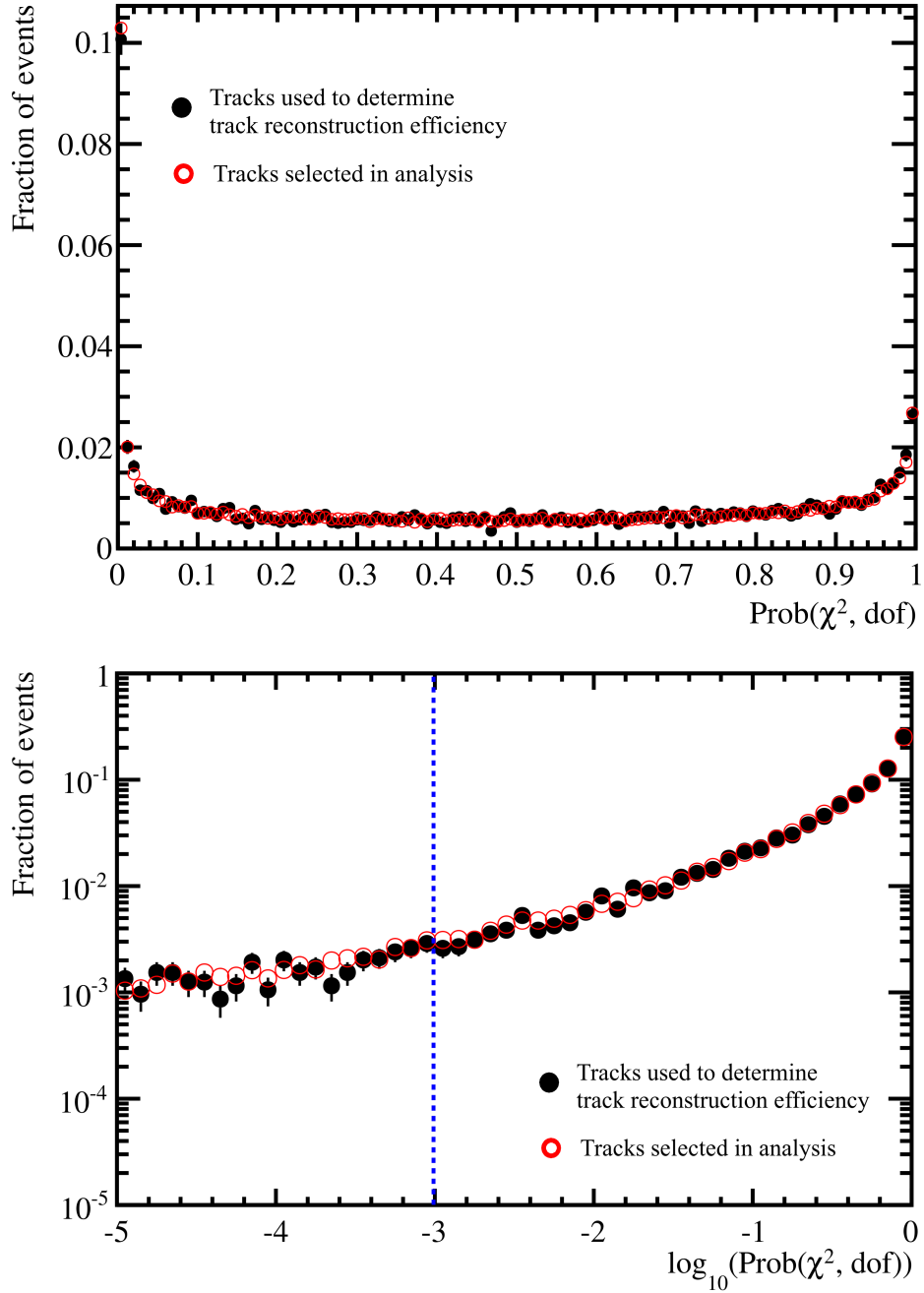


Figure 5.25: The χ^2 probability distribution for the tracks use to determine the track reconstruction efficiency, and the muon tracks found using the nominal selection (neglecting the χ^2 probability requirement). The bottom plot shows the same distribution on log scales, with the blue line showing the requirement made in the nominal selection. The agreement between the distributions continues for $\text{Prob}(\chi^2, \text{dof}) < 10^{-5}$. The uncertainties shown are statistical.

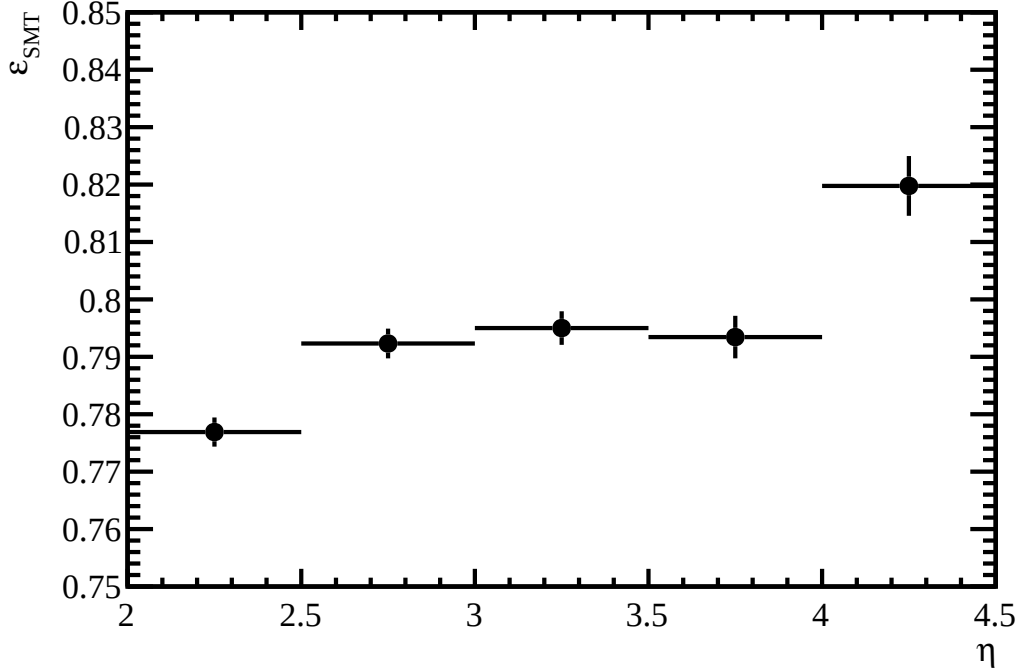


Figure 5.26: The single muon trigger efficiency, binned as a function of the muon η . The uncertainties shown are statistical.

5.4.3 Muon trigger efficiency

We also seek to determine the single muon trigger efficiency, ϵ_{SMT} . This is the probability that a muon passes our trigger requirements. We measure this using data, again using a tag-and-probe method. The same selection requirements are made to those outlined for the nominal selection for the Z+jet sample (see Section 5.2.2), except that the trigger requirements are changed. Here, the tag is required to be TOS on the muon triggers. The probability that the probe is also TOS on these trigger lines sets the trigger efficiency for a single muon to pass the muon lines.

This efficiency is shown as a function of the muon pseudorapidity in Fig. 5.26. There is no obvious dependence of the efficiency on the jet multiplicity, the muon p_T , the muon charge, or the magnet polarity. The efficiency calculation is also repeated for a tighter dimuon invariant mass selection ($80 < M^{\mu^+\mu^-} < 100$ GeV). The efficiencies found with this selection are consistent with those found using the wider mass window to within 0.1%.

Having found the single muon trigger efficiency, we are able to calculate the efficiency for either muon in our selection passing the single muon triggers. This overall trigger

efficiency ($\varepsilon_{\text{Trigger}}$) is given by

$$\varepsilon_{\text{Trigger}}(\mu_1, \mu_2) = \varepsilon_{\text{SMT}}(\mu_1) + \varepsilon_{\text{SMT}}(\mu_2) - \varepsilon_{\text{SMT}}(\mu_1) \cdot \varepsilon_{\text{SMT}}(\mu_2), \quad (5.1)$$

where $\varepsilon_{\text{SMT}}(\mu_n)$ is the single muon trigger efficiency associated with the kinematics of the n^{th} decay muon of the Z boson. This combination assumes that the single muon trigger efficiencies associated with each individual muon in the event are uncorrelated: triggering on one muon does not change the probability of triggering on the other.

It is also worth noting here that this combination reduces any dependence of the overall trigger efficiency on any other variables. If the single muon trigger efficiency was to change by $\sim 1\%$, then the overall trigger efficiency would change by only $\sim 0.4\%$.

Validation of method

We check for a possible bias in the tag-and-probe method by validating the method on simulation. The trigger efficiency is found on simulated events, using the tag-and-probe method, and compared to the efficiency found from MC truth information. Again, for the two methods, statistically independent samples are used. No significant difference is seen between the two methods (see Fig. 5.27).

The assumption that the probability of triggering on each muon in the event is uncorrelated is also checked. The overall trigger efficiency on simulation is found directly (by finding the ratio of the number of events which pass the trigger requirements to the total number of events considered). This is then compared to the efficiency found by combining the single muon trigger efficiencies. These two methods of finding the overall trigger efficiency are applied to statistically independent simulation samples. The combination method and the overall method give consistent efficiencies (within one standard deviation), so we conclude that the method of combining the single muon trigger efficiencies into an overall efficiency is valid, and that the probabilities of triggering on either of the muon candidates in the event are uncorrelated.

5.4.4 Global event cut efficiency

The lowest level trigger at LHCb also includes global event cuts (GEC) that are designed to reject very high occupancy events that will take too long to process. It is clear that the GEC efficiency will have a dependence on the jet multiplicity. The greater the number

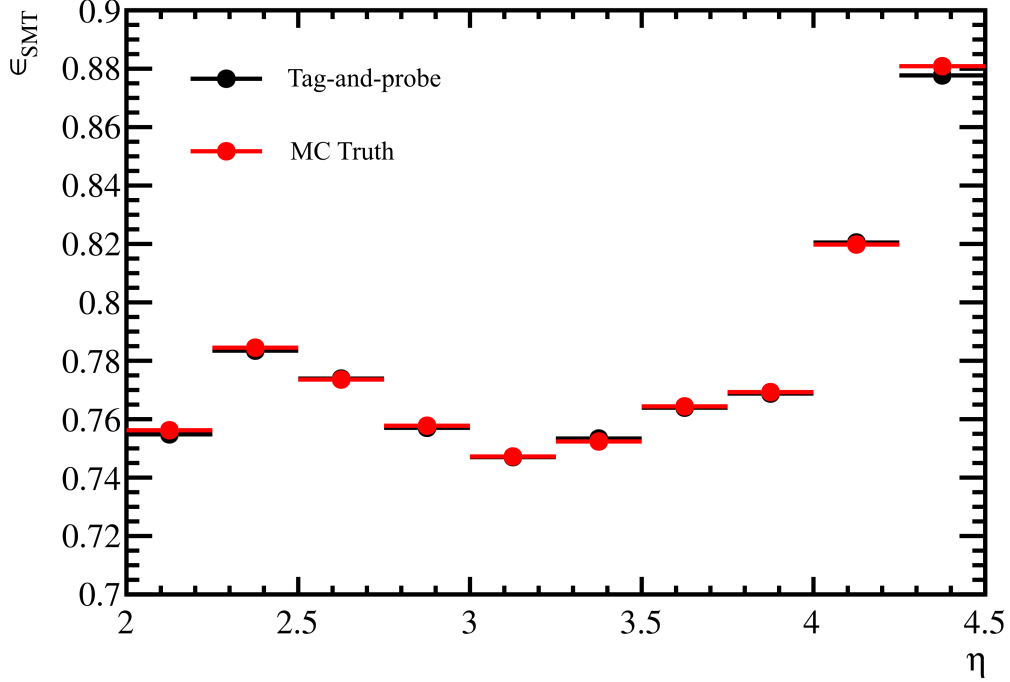


Figure 5.27: The single muon trigger efficiency, measured on simulation, using the tag-and-probe method, and from MC truth information. The uncertainties shown are statistical.

of jets in the event, the greater the occupancy, and the more likely it is that the event will not pass the GEC.

The trigger path used in this analysis contains a GEC requirement placed on the number of hits in the SPD: $n_{\text{SPD}} < 600$. However, there is an alternative path possible since trigger lines requiring DiMuon final states are available at each level of the trigger (L0, HLT1 and HLT2). This trigger path is not used in the nominal selection as it has a lower efficiency than the single muon trigger. However, the DiMuon trigger has a looser GEC requirement, $n_{\text{SPD}} < 900$, and we can therefore use it to estimate the number of events rejected with $600 < n_{\text{SPD}} < 900$. The distribution of n_{SPD} agrees very well between the two different trigger paths in the region where both paths select events ($n_{\text{SPD}} < 600$); the n_{SPD} distributions for both trigger routes are overlaid and shown in Fig. 5.28.

We also need to consider the number of events we reject with $n_{\text{SPD}} > 900$. This is a small contribution, which we determine by fitting the n_{SPD} distribution (for the DiMuon

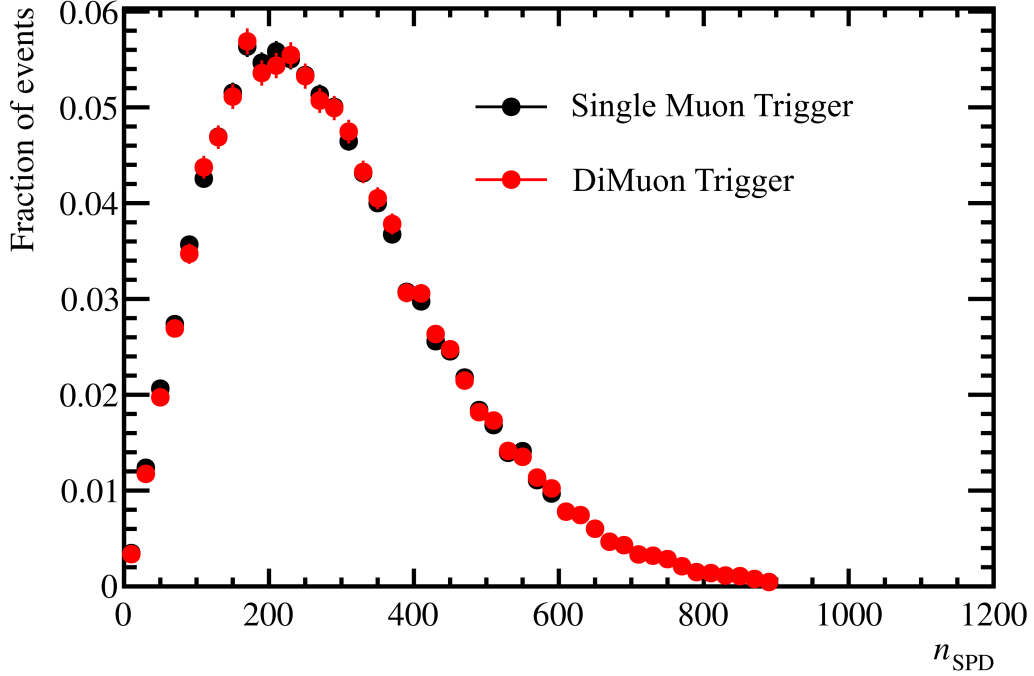


Figure 5.28: The distribution of n_{SPD} for the two different trigger routes. The uncertainties shown are statistical.

trigger route) with a gamma function [162], as in Ref. [6]:

$$y = A \cdot n_{\text{SPD}}^{\alpha-1} \cdot \exp(-n_{\text{SPD}}/\beta), \quad (5.2)$$

where A , α and β are the parameters that are used to fit y , the fraction of events. The fit is performed using the binned-likelihood method. An example fit is shown in Fig. 5.29. The fit result is then extrapolated to calculate the expected fraction of events with $n_{\text{SPD}} > 900$. This fraction is typically much smaller than the fraction of events with $600 < n_{\text{SPD}} < 900$. As an alternative, we also fit the tail of the distribution with an exponential decay. The difference between efficiencies measured using the two different fits gives a measure of the uncertainty associated with the fit. This difference is typically much smaller than the fraction of events with $n_{\text{SPD}} > 900$, and is assigned as a systematic uncertainty on the efficiency.

These two steps together enable a determination of the fraction of events with $n_{\text{SPD}} > 600$ that do not pass the GEC in the required single muon trigger path. The efficiency is measured as a function of the number of primary vertices and the jet multiplicity and is recorded in Table 5.1 (the distribution of the number of primary

vertices in reconstructed Z+jet events is shown in Figure 5.30). Whilst the efficiency is seen to drop with increased jet multiplicity, few events contain more than one jet.

No Z+jet events are observed to contain zero reconstructed primary vertices; there are only 59 events with zero primary vertices in the entire nominal $Z \rightarrow \mu^+ \mu^-$ sample. Studies using simulated data (without global event cuts applied) show no Z+jet events containing zero primary vertices, so the lack of events in recorded data is not associated with a significant detection inefficiency (instead, the presence of a Z boson and a jet in an event typically ensure that a primary vertex is reconstructed). No events containing only one primary vertex are seen to contain $n_{\text{SPD}} > 600$, so this efficiency is set to unity, and an appropriate uncertainty is assigned using the Feldman-Cousins approach [158]. The mean GEC efficiency for the Z+jet sample is $91.3 \pm 0.1\%$.

The global event cut efficiency shows no significant dependence on the kinematic parameters of the muons [155]. The global event cut efficiency was also studied as a function of the jet p_{T} . The efficiencies found when binning in the jet p_{T} were consistent with each other, within the statistical precision achievable (1-2%), once the dependence on the jet multiplicity was accounted for.

Validation of extrapolation method

The validity of fitting the n_{SPD} distribution to determine the global event cut efficiency is checked by performing a similar fit to the same distribution between $200 < n_{\text{SPD}} < 450$, for events containing 1 primary vertex. Again, the fit function used is the gamma function outlined in equation 5.2. This fit is then used to estimate the number of events with $n_{\text{SPD}} > 450$. However, this quantity can also be directly measured. The fit suggests that there would be 97 such events, whereas 88 events are observed. These two numbers are compatible within Poisson fluctuations. We therefore have confidence in the extrapolation of the fit.

5.5 Summary

We have considered the selection of Z+jet events at LHCb, for two jet transverse momentum thresholds, 10 and 20 GeV. We measure roughly 10000 and 4000 events containing jets with above these respective thresholds. The background levels are seen to

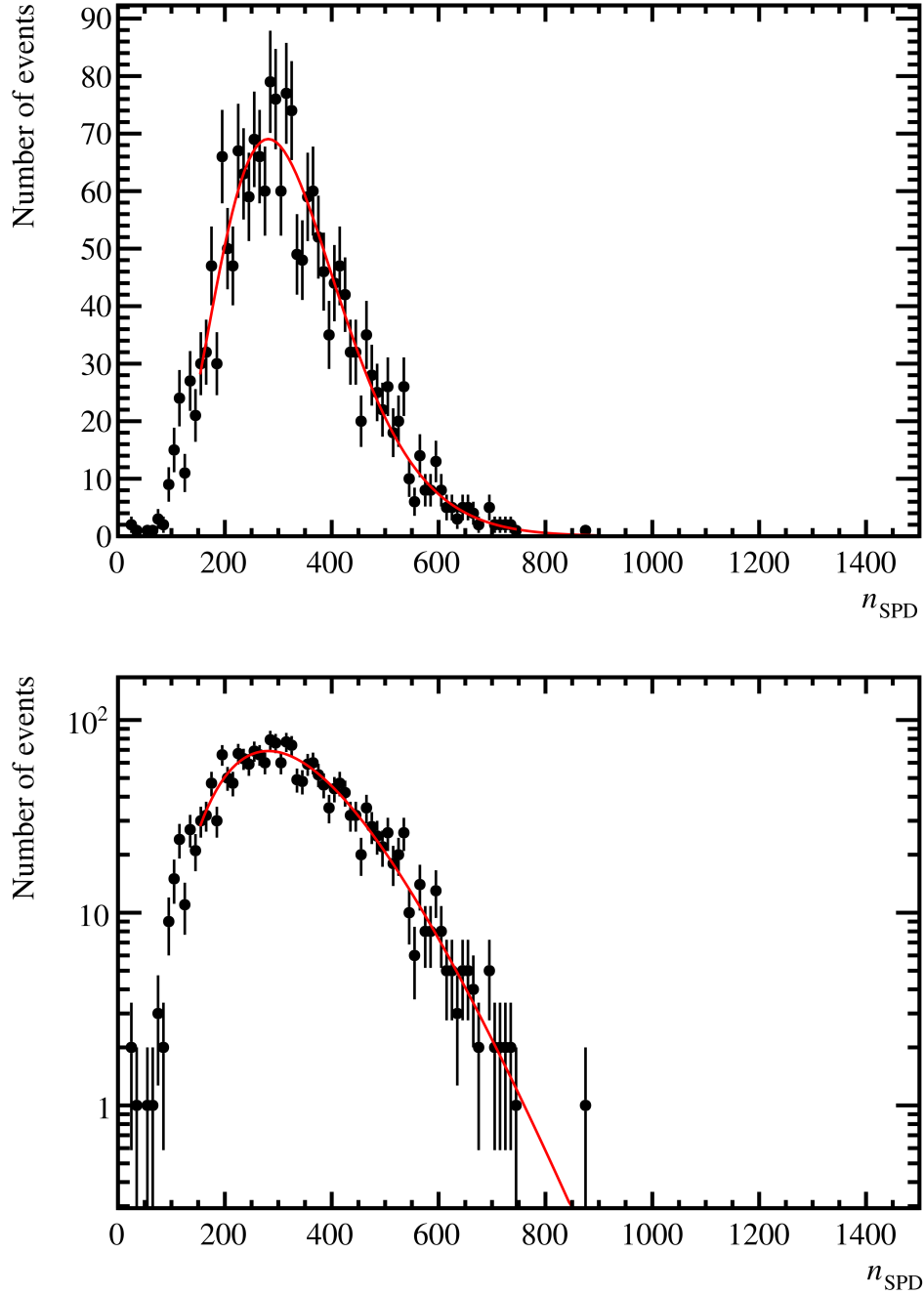


Figure 5.29: An example of the gamma function fit used to determine the number of candidates that would have had more than 900 SPD hits. This plot shows events with a jet multiplicity (for jets with $p_{\text{T}} > 10$ GeV) of 1, and with 2 reconstructed primary vertices. The uncertainties shown are statistical. The bottom plot shows the graph on a logarithmic scale. In this case, it is clear that the effect of events with $\text{SPD} > 900$ is very small, and the change in measured efficiency with the change in fit function is small. The fit shown in the plot is performed using the binned-likelihood method. A compatible fit using the minimisation of χ^2 method returns a best fit value of $\chi^2/\text{dof} = 60/47$.

	Jet Multiplicity			
	0	1	2	≥ 3
n_{PV}	0	$1.00^{+0.00}_{-0.02}$	-	-
	1	$1.0000^{+0.0000}_{-0.0001}$	$1.0000^{+0.0000}_{-0.001}$	$1.00^{+0.00}_{-0.02}$
	2	0.9932 ± 0.0008 (0.9931)	0.974 ± 0.003 (0.974)	0.924 ± 0.015 (0.920)
	≥ 3	0.895 ± 0.004 (0.894)	0.782 ± 0.011 (0.772)	0.59 ± 0.03 (0.56)

Table 5.1: The global event cut efficiency, measured as a function of the number of primary vertices (n_{PV}) and the jet multiplicity (for jet $p_{\text{T}} > 10$ GeV). The main numbers shown are for the fits to the n_{SPD} distribution using a gamma function, whilst those in brackets correspond to fitting the distribution using an exponential decay. As can be seen, the effect of changing the fit is small in most bins. This is for two reasons: the number of events with $n_{\text{SPD}} > 900$ is small, and the n_{SPD} distributions are well described by both gamma functions and exponential decays. The number of bins used is determined by the number of events in each bin - there are too few events to bin more finely. This means that we combine the 2 and ≥ 3 primary vertex bins together for events with jet multiplicity ≥ 3 .

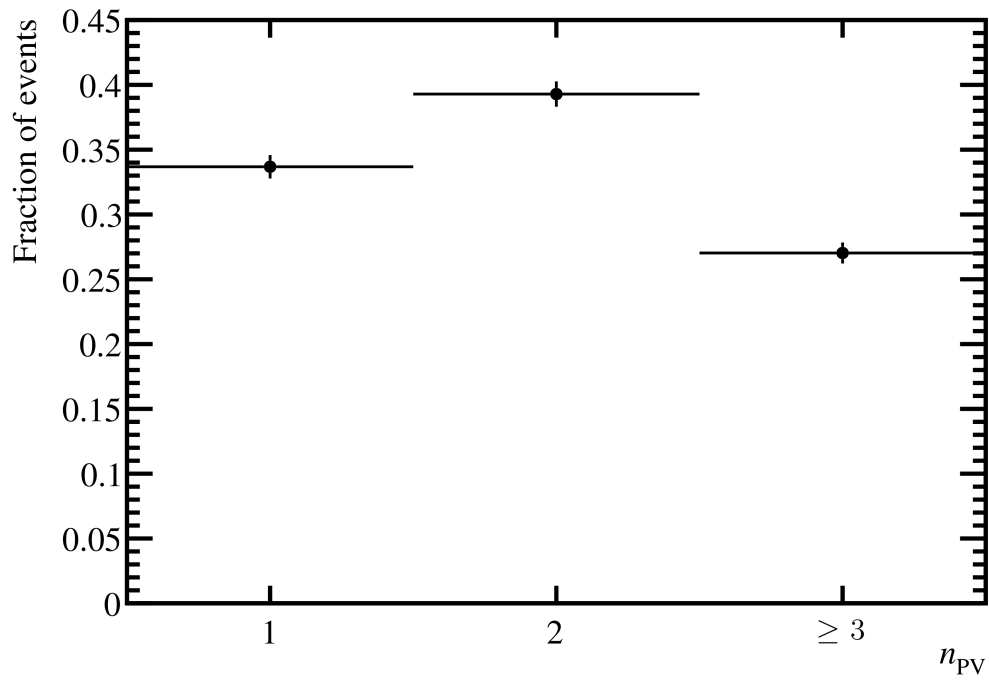


Figure 5.30: The distribution of the number of primary vertices in the reconstructed Z+jet events. The uncertainties shown are statistical. The distribution falls rapidly for events with more than 3 primary vertices.

be low ($\sim 0.3\%$). The efficiency with which Z bosons are detected and pass the nominal selection was measured as a function of the kinematic parameters of the decay muons. This detection efficiency is typically around 70%, and is dominated by the ability of the LHCb detector to reconstruct and select the tracks formed by the decay products of the Z boson (typically 90% per track).

The next chapter will focus on taking these events and efficiencies and measuring cross-sections for the production of Z+jet events at LHCb.

Chapter 6

Measuring the Z+jet production cross-section at LHCb at $\sqrt{s} = 7$ TeV

“When you can measure what you are speaking about, and express it in numbers, you know something about it.”

— Lord Kelvin

In this chapter we discuss the measurement of the Z+jet production cross-section, building significantly on the work discussed in previous chapters. The work of the author in this chapter is also documented in Ref. [4]. We consider the challenges of accurately measuring the jet p_T spectrum, due to the resolution effects discussed in Chapter 4. In addition, we correct the distributions shown in Chapter 5 for Z and jet reconstruction efficiencies, enabling the measurement. We also focus our attention on the systematic uncertainties associated with the measurement. Before we discuss these matters, however, we begin by defining the distributions we will measure, and the theoretical acceptance that these measurements correspond to.

6.1 Defining the measurement

We measure the Z+jet production cross-section and the ratio of this cross-section to the inclusive Z boson production cross-section. In addition, we measure the differential cross-sections for Z+jet production. The differential measurements are made as a function of the transverse momentum and pseudorapidity of the leading jet in the event, the Z

boson transverse momentum and rapidity, and the difference in rapidity and the difference in azimuthal angle between the Z boson and the leading jet. The data are compared to $\mathcal{O}(\alpha_s^2)$ and $\mathcal{O}(\alpha_s)$ predictions from POWHEG [7–11] using different PDF parametrisations that are matched to PYTHIA [12] for simulation of the parton shower development and hadronisation. These predictions will be discussed in Section 6.6.

We choose to normalise the differential measurements to the total Z+jet cross-section, so that the integral of each differential distribution is equal to unity. This choice reduces the impact of some systematic uncertainties (like the luminosity), and also enables us to concentrate on the shape of the differential distributions measured while the normalisation is probed by the measurement of the total cross-section. The total cross-section and cross-section ratio are measured at the Born level; a correction is applied to remove the effects on the cross-section from FSR associated with the final state muons. This correction will be discussed in detail in Section 6.3.1. This correction enables direct comparison of the result to the measurement of inclusive Z boson production [5]. However, no correction is applied to the shapes of the differential distributions, since the programs providing theoretical predictions can be configured to include the effects of FSR.

In our case the fiducial acceptance matches very closely the kinematic selection requirements placed on the events in Chapter 5¹. The fiducial selection is defined for the final state muons from the Z boson decay by:

- $p_T^\mu \geq 20$ GeV,
- $2.0 \leq \eta^\mu \leq 4.5$,
- $60 < M^{\mu\mu} < 120$ GeV.

This fiducial acceptance is the same as that used for the measurement of the inclusive Z cross-section [5]. The final state jets, defined at the hadron level, using the anti- k_T algorithm [52] with $R = 0.5$, are required to have:

- $2.0 \leq \eta^{\text{jet}} \leq 4.5$,
- $p_T^{\text{jet}} \geq 10(20)$ GeV,
- $\Delta R(\text{jet}, \mu) > 0.4$.

¹However, there is no *a priori* reason that the selection requirements should be the same as the fiducial cross-section: for example, it is possible to consider having loose selection requirements, and correcting to a tighter fiducial acceptance using some correction factor (as was done in Ref. [6]).

6.2 Jet transverse momentum resolution and jet detection efficiencies

Before we consider the Z+jet cross-section, we turn our focus to the ability of LHCb to accurately reconstruct jets, and the impact any inaccuracies have on the measurement of the Z+jet production cross-section. We consider different effects which mean that the hardest reconstructed jet p_T may not fall in the same bin as the hardest true jet p_T in the same event.

The jet p_T resolution smears the true jet p_T distribution. The resolution was discussed in Section 4.4, and is typically 11-13%. This causes migrations in the jet p_T distribution between true jet p_T bins and reconstructed jet p_T bins. This is the dominant migration effect on the jet p_T distribution.

In addition, the leading jet may not be reconstructed, or if it is reconstructed, it may fail the jet identification requirements. If there is no other jet in the event, then the event will not be selected. This shows up as an inefficiency in jet detection. However, if the event contains an additional jet, which is reconstructed and passes the identification requirements, then the net effect will be a migration in the jet p_T distribution between the true and reconstructed levels. This latter effect is small: fewer than 0.5% of Z+jet events contain two or more jets where the leading jet is not reconstructed and identified.

Another similar effect can also influence the jet p_T distribution. If two jets in an event have similar p_T , then the wrong jet may be reconstructed as the hardest jet in the event. This will lead to a migration in the jet p_T distribution between the true and reconstructed levels. This effect is also small: first, the jets must have similar p_T , which reduces the effect of any migration. Second, the probability of having two similar p_T jets in the same event is itself small: typically 3% of Z+jet events with one jet in the range $20 < p_T < 30$ GeV contain a second jet in the event with similar p_T . Fewer than half of these events will have the second jet incorrectly identified as the hardest jet.

In addition, there is the possibility that jets just outside the LHCb acceptance might be partially reconstructed, with the direction of the jet incorrectly identified as inside the acceptance. However, studies of the jet direction resolution, discussed in Section 4.4, show that this effect should be small. Simulation studies of the migration across the edge of the acceptance $\eta = 2.0$ show that any possible effect is smaller than 1%. This migration across this border is an example of a wider class of events where there is no

jet at truth level in the event, but where a jet is reconstructed. Such jets are labelled as ‘fake jets’. Most fake jets are typically due to the presence of one poorly measured particle. The jet identification requirements reduce the fake rate to about 4% for low p_T jets, and 1% for high p_T jets.

The migrations in the leading jet p_T distribution between the truth level and the reconstructed level are shown in Figure 6.1. For most values of jet p_T , the probability of an event in a given truth bin being in the same reconstructed bin is $\sim 60 - 70\%^2$. We must undo the effect of these migrations, and measure the true jet p_T distribution (defined at the hadron level).

We select events where the reconstructed jet p_T is above some cut-off p_T^{cut} , and split the corrections for the differences between the true distribution and the reconstructed distribution into two stages. First we consider migrations in the jet transverse momentum distribution (correcting events where the truth jet transverse momentum bin is different from the reconstructed jet transverse momentum bin). We correct for these migrations by unfolding, as explained below, giving the ‘truth’ p_T distribution for the selected Z+jet events with $p_T^{\text{jet}}(\text{reco}) > p_T^{\text{cut}}$. Then we find the efficiency, ε_{jet} , to correct for the fraction of true events where no jet was reconstructed, or where the reconstructed jet p_T was below p_T^{cut} or where the reconstructed jet failed to pass the jet identification requirements. This efficiency measurement allows for the p_T distribution to be corrected for selection inefficiencies.

The threshold p_T^{cut} is chosen to be 7.5 GeV: we unfold the p_T distribution to find the true p_T distribution associated with the selected events with a reconstructed p_T above this threshold. Corrections for the fraction of true events that do not contain a reconstructed jet above the threshold are then found as efficiencies. An alternative approach would be to consider a 10 GeV threshold. However, many true events above 10 GeV may have migrated to a transverse momentum below 10 GeV and consequently be lost (indeed roughly half of events with true $p_T = 10$ GeV will migrate to have a reconstructed p_T below a 10 GeV threshold). Placing the threshold at 7.5 GeV enables us to consider more of these events within the unfolding. We consider this unfolding next, before considering the jet reconstruction and selection efficiencies in Section 6.2.2. The net effect of placing the threshold at 7.5 GeV is that the selection outlined in Section 5.2.2 and considered in Chapter 5 is modified for this chapter: the jet p_T requirement is now $p_T^{\text{jet}} > 7.5$ GeV. The selected events must still pass all the other selection criteria listed in Section 5.2.2.

²This is dependent on the bin widths chosen.

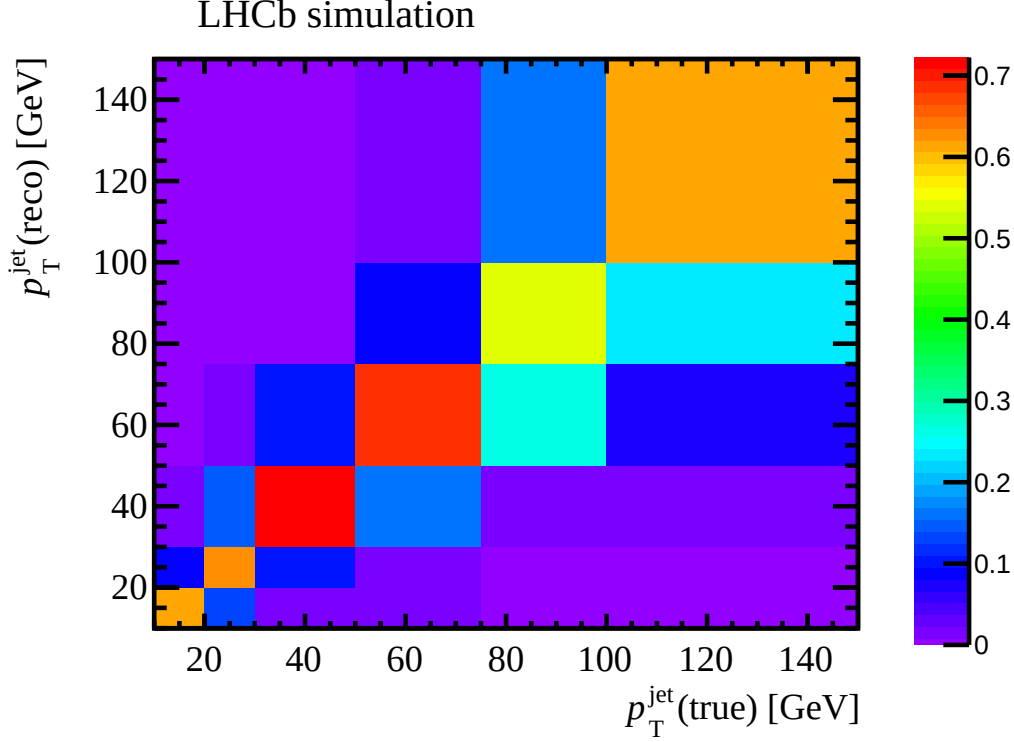


Figure 6.1: The probability that an event with a given true p_T of the leading jet (x axis) is reconstructed in a particular bin of the measured jet p_T (y -axis). Columns therefore sum to unity. This probability is shown by the colour scheme.

6.2.1 Unfolding the jet p_T distribution

We seek to correct the jet p_T distribution from the binned quantities we reconstruct to a reliable estimate of the corresponding binned distribution at truth level. This is typically known as unfolding the data.

We start by defining terminology. We refer to migrations when we mean the changes in the binned distribution between the truth and reconstructed levels. The matrix describing this, $\mathbf{R}$, is the response matrix, which is typically measured on simulation using some training sample. If the (binned) truth distribution is labelled $\mathbf{y}$ and the (binned) measured distribution is labelled $\mathbf{x}$ then we can write

$$\mathbf{x} = \mathbf{R}\mathbf{y}, \quad (6.1)$$

where it is clear that we can associate the elements of the response matrix with the probability that an event in a given true p_T bin is measured in a particular reconstructed

p_T bin. We use the term corrections to refer to the changes we apply to the reconstructed distribution to measure an estimate of the true distribution. This process of correcting the reconstructed distribution is the unfolding, and the matrix for these corrections, $\mathbf{U}$, is the unfolding matrix. We can calculate the unfolded distribution, $\mathbf{u}$ (our estimate of $\mathbf{y}$) from

$$\mathbf{u} = \mathbf{U} \mathbf{x}, \quad (6.2)$$

where the elements of the unfolding matrix contain the probabilities that events measured in bins of the reconstructed p_T distribution correspond to particular true p_T bins.

We therefore discuss below different methods to determine $\mathbf{U}$ and find the distribution $\mathbf{u}$ that provides a reliable estimate of $\mathbf{y}$. We then turn our attention to their application in our case.

Unfolding techniques

We could take the response matrix $\mathbf{R}$ describing the migrations between bins of p_T from simulation. The inverse of the response matrix will enable an estimate of the true distribution in data to be recovered from the reconstructed distribution: we could set the unfolding matrix $\mathbf{U} = \mathbf{R}^{-1}$, and $\mathbf{u} = \mathbf{R}^{-1} \mathbf{x}$. However, the inverse of $\mathbf{R}$ is not always well determined. If the response matrix has small off-diagonal elements, these elements may have a large statistical uncertainty. These elements become important when the inverse is found as they are used to calculate large elements in the inverse matrix, which will as a consequence have large uncertainties. This means that the inverse matrix we would use to correct the reconstructed distribution is not necessarily well determined; this method of determining the unfolding matrix yields large (anti)correlations between bins in the unfolded distribution.

Instead, the unfolding matrix $\mathbf{U}$ could be directly measured in simulation (without the need to invert the response matrix). This matrix contains elements which describe the probability of an event being in a particular bin of the true p_T distribution given the presence of the event in a particular bin of the reconstructed p_T distribution. However, this probability depends on the true distribution used in the simulation. This can be directly seen through Bayes' theorem [163]:

$$\begin{aligned}
 U_{ij} &= \text{Prob}(\text{true bin } i \mid \text{reco bin } j) \\
 &= \text{Prob}(\text{reco bin } j \mid \text{true bin } i) \frac{\text{Prob}(\text{true bin } i)}{\text{Prob}(\text{reco bin } j)} \\
 &= R_{ji} \frac{\text{Prob}(\text{true bin } i)}{\text{Prob}(\text{reco bin } j)},
 \end{aligned} \tag{6.3}$$

where we have chosen to identify³ the elements of the response matrix with the migration probabilities as $R_{ji} \equiv \text{Prob}(\text{reco bin } j \mid \text{true bin } i)$ and the elements of the unfolding matrix as $U_{ij} \equiv \text{Prob}(\text{true bin } i \mid \text{reco bin } j)$. Even if the response matrix is accurately simulated, it is clear that a direct determination of the unfolding matrix from simulation relies on the true distribution also being accurately simulated. Therefore simply taking the unfolding matrix directly from simulation does not guarantee an accurate result.

Various solutions to these problems have been proposed, based on different potential assumptions. We discuss the two most commonly used here: the D’Agostini/Bayesian unfolding [164] and the single value decomposition (SVD) [165] unfolding methods.

The D’Agostini approach takes the unfolding matrix from simulation, and uses this matrix to correct the reconstructed distribution. This yields a truth level distribution in the measured data. This distribution is not necessarily accurate, for reasons outlined earlier. However, this truth distribution can be used to reweight the unfolding matrix that was taken from simulation, effectively using equation (6.3), so that it better describes the recorded data⁴. This process can be repeated iteratively, with a new truth level distribution found for data with each iteration. A small number of iterations typically tends to give a reasonable result [166]; a large number of iterations provides a result which is identical with inverting the response matrix, containing large (anti)correlations between bins in the unfolded distribution and large uncertainties on the final result. Therefore, the iteration process must be stopped after an appropriate (small) number of iterations. The number of iterations is therefore a key parameter which regularises the unfolding procedure for the D’Agostini method, and is referred to as the ‘regularisation parameter’. If it is set to unity, then the unfolding matrix is not reweighted: the unfolding

³We use here the notation that the probability of an event a given knowledge that an event b has already occurred is $\text{Prob}(a|b)$. We also use the notation that elements of the matrix $\mathbf{M}$ are referred to as M_{ij} .

⁴This use of Bayes’ theorem gives the unfolding method its alternative name - ‘Bayesian unfolding’.

matrix is simply taken directly from simulation. Historically, the D’Agostini method has shown issues with the propagation of uncertainties. According to Ref. [166], the original paper by D’Agostini [164] neglected a term in the uncertainty propagation: after two iterations the unfolding matrix itself obtains an uncertainty associated with the reweighting procedure. The author of Ref. [166] has expanded on the original paper to include this term in the uncertainty propagation and demonstrated the accuracy of this approach by comparing the propagation with toy studies.

The SVD unfolding method provides an alternative approach to inverting the response matrix using the language of linear algebra [165]. The response matrix and the associated truth and reconstructed distributions form a set of linear equations [165], as shown in equation (6.1). The SVD approach rotates the co-ordinate system so that the response matrix is diagonal. This approach implicitly brings in the problems associated with directly inverting the response matrix. Some of the diagonal elements in the rotated matrix are not well determined as they are consistent with zero: the matrix is consistent with being singular [165]. The SVD approach to unfolding therefore sets these elements to zero, and rescales the linear equations⁵ so that the equations which are best determined carry the most weight [165]. The presence of a singular response matrix is equivalent to stating that the problem is under-constrained. The SVD approach takes other information to provide additional constraints [165]. Usually the condition that the truth distribution varies smoothly is applied, though due to the rescaling of the linear equations, this condition is entirely equivalent to the assumption that the differences between the training sample (used to determine the response matrix) and data vary smoothly [165]. The weight given to this assumption is set by some integer regularisation parameter, which typically varies between unity and the number of bins being considered [165]. The choice of regularisation parameter is important: a choice of unity returns an unfolded distribution identical to the truth distribution in the training sample, whereas the choice where the parameter equals the number of bins keeps many of the large statistical fluctuations associated with inverting the response matrix [166]. Intermediate values work best; a reasonable choice can be made by studying the system in question [166].

The two methods both take the response matrix and provide an unfolded distribution using different approaches to avoid the pitfalls associated with directly inverting the response matrix. They therefore provide independent approaches to unfold the reconstructed jet p_T distribution measured at LHCb.

⁵Rescaling linear equations will not change the solution obtained for non-singular problems. However, when external constraints are necessary and the solutions found are not exact, as in singular problems, the scaling has a noticeable effect. More detail on this is available in Ref. [165].

Unfolding the LHCb data

We unfold the LHCb data using the D’Agostini [164] and SVD [165] approaches discussed above, as implemented in the ROOUNFOLD package [166]. The training sample for the unfolding was the Z+jet simulation sample discussed in the previous chapter. The unfolding is performed using different binning schemes for the reconstructed and truth levels. At the reconstructed level, we consider the bins defined by the bin edges (and an inclusive high p_T bin) [7.5, 10, 15, 20, 25, 30, >30 GeV], whereas at the truth level we consider the bin edges (and inclusive high p_T bin) [0, 5, 10, 15, 20, 25, 30, >30 GeV]⁶. This unfolding automatically corrects for the presence of fake jets, correcting for events where there is no jet at truth level, but where there is a reconstructed jet.

The regularisation parameters to be used in the D’Agostini method and the SVD method were both determined by investigating the performance of the unfolding. An independent simulation sample was studied: this sample was based on the PYTHIA 6 [12] simulation of inclusive Z boson production⁷, with the simulation of the LHCb detector included. We will refer to this as the pseudo-data sample, as the sample was then treated like data and unfolded using the two approaches. This unfolding could then be used to estimate the total number of the pseudo-data events containing a jet with p_T above the 20 GeV threshold. This number was then compared to the known number in the pseudo-data sample using truth level information. The same number of events were used in the pseudo-data as were present in data. For the D’Agostini method the use of two iterations was found to give the best agreement between the unfolded distribution and the truth distribution in pseudo-data. For the SVD method, a regularisation parameter of three gave the best agreement. These choices of regularisation parameters also gave the best agreement in the pseudo-data sample between the number of unfolded events and the number of truth events containing a jet with p_T greater than 10 GeV.

When applied to the recorded LHCb data, the two methods agree impressively in the unfolded p_T distribution of the selected events, as shown in Figure 6.2. The numbers of events above the 20(10) GeV threshold agree within 0.3(0.2)%. This level of agreement is a consequence of the data and simulation agreeing well (as was shown in Figure 5.7). In addition, the effect of changing the number of iterations used when unfolding the data using the D’Agostini method is shown in Figure 6.3, where one iteration is used

⁶Two bins are used for $p_T(\text{true}) < 10$ GeV, as this region sees reasonably large variations, and it was felt such a binning scheme might be necessary. However, other binning choices here were found to give consistent results.

⁷In this simulation sample all jets are produced using the parton shower.

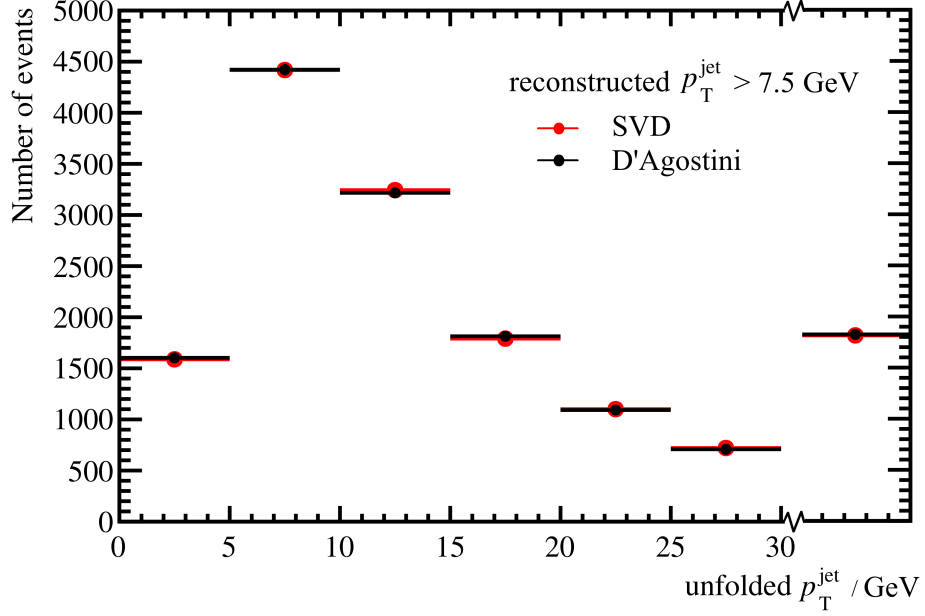


Figure 6.2: The unfolded jet p_T distribution. The black distribution uses the D'Agostini method for the unfolding, whereas the red distribution uses SVD. The final bin contains all events with $p_T^{\text{jet}} > 30$ GeV. The uncertainties shown are statistical.

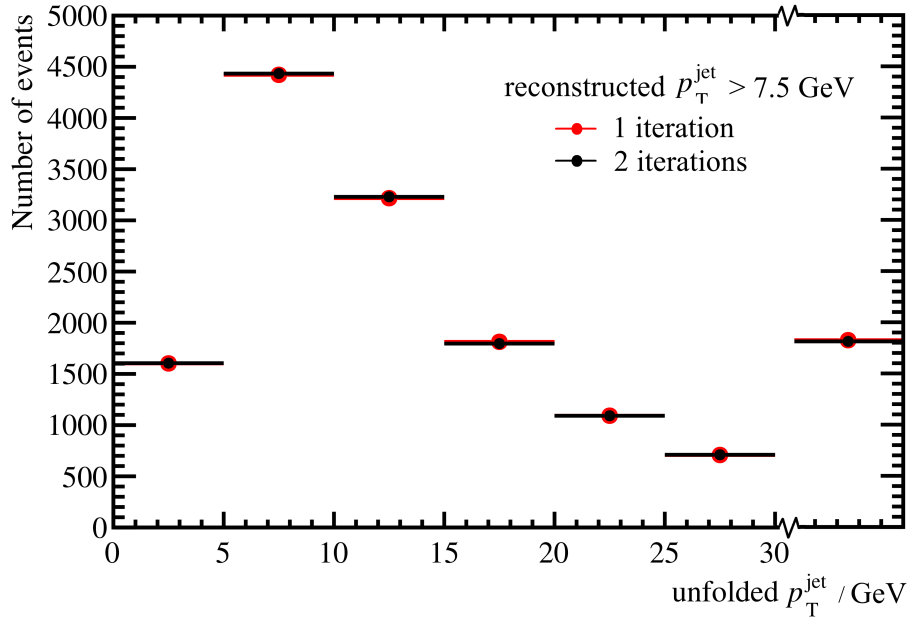


Figure 6.3: The unfolded jet p_T distribution. Both distributions are unfolded using the D'Agostini method, but with different numbers of iterations used to regularise the procedure. The final bin contains all events with $p_T^{\text{jet}} > 30$ GeV. The uncertainties shown are statistical.

and the resulting distribution is compared to the distribution found using two iterations. As discussed earlier, one iteration of the D'Agostini method corresponds to taking the unfolding matrix directly from simulation, and not reweighting it. Once again, excellent agreement is seen: the result is stable to this change in the number of iterations. This stability is again due to data and simulation already agreeing well. Repeating the unfolding using different binning schemes also gives consistent results.

6.2.2 Jet reconstruction and selection efficiency

We also need to correct for the events where there was a truth jet but no reconstructed (and selected) jet above the 7.5 GeV threshold, in which case the event was not considered using the migrations discussed above. This could be because there was no reconstructed jet in the event, or because the reconstructed jet migrated below the threshold, or because the reconstructed jet failed to pass the jet identification requirements. Explicitly, we define this jet reconstruction and selection efficiency, ε_{jet} , to be the probability that given a true jet in our acceptance, a jet is reconstructed in the event with $p_T > 7.5$ GeV, passing the jet identification selection.

This efficiency is directly measured from the Z+jet simulation, and is shown in Figure 6.4 as a function of the true jet p_T . We see a very strong dependence on p_T , due to two effects. At low p_T there is a larger probability that our jet p_T resolution can lead to us not reconstructing the jet with p_T above the 7.5 GeV threshold. The efficiency therefore drops at low p_T . In addition, at low p_T , more events fail the jet ID requirements. This is the larger of the two effects.

The efficiency is shown as a function of the jet pseudorapidity in Figure 6.5. The efficiency shows a variation of about 2-3% across the detector acceptance, with the largest effect seen closest to the edge of the acceptance. This is primarily due to the increased impact of the jet identification requirements in this region. Figure 4.6 shows that less of the jet content is reconstructed near the edge of the detector acceptance. This in turn increases the impact of the jet identification requirements, so the jet reconstruction and selection efficiency drops.

However, we know that simulation and data do not agree exactly (this was discussed in detail in the previous chapter). We therefore consider reweighting the events in simulation to better match the events measured in the recorded data. The effect of reweighting the jet multiplicity distribution to be the same in simulation as in the recorded data

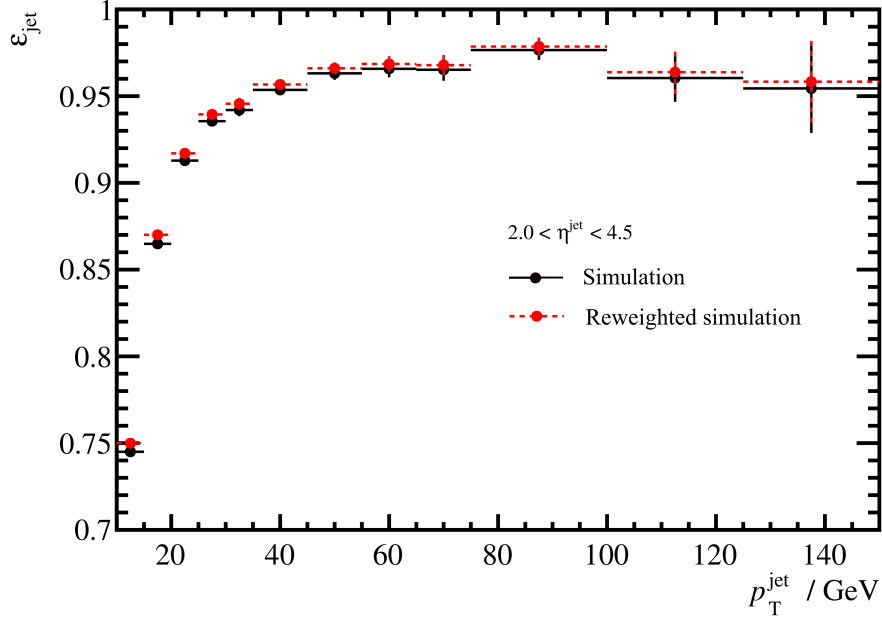


Figure 6.4: The jet reconstruction and selection efficiency as a function of the true jet p_T . The black distribution shows the efficiency evaluated directly using simulation. The red distribution shows this same efficiency reweighted so that the jet multiplicity in simulation is the same as the jet multiplicity in data. The uncertainties shown are statistical.

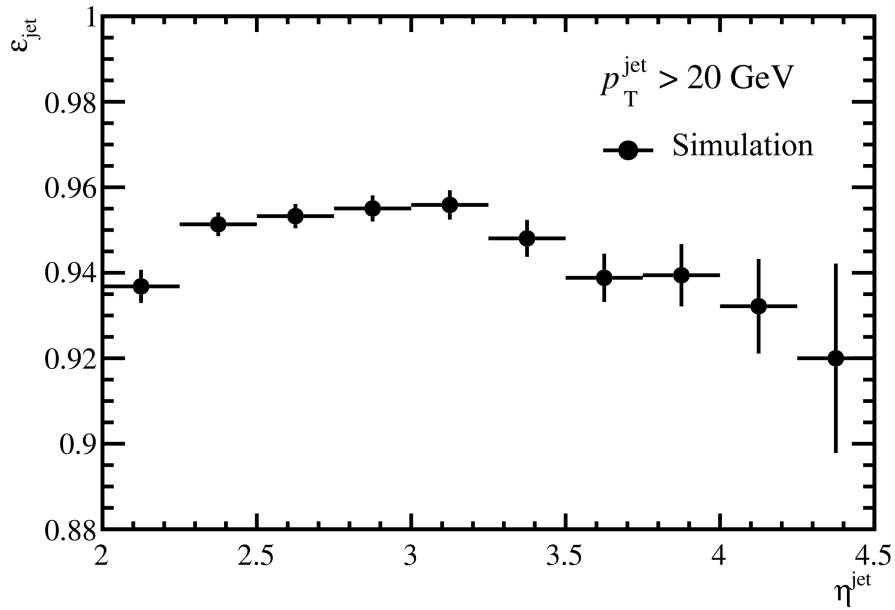


Figure 6.5: The jet reconstruction and selection efficiency as a function of the jet pseudorapidity. The reconstruction efficiency is highest in the centre of the LHCb detector, falling slightly close to the edge of the LHCb acceptance. The uncertainties shown are statistical.

is $\sim 0.5\%$, and is also shown in Figure 6.4. Since the efficiency is seen to have a small variation with the jet pseudorapidity, we also consider the effect of differences in the jet pseudorapidity distribution between data and simulation. If we reweight the events used to measure ε_{jet} so that the true jet pseudorapidity distribution in simulation is the same as the measured jet pseudorapidity distribution in data, then we see a negligible change in the efficiencies ($< 0.1\%$). The efficiencies found are also consistent with those found from the inclusive Z boson simulation sample.

There are also uncertainties associated with the degree to which simulation provides an accurate model for jets in the recorded data and therefore the degree to which the simulation provides an accurate determination of the jet reconstruction and selection efficiency. We will discuss this in Section 6.4.6, when we consider systematic uncertainties on the measurement of the Z+jet production cross-section.

6.3 Cross-section measurement

So far we have discussed the selection of Z+jet events, the background sources to this selection, the Z and jet selection and reconstruction efficiencies, and the effect of the jet transverse momentum resolution. We now take these different studies, and combine them, using this information to measure the Z+jet production cross-section.

We determine the number of events within each bin of the unfolded jet transverse momentum, $N(p_{\text{T}}^{\text{unf}})$, corrected for the Z detection efficiency⁸ ε_{Z} , determined on an event-by-event basis from the kinematics of the particular event:

$$N(p_{\text{T}}^{\text{unf}}) = \sum_{\text{selected Z+jet events}} \frac{U(p_{\text{T}}^{\text{unf}}, p_{\text{T}}^{\text{reco}})}{\varepsilon_{\text{Z}}}, \quad (6.4)$$

where $U(p_{\text{T}}^{\text{unf}}, p_{\text{T}}^{\text{reco}})$ is an element of the unfolding matrix $\mathbf{U}$: it determines the probability that an event containing a jet with reconstructed transverse momentum in the bin $p_{\text{T}}^{\text{reco}}$ contains a truth jet with transverse momentum in the bin $p_{\text{T}}^{\text{unf}}$. We determine the matrix using the D'Agostini scheme, with the SVD method considered earlier providing a cross-check. For the differential jet p_{T} distribution, a different binning scheme is used, with a larger number of bins at high p_{T} . The results from this binning scheme are

⁸Since the efficiencies associated with each event vary within a narrow range between about 65% and 75% the effect of using weighted events on the (fractional) statistical uncertainty is negligible, and is therefore neglected. This was confirmed using studies in ‘toy’ simulation.

consistent with the results using the standard scheme described above. For measurements of the other differential distributions the analysis is repeated separately in each bin of the relevant distribution. Consequently the unfolding matrix is re-determined for each bin in each distribution, only using events reconstructed in that bin⁹.

We then correct the $N(p_T^{\text{unf}})$ distribution to account for the jet reconstruction efficiency, ε_{jet} . For the differential measurements, as with the unfolding matrix, ε_{jet} is re-determined based on events in the relevant bin. In addition, for the differential measurements, a factor A_{mig} is applied to account for migration between different bins (for example, in the jet pseudorapidity distribution)¹⁰. This correction typically has a small net effect of less than 2-3% (see for example Figure 6.6), and is taken from simulation. Dividing by the integrated luminosity, $\int \mathcal{L} dt$, then determines the cross-section,

$$\sigma_{\text{Z+jet}} = \frac{\rho}{\int \mathcal{L} dt} \sum_{p_T^{\text{unf}} > p_T^{\text{thr}}} \frac{A_{\text{mig}}}{\varepsilon_{\text{jet}}} N(p_T^{\text{unf}}), \quad (6.5)$$

where p_T^{thr} is the relevant threshold, 20 or 10 GeV, and ρ is the purity of the sample, accounting for the presence of background.

Having found the total cross-section in each bin of a differential distribution, we then normalise the distribution to the total cross-section so that each differential distribution has an integral of unity. The integrals of the differential distributions (before this normalisation) are consistent with the total cross-section for all the distributions considered. Finally, the measurement of the overall Z+jet cross-section is corrected to the Born level: the effect of FSR from the decay muons of the Z boson is removed. This will be discussed below.

The work discussed so far also enables us to find the inclusive Z boson production cross-section (in the muon decay channel). This cross-section can be found by summing over all the selected Z boson events (i.e. neglecting the jet selection requirements) and evaluating

$$\sigma_{\text{Z}} = \frac{\rho}{\int \mathcal{L} dt} \sum_{\text{selected Z events}} \frac{1}{\varepsilon_{\text{Z}}}. \quad (6.6)$$

⁹We give an example to ensure clarity: for the jet η distribution, for the bin $2.0 < \eta^{\text{jet}} < 2.5$, we determine the unfolding matrix for the jet p_T solely using events reconstructed in that bin. We carry this out separately for each bin in the jet η distribution.

¹⁰This is not done for the differential p_T distribution, as the unfolding already accounts for these migrations.

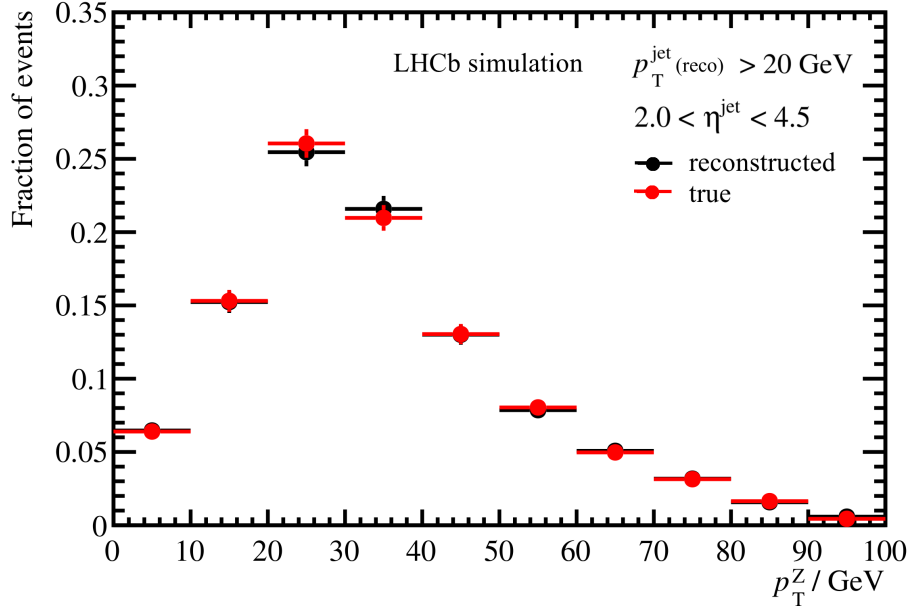


Figure 6.6: The reconstructed and true p_T^Z distributions in simulation, for events where the reconstructed $p_T^{\text{jet}} > 20$ GeV. This distribution has larger migrations between bins than the other measured distributions, save for the p_T^{jet} distribution. This is because angular variables like rapidities and $\Delta\phi$ tend to be better measured than momenta. However, even in this distribution the net effect of migrations between bins is clearly small (less than 2-3%); the reconstructed p_T^Z lies in the correct bin in the large majority of cases.

Comparison of this result with existing measurements at LHCb allows for a cross-check of the Z+jet analysis. It also lets us find the ratio of the Z+jet production cross-section to the inclusive Z boson production cross-section, for both jet transverse momentum thresholds.

6.3.1 Final State Radiation (FSR)

Charged particles radiate photons. This means that the particles we reconstruct do not strictly correspond to those in the Feynman diagrams shown in Fig. 5.4 or Fig. 5.5. Instead, we measure particles after they have radiated photons. The ingredients we have discussed so far allow us to make measurements at the ‘bare’ level, as we have not considered the effect of these photons¹¹.

In order to aid comparison with previous results for inclusive Z boson production at LHCb, we choose to correct the total cross-section measurement for Z+jet production, and the ratio of this cross-section with the inclusive Z boson production cross-section for the effect of the photons radiated from the muons produced by the Z boson decay. This corrects the measurement to the Born level in QED. However, since the simulation that we compare to data has the option to include FSR, we choose not to correct the differential distributions, and simply compare our results at the bare level to simulation with FSR included.

We determine the corrections to the overall cross-section from bare level to Born level by taking the ratio of the cross-section in simulation with FSR turned on to that with FSR turned off. To perform the correction, we simply multiply the bare level cross-section by the correction factor.

The correction factors have been measured in Herwig++ [167] for both the inclusive Z and the Z+jet processes. The model implemented in Herwig++ is based on YFS QED radiation theory and also includes fixed order QED matrix element corrections [168–170]. Using this simulation the correction factor for the inclusive Z boson production cross-section is evaluated to be 1.02309 ± 0.0008 , where the uncertainty quoted is statistical. The overall correction factors for the Z+jet process evaluated using Herwig++ are: 0.9994 ± 0.0016 ($p_T^{\text{jet}} > 10$ GeV) and 1.0096 ± 0.0027 ($p_T^{\text{jet}} > 20$ GeV).

¹¹FSR, and the differences between Born and bare level calculations, were discussed briefly in Section 2.2.

Source of Systematic Uncertainty	Relative Uncertainty (%)
Z boson selection and reconstruction efficiencies	3.5
Jet energy scale	6.3
Jet energy resolution	0.4
Unfolding	0.3
Background and fake jets	1.5
Jet identification and reconstruction efficiency	4.3
FSR corrections	0.2
Luminosity	3.5
Total	9.3

Table 6.1: The relative uncertainty from each source of possible systematic uncertainties considered for the Z+jet cross-section for $p_T(\text{jet}) > 20$ GeV. The relative uncertainties are similar for the 10 GeV threshold.

The correction factors have also been found using Pythia8 interfaced with Photos [42, 74, 171]. This uses the same theoretical ingredients as Herwig++, but simulates higher order effects differently [168]. The corresponding corrections [171] agree with those found using Herwig++ at the level of 0.15%. Other uncertainties associated with the calculation are known to be much smaller [171].

6.4 Systematic uncertainties

Statistical uncertainties are well understood, and their size can be evaluated precisely. However, we must also consider systematic uncertainties. These are, inevitably, hard to determine precisely: if a clear bias was present then it would be the scientist's role to quantify and correct for it. Instead, systematic uncertainties describe the degree to which we trust a calculated result. In this section we discuss the treatment of systematic uncertainties in the Z+jet cross-section measurement. The relative contributions to the measurement are summarised in Table 6.1. Before we start discussing the methods used to estimate these uncertainties, we note that the normalisation of the differential distributions such that their integral is unity reduces the effect of some systematic uncertainties. Uncertainties that are fully correlated between bins, like the luminosity uncertainty, have no effect on the distributions. Other uncertainties which are largely uncorrelated with the measured distribution also have their effect reduced.

6.4.1 Z boson reconstruction and selection systematic uncertainties

We take the statistical uncertainties associated with the efficiency determination and use these as the systematic uncertainties on each efficiency associated with the precision to which we determine that efficiency, as considered in Refs. [5, 155]. This is done for each Z detection efficiency ($\varepsilon_{\text{SMT}}, \varepsilon_{\text{GEC}}, \varepsilon_{\text{ID}}, \varepsilon_{\text{Track}}$) in turn.

In addition to these uncertainties, we consider other possible sources of uncertainty. As discussed in Section 5.4, we have checked for the possible effect of background contamination by re-evaluating the efficiencies using different mass windows in the tag-and-probe method. The calculations are therefore repeated using each of the re-evaluated efficiencies in turn. The change in the final result with respect to the nominal central value sets an additional systematic uncertainty for each of the efficiencies. The data-driven measurement of the tracking efficiency was also found to be accurate at the level of 0.3% per track. This is also assigned as a systematic uncertainty associated with $\varepsilon_{\text{Track}}$. This therefore gives an overall systematic uncertainty of 0.6%, as there are two muon tracks we must reconstruct and select in the $Z \rightarrow \mu^+ \mu^-$ decay. We also consider the systematic uncertainty associated with the fit used to determine ε_{GEC} . This is evaluated by replacing the efficiencies found using the gamma function extrapolation with those found using the exponential decay fit, and determining the change in the calculated cross-sections from the nominal results. The largest uncertainties associated with the Z boson reconstruction and selection efficiencies are due to determination of the tracking and global event cut requirements, which are both typically $\sim 2\%$ for the overall cross-section.

6.4.2 Jet energy scale

The jet energy scale (see Section 4.3) plays an important role in the distributions we present: it determines explicitly the number of jets we see in each event above a given transverse momentum threshold. The LHCb collaboration have provided standard uncertainties associated with the jet energy scale based on the performance in simulation [4, 123]. These uncertainties, corresponding to uncertainties in the methods used to set the scale, the flavour content of jets and the statistical uncertainty in the scale evaluation, were discussed in Section 4.3 and typically provide a $\sim 2 - 3\%$ uncertainty on the energy scale within the majority of our acceptance. The analysis is repeated with the energy scale in each bin (used to determine the scale) simultaneously¹² varied in data by one standard

¹²The variation is done simultaneously for all bins as these uncertainties are largely correlated [123, 132].

deviation. The differences in the measured cross-sections with respect to the nominal results set the systematic uncertainty associated with these sources. For the 20 GeV threshold this introduces a systematic uncertainty of $\sim 3.5\%$.

The numbers used in the jet energy correction are determined from simulation, and are therefore optimised to provide the best performance based on simulation. However, there is a further systematic uncertainty associated with the degree to which such an approach is valid: we need to determine whether the jet detector response is accurately modelled in Monte Carlo (MC) simulation. If, for example, there are differences in the alignment between data and simulation, the response of the detector to jets might be different. Discrepancies in the particle reconstruction performance in data and simulation would also create differences in the jet reconstruction.

We therefore seek to determine how well simulation and data agree. In order to do this, we must measure jets in both data and simulation. We choose to do this by considering events where the physics of the production mechanism is believed to be well understood, so that any discrepancies between data and simulation can be associated with the simulation of the detector. The size of any possible discrepancies between data and simulation then provide systematic uncertainties on the jet energy scale based on how well the detector is simulated.

Conventionally, the p_T balance of γ +jet events would be considered for such a study. However, the LHCb calorimeter saturates at high p_T , and the resolution degrades. We therefore choose to study the p_T balance of Z+jet events instead. We consider the p_T^{jet}/p_T^Z distribution. In events with little other activity, and where the Z boson and jet are produced back-to-back in ϕ , the Z boson and the jet should balance in p_T . In effect, we can use the Z boson p_T as a proxy for the true jet p_T . In these events the p_T^{jet}/p_T^Z distribution is expected to be well modelled at generator level by simulation [150], so that any discrepancy between data and simulation can be assigned to the detector response. By comparing this distribution between data and simulation, we can gauge the extent to which the detector response to jets agrees between data and simulation.

Before proceeding with this investigation, we note that the correlations between the p_T^{jet}/p_T^Z distribution in the selected back-to-back events and the distributions we seek to measure in the analysis are small, and so assigning a systematic uncertainty based on the agreement between data and simulation does not render the comparison of the measured distributions with theoretical predictions uninteresting. We also note that the selection of back-to-back events does not bias our comparison to events where the jet

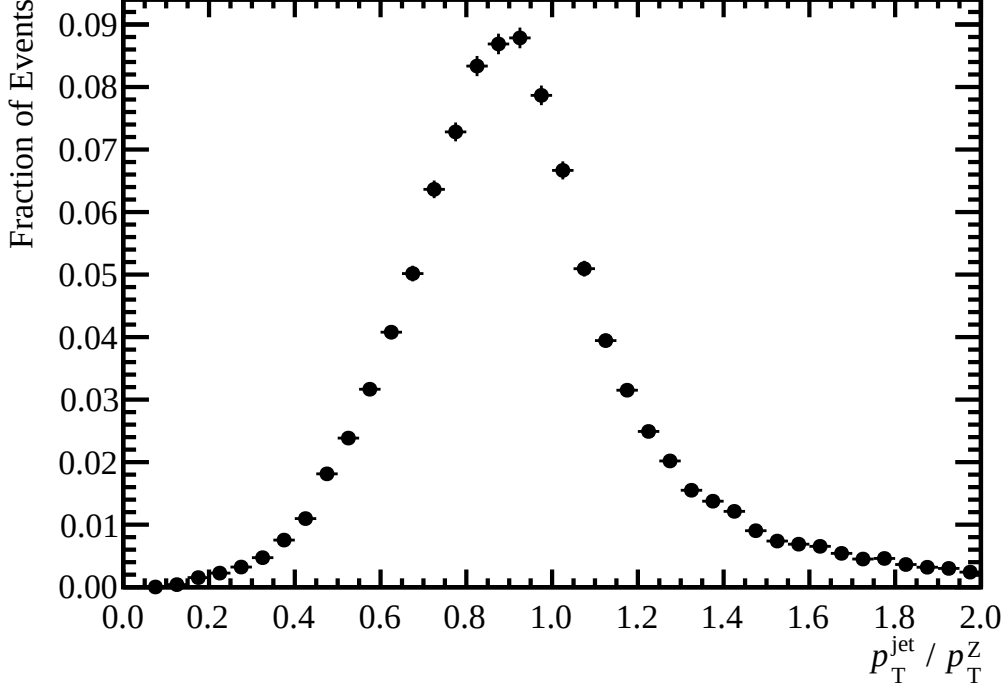


Figure 6.7: The p_T balance of selected Z+jet events at the truth level, for $p_T^{\text{jet}} > 10$ GeV.

reconstruction performs best: the jet energy scale and resolution show no dependence on the difference in azimuthal angle between the Z and the jet.

In order to select back-to-back events with little other activity, we place additional selection requirements to the nominal selection of Z+jet candidate events discussed in the previous chapter. We require that $|\Delta\phi(\text{Z}, \text{Hardest jet})| > 7\pi/8$, and $\frac{p_T(\text{Second jet})}{p_T(\text{Hardest jet})} < 0.25$. We place no jet identification requirements on the second (next highest p_T) ‘jet’ in the event. This means we are selecting events with little other activity, rather than just events where there is no second jet. In these events, the Z boson p_T is expected to largely balance the jet p_T at the truth level. This is shown in simulated events in Fig. 6.7.

The p_T^{jet}/p_T^Z distributions in data and simulation are overlaid and plotted in Figure 6.8. Good agreement is seen between data and simulation, and similar agreement is seen when the distribution is considered as a function of the jet p_T and η . This agreement provides confidence that the jet reconstruction is well modelled in simulation, so we proceed to quantify the level to which we can trust simulation, and measure a systematic uncertainty associated with the simulation of the detector response to jets.

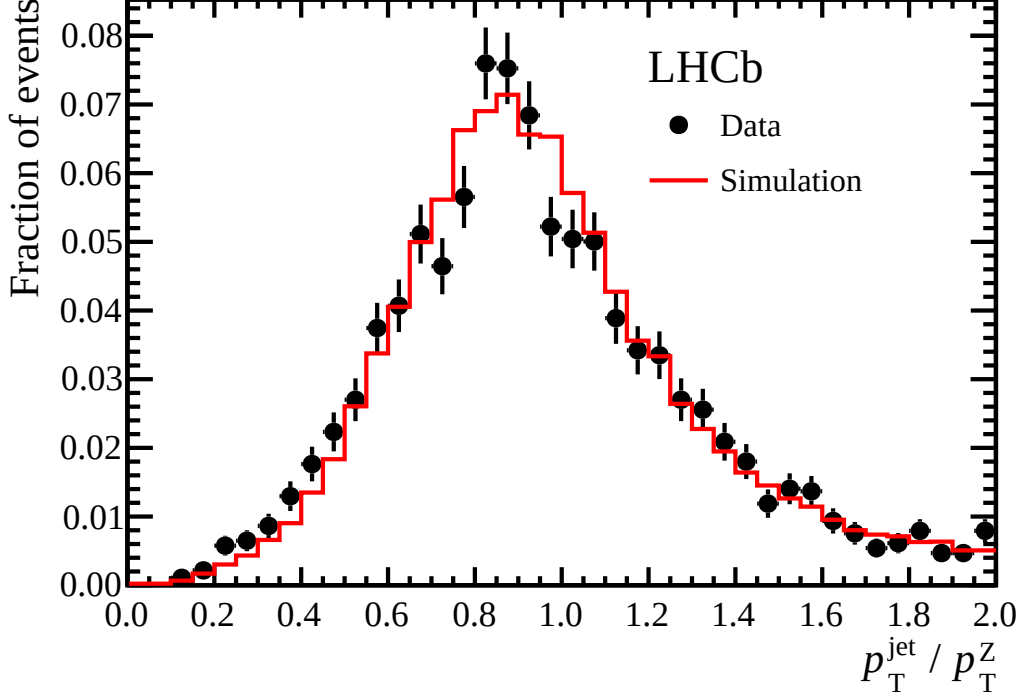


Figure 6.8: The p_T balance of selected Z+jet events for $p_T^{\text{jet}} > 10$ GeV. The data is shown as the black, with simulation in red.

We consider the effect of varying the detector response in simulation. We vary the reconstructed jet p_T from its original value, $p_{T,\text{orig}}^{\text{jet}}$, to a new value, $p_{T,\text{new}}^{\text{jet}}$, through $p_{T,\text{new}}^{\text{jet}} = \beta \times p_{T,\text{orig}}^{\text{jet}}$. By comparing the p_T^{jet}/p_T^Z distribution between data and the varied simulation we can fit for the parameter β which gives the best agreement between data and simulation. If $\beta = 1$, then the simulation has not been varied; any offset of the ‘best’ value from unity is used to set the systematic uncertainty.

We quantify the agreement between data and simulation by performing a χ^2 comparison between data and simulation as a function of β . The results are shown in Fig. 6.9: a clear minimum is seen, occurring at $\beta_{\text{best}} = 1.02$, with an uncertainty of about 0.02. The best value of β is therefore consistent with unity (for this value of β the χ^2 test suggests that the distributions in data and simulation are consistent at the level of $\sim 10\%$). We therefore consider a global uncertainty in the jet energy scale of 3%, in order to account for the offset and the potential uncertainty in the offset. This uncertainty is conservative, as the systematic uncertainty may be smaller than the numbers we assign. However, without evidence to this fact, we choose to assign an uncertainty that corresponds to the level at which we can be sure that the simulation accurately models data.

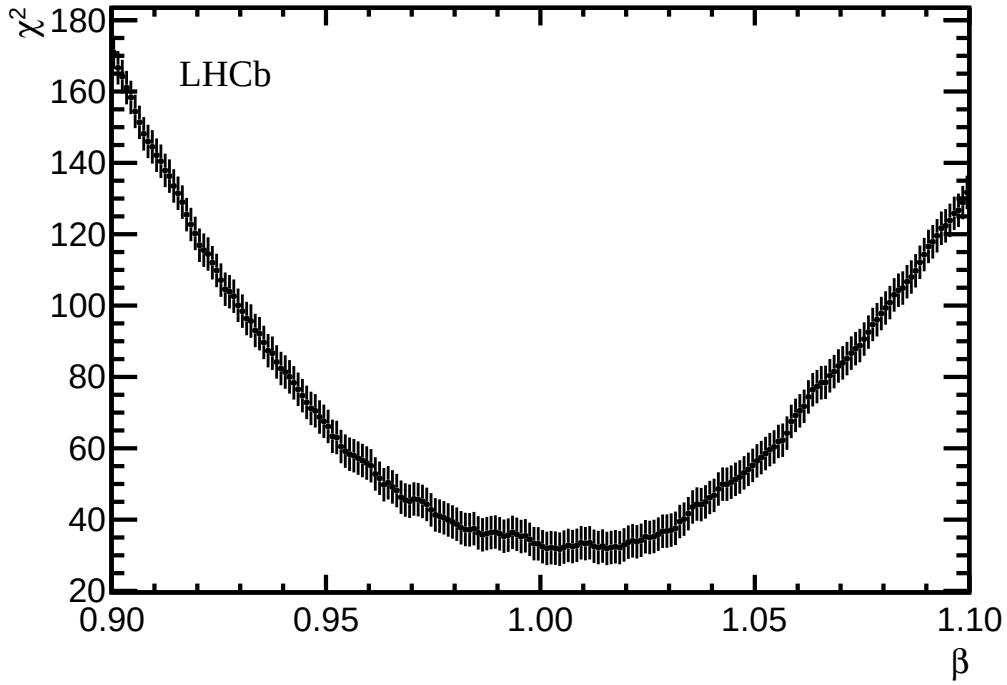


Figure 6.9: The value of χ^2 from a comparison of the p_T^{jet}/p_T^Z distribution in data and simulation as a function of β , for $p_T^{\text{jet}} > 10$ GeV. The minimum χ^2 is 32 (based on 23 degrees of freedom), and occurs at $\beta_{\text{best}} \sim 1.02$. For this value of β data and simulation are consistent at the level of 10%.

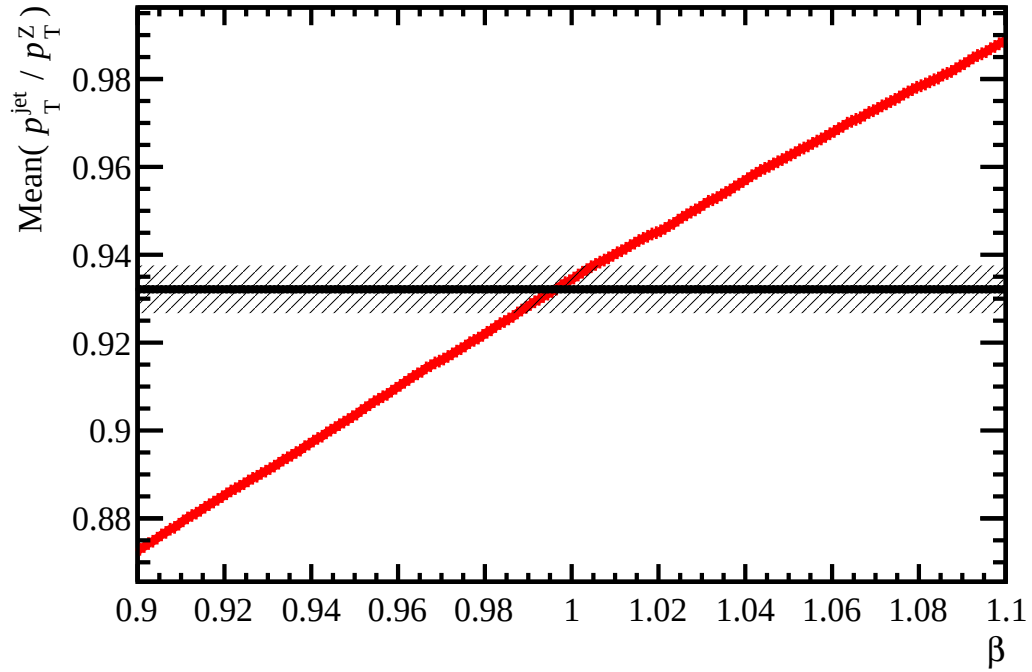


Figure 6.10: The comparison of the mean of the p_T^{jet}/p_T^Z distribution in data and simulation for $p_T^{\text{jet}} > 10$ GeV. The mean in data is shown in black (with the uncertainty given by the shaded region), with the mean in simulation shown in red as a function of β . The best agreement between data and simulation occurs for $\beta_{\text{best}} = 0.996 \pm 0.011$.

Bin	β_{best}
$10 < p_{\text{T}} < 20$ GeV	0.98 ± 0.02
$20 < p_{\text{T}} < 30$ GeV	1.02 ± 0.02
$30 < p_{\text{T}} < 70$ GeV	0.98 ± 0.02
$p_{\text{T}} > 70$ GeV	0.98 ± 0.04
$2.0 < \eta < 2.5$	0.99 ± 0.02
$2.5 < \eta < 3.5$	1.00 ± 0.02
$3.5 < \eta < 4.5$	1.00 ± 0.03

Table 6.2: The values of β where we see best agreement between data and simulation, for the different bins considered, based on comparing the means of the $p_{\text{T}}^{\text{jet}}/p_{\text{T}}^{\text{Z}}$ distribution between data and simulation. A two dimensional binning does not provide sufficient precision to probe the performance of simulation.

We also desire to test the performance of simulation as a function of the jet p_{T} and η . We investigate this by finding the difference in the means of the $p_{\text{T}}^{\text{jet}}/p_{\text{T}}^{\text{Z}}$ distribution in data and simulation. This mean is truncated by considering events that lie in the range $0.4 < p_{\text{T}}^{\text{jet}}/p_{\text{T}}^{\text{Z}} < 1.6$, since outlying events in this distribution are particularly sensitive to multi-jet events which are not expected to be well modelled in simulation [150] and could make the distribution less sensitive to the jet detector response. This test is used to determine the agreement between data and simulation in the binned samples as it is more straightforward and robust at probing potential discrepancies than the χ^2 comparison when the sample sizes decrease. The test is consistent with the χ^2 comparison within the assigned global uncertainty when the overall sample is considered (see Figure 6.10).

The bins in p_{T} are chosen to maximise sensitivity to the 10 and 20 GeV thresholds, but to also confirm the behaviour at higher p_{T} , whilst the bins in η are chosen to check the performance close to the edge of the LHCb detector (at low η), and in the regions where track finding is dominated by the outer and inner trackers ($2.5 < \eta < 3.5$ and $\eta > 3.5$ respectively). The values of β which give the best agreement between the means of the $p_{\text{T}}^{\text{jet}}/p_{\text{T}}^{\text{Z}}$ distribution in data and simulation are consistent with unity, varying between 0.98 and 1.02 as a function of the jet η and p_{T} , as shown in Table 6.2. To account for potential differences in the detector response to jets between data and simulation we consider the effects of a 3% systematic uncertainty in each bin of the jet pseudorapidity and p_{T} considered. This covers the measured offsets from unity. This uncertainty is also statistically limited, and were more data available the uncertainty would be expected to decrease.

We note that taking a global uncertainty of 3% and an uncertainty of 3% in bins of η or p_T at the same time might lead to a possible double counting of the uncertainty from potential discrepancies between data and simulation. We therefore investigate this. The effect of the global uncertainty on the normalised differential distributions is negligible, as globally varying the energy scale tends to change the number of events but not the relative number in each bin. Likewise the binned uncertainties have a small effect on the total cross-section when their effects are combined in quadrature. We are therefore comfortable proceeding with this approach.

Repeating the analysis with the energy scale varied in data yields a difference in the overall cross-section at the level of 5.3% for the 20 GeV p_T threshold with respect to the nominal cross-section. This sets the systematic uncertainty associated with the degree that the detector response to jets in simulation and data agree. This uncertainty on the energy scale has since become the standard uncertainty from this source considered by the LHCb collaboration [127, 172].

Combining this uncertainty with the other uncertainties from the jet energy scale discussed earlier sets a total uncertainty from the jet energy scale on the overall Z+jet cross-section of 6.3% above the 20 GeV threshold. This provides the dominant uncertainty. The jet energy scale uncertainties are a similar size for most of the differential distributions, but the uncertainties are smaller for the Z boson rapidity, as this variable has less direct dependence on the jet kinematics.

6.4.3 Jet energy resolution

We also assign a systematic uncertainty based on how well the jet energy resolution agrees between data and simulation. The same selection of events described in Section 6.4.2 is then used to select back-to-back events where the jet p_T / Z p_T distribution is strongly influenced by the detector response to jets. The jet energy in simulation is then smeared, and the data and smeared simulation are compared to determine the level of smearing where data and simulation agree best. The smearing is performed by taking a random number from a Gaussian distribution of unit mean and width α , for each jet in simulation, and multiplying the jet p_T by this random number to get a fluctuated p_T for each jet. We compare the width (standard deviation) of the distribution in the smeared simulation and data. The width of the distribution in data and simulation already agree at the $\sim 0.6\sigma$ level with no smearing ($\alpha = 0$). We therefore take as the systematic uncertainty the level of smearing where we see disagreement between data and simulation.

This occurs for $\alpha \sim 0.1$. This is a conservative approach, but the systematic uncertainty from the jet energy resolution is not dominant. There is no evidence for any variation of this value with the jet p_T and η . The analysis is then repeated a large number of times with the simulated jet transverse momentum smeared by this additional factor, with the absolute mean change in each cross-section setting the relevant uncertainty. This uncertainty tends to be small as the net effect of the smearing is small (the jet p_T is as likely to fluctuate upwards as downwards).

6.4.4 Unfolding

The difference in the number of unfolded events above the 10 and 20 GeV thresholds between the D'Agostini and SVD methods is taken as the systematic uncertainty associated with the unfolding method. For the 20 GeV threshold this gave a 0.2% uncertainty.

The unfolding was validated on an independent simulation sample (as discussed in Section 6.2.1). The level to which the unfolded distribution in the validation sample agrees with the simulated truth distribution in the same sample is set as an additional systematic uncertainty associated with the unfolding procedure. This systematic uncertainty is conservative since the differences between the two simulation samples studied are larger than the differences between the Z+jet simulation used in the measurement and the recorded data. This provides a 0.1% uncertainty for the 20 GeV threshold¹³. For the differential cross-sections, the uncertainties are larger, but are typically much smaller than the statistical uncertainty in the relevant bin. The exception to this is at small $|\Delta\phi|$ and small p_T^Z , where this systematic uncertainty is similar in size to the statistical uncertainty. Even in these regions it is not the dominant uncertainty.

6.4.5 Background and Fake Jets

The purity was estimated to be $\sim 0.997 \pm 0.001$ (discussed in Section 5.3), where the uncertainty quoted is statistical. Other possible systematic uncertainties on the background level were discussed in Section 5.3 and are much smaller than the statistical uncertainty. The statistical uncertainty on the number of background events is therefore treated as a systematic uncertainty in the cross-section analysis. This matches the procedure used in Ref. [5], and provides a negligible systematic uncertainty.

¹³This can also be viewed as a closure test; the method provides the truth distribution at the level of 0.1%.

Another potential source of ‘background’ comes from the presence of fake jets. The jet identification requirements we use to select well reconstructed jets and remove fake jets were introduced in Section 4.2. Fake jets are typically caused by a single particle mis-measured as having high transverse momentum. The jet identification requirements, introduced in Section 4.2, make use of this to select well reconstructed jets, reducing the fake rate to $\sim 4\%$ of jets above the 10 GeV threshold, $\sim 1\%$ of jets above the 20 GeV threshold, and $\sim 0.7\%$ of jets above 50 GeV. These are then removed by the unfolding. Data and simulation agree well in the distributions which are most sensitive to these remaining fake jets, as shown in Figure 6.11, so the presence of fake jets appears to be well-modelled in simulation.

By changing the requirement on mtf to < 0.4 or n_{pointing} to > 4 we reject the majority of the remaining fake jets (see Figure 6.11). If we do this, the fraction of events rejected in data and simulation agrees to the level of $2\%(mtf)$ or $1\%(n_{\text{pointing}})$. This difference could be due to the fake jets being simulated incorrectly; we therefore assign a 1.5% systematic uncertainty for jets with $p_T < 50$ GeV. Above 50 GeV we assign a 0.7% uncertainty for this contribution (the size of the effect). This is done as there are not enough events to carry out the same analysis described above to a high enough statistical precision. This is clearly a conservative estimate, but the size of the effect is small and the uncertainty is not dominant.

6.4.6 Jet reconstruction and selection efficiency

In addition to the uncertainty associated with the jet identification cuts and fake jets, we seek to determine the systematic uncertainty associated with the value of ε_{jet} we measure from simulation. We consider contributions from a variety of effects.

The effect of reweighting the simulation sample to match the jet multiplicity distribution in data is to change the efficiency by $\sim 0.5\%$. This is assigned as a systematic uncertainty. Reweighting the simulation sample based on other distributions yields negligible changes to ε_{jet} .

In addition, we note that ε_{jet} varies rapidly at low p_T . If the p_T distribution is incorrectly simulated, then the values of ε_{jet} in each bin could be biased. We therefore repeat the analysis with the bin sizes halved. This change in binning makes the measurement of ε_{jet} less sensitive to the p_T distribution in each bin, and so provides an estimate of the size of any possible effect. The change in the final Z+jet production cross-section

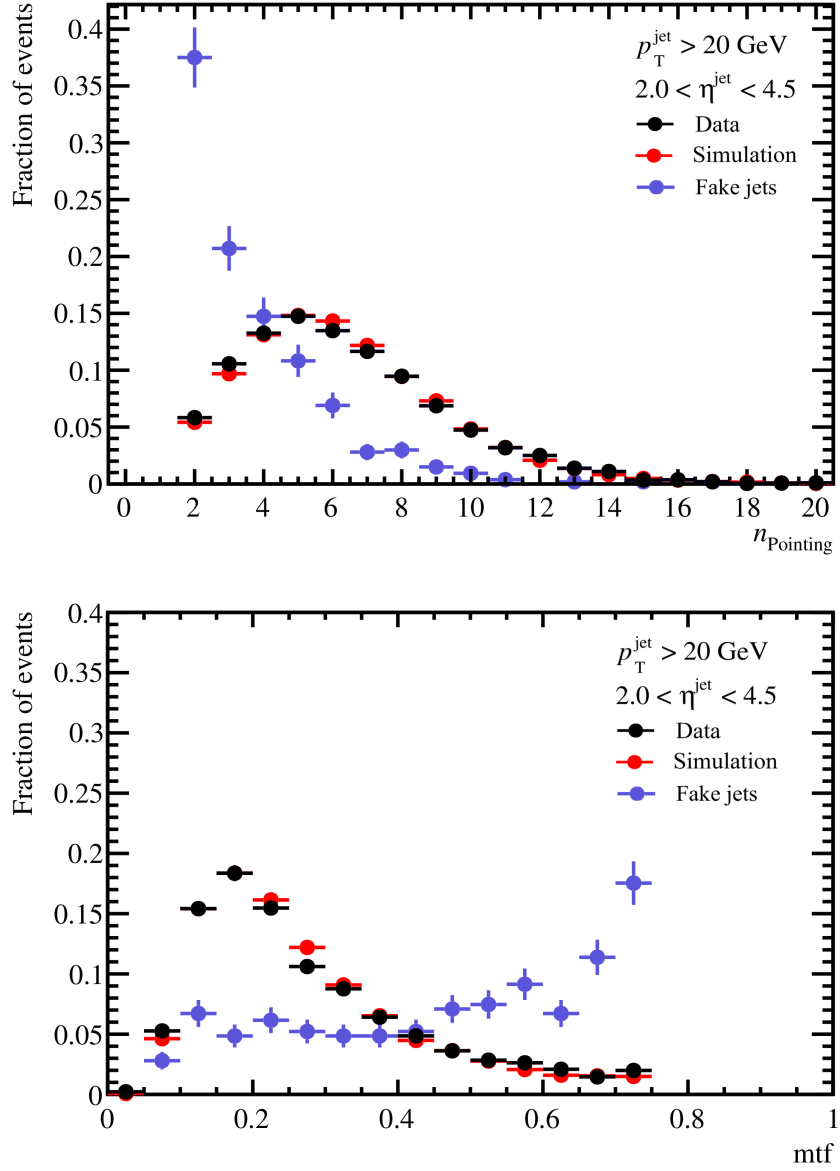


Figure 6.11: The n_{Pointing} (top) and mtf (bottom) distributions for fake jets (purple) in simulation, and for all jets in data (black) and simulation (red) that pass the jet identification requirements.

with the modified binning is small (0.2%), and is assigned as an additional systematic uncertainty. The effect is of a similar size for the differential distributions.

We also calculate ε_{jet} using the inclusive $Z \rightarrow \mu^+ \mu^-$ simulation sample. The measured efficiencies agree at the level of 0.1% for $p_{\text{T}}^{\text{jet}} > 20$ GeV. We assign this as an additional systematic uncertainty on the method. We also carry this out for the differential distributions, and assign the difference in each bin as a systematic uncertainty. This uncertainty is typically smaller than the statistical uncertainty in each bin.

We also assign a systematic uncertainty associated with whether simulation accurately models the fraction of true jets where we fail to reconstruct a jet, or events where a reconstructed jet is rejected by the jet selection requirements. We consider these separate contributions in turn.

The fraction of events containing a reconstructed jet (neglecting the jet identification requirements), given that there is a true jet in the event, is about 98.5% (for $p_{\text{T}}^{\text{jet}} > 10$ GeV). This efficiency is seen to increase to about 99% for $p_{\text{T}}^{\text{jet}} > 20$ GeV. We have considered different approaches to determine how well this quantity is modelled in simulation, but we do not achieve a statistical sensitivity better than the size of the effect. We therefore assign a 1.5% systematic uncertainty to the jet reconstruction efficiency. This is a conservative estimate - we are stating that the simulation could be incorrect by a factor of two.

The dominant contribution to ε_{jet} comes from the jet identification requirements. We seek to determine how well the effect of these requirements is modelled in simulation. The approach we take to determine the accuracy of the simulation of the n_{Pointing} , mtf and mpt distributions will have some overlap with the discussion above concerning fake jets, as these same distributions enable us to estimate how well fake jets are simulated.

Specifically, we want to consider how well the regions we remove with the jet identification requirements are modelled in simulation. We choose not to simply compare data and simulation in these regions: the regions were removed for a reason; the fraction of fake jets in these regions is known to be large and would ‘pollute’ any comparison of data and simulation. Instead, we consider how well data and simulation agree close to the ‘cut’ value. If data and simulation agree at the level of $x\%$ in this region, and the fraction of true Z+jet events rejected by the jet identification requirements is y , then we set the systematic uncertainty on the overall efficiency from the agreement of data and

Bin	$x(\text{mtf})$ (%)	$x(\text{mpt})$ (%)	$x(n_{\text{Pointing}})$ (%)
$10 < p_{\text{T}}(\text{jet}) < 20$ GeV	7 ± 8	9 ± 3	4 ± 2
$20 < p_{\text{T}}(\text{jet}) < 30$ GeV	0 ± 17	5 ± 8	2 ± 4
$p_{\text{T}}(\text{jet}) > 30$ GeV	30 ± 10	0 ± 15	15 ± 4

Table 6.3: The agreement between data and simulation in each of the distributions used to select jets, close to the ‘cut’ value for the jet identification requirements.

simulation to be $y \times x\%$.¹⁴ Since the efficiency varies strongly with p_{T} we perform this analysis as a function of the jet p_{T} .

We have already seen that the jet identification requirements are placed on distributions that shown good agreement between data and simulation. For the mtf distribution, we consider the agreement between data and simulation in the region $0.6 < \text{mtf} < 0.75$. For mpt, we consider the region $1.8 < \text{mpt} < 3.6$ GeV. For n_{Pointing} , we consider the region $2 \leq n_{\text{Pointing}} \leq 4$. We consider the relative numbers of events in data and simulation in this region near the cut value, and express how well these numbers agree as a percentage, x , given by $x = 100\% \times |f_{\text{sim}} - f_{\text{data}}|/f_{\text{sim}}$, where f_{sample} is the relative number of events in the region close to the cut value in the relevant sample.

The values of x as a function of jet p_{T} for the three different distributions we cut on are shown in Table 6.3. Whilst some of the numbers are significantly different from zero, the differences between data and simulation tend to be small. The largest values of x are for the mpt distribution at low jet p_{T} ($\sim 10\%$) and for the mtf distribution at high jet p_{T} ($\sim 30\%$). We therefore assume we determine the relative number of events lost through the jet ID requirements to an accuracy of $\sim 10\%$ at low p_{T} and $\sim 30\%$ at higher p_{T} . This leads to a conservative estimate of the systematic uncertainty as we implicitly assume that all the events that fail the ID requirements are failing the particular requirement where data and simulation agree the least well. However, the effect of this assumption is small. The agreement between data and simulation is similar at low p_{T} for all the variables, and at high p_{T} the majority of events which fail to pass the requirements are rejected by the requirement on mtf.

The values of the jet efficiency vary between $\sim 75\%$ at low jet p_{T} and $\sim 95\%$ at higher jet p_{T} . A $\sim 10\%$ uncertainty on the 25% of events rejected at low p_{T} yields a $\sim 2.5\%$

¹⁴We give here a numerical example: if data and simulation agree at the level of 10% close to the requirement, and the requirement rejects 15% of events, then we set the systematic uncertainty associated with the efficiency to be 10% on that 15% of events, i.e. the efficiency is $85.0 \pm 1.5\%$.

uncertainty on the overall efficiency. Likewise, at higher p_T the $\sim 30\%$ uncertainty on the 5% of events rejected yields a $\sim 1.5\%$ uncertainty on the overall efficiency. Similar agreement (and uncertainties) are seen when the distributions are considered as a function of the jet pseudorapidity. For simplicity, we estimate our uncertainty on the jet efficiency to be 2.5%.

For the overall cross-section we treat this 2.5% as being correlated between the p_T bins: this simply translates into a 2.5% systematic uncertainty on the overall cross-section measurement. For the differential distributions we take a conservative approach and consider this uncertainty as uncorrelated between the different bins, providing coverage of the ‘worst’ possible situation and increasing confidence in our final results. However, we note that the impact of this choice is small, as the uncertainty considered here is not dominant.

6.4.7 Final State Radiation

We correct the total Z+jet cross-section from the bare level to the Born level by multiplying by a correction factor. This correction factor is taken from simulation. Correction factors for the inclusive Z process produced by different theoretical programs agree at the level of 0.15% (discussed in Section 6.3.1). The difference between the correction factors predicted by different programs is a good indicator of the theoretical uncertainty in such predictions[168]. We therefore consider this 0.15% to be the systematic uncertainty in the FSR correction. A similar approach has been taken in other LHCb measurements [155].

6.4.8 Luminosity

We measure the luminosity to an accuracy of 3.5%, as detailed in Ref. [97], and discussed in Section 3.2.10. This therefore contributes a 3.5% systematic uncertainty on any cross-section measurement.

6.5 Cross-checks

6.5.1 Measuring the Z production cross-section

We can take the efficiencies we have outlined in the previous chapter, and, using the corrections for FSR we outline in Section 6.3.1 and the relevant systematic uncertainties outlined in Section 6.4, we can calculate the inclusive Z cross-section using events recorded in 2011 where the Z boson decays to a dimuon final state. This cross-section is

$$\sigma_Z = 76.5 \pm 0.3(\text{stat.}) \pm 1.2(\text{syst.}) \pm 2.7(\text{lumi.}) \text{ pb},$$

where the first uncertainty is statistical, the second is systematic and the third is the uncertainty due to the luminosity determination. This result agrees very well with the values published by LHCb for Z bosons decaying to a dimuon final state,

$$\sigma_Z = 76.7 \pm 1.7(\text{stat.}) \pm 3.3(\text{syst.}) \pm 2.7(\text{lumi.}) \text{ pb},$$

based on data recorded in 2010 [5], and for Z bosons decaying to a dielectron final state in data recorded in 2011 [6],

$$\sigma_Z = 76.0 \pm 0.8(\text{stat.}) \pm 2.0(\text{syst.}) \pm 2.6(\text{lumi.}) \text{ pb}.$$

These results are compared to each other and a theoretical prediction in Figure 6.12: clearly good agreement is seen, increasing confidence in the presented Z boson reconstruction. The main difference between the results based on dimuon decays of the Z boson measured in 2010 and those measured in 2011 is the improvement in the systematic uncertainty quoted for the 2011 result. The systematic uncertainties in Ref. [5] were statistically limited: the tag-and-probe samples used to determine the efficiencies were small in comparison to the samples now available.

6.5.2 Magnet polarity and charge asymmetries

The efficiencies presented here were checked for any bias between the two muon charges, and for differences between different magnet polarities. No such effects were seen, providing confidence in the results quoted here. In addition the reconstructed distributions were compared for the different magnet polarities. No significant variations were seen with magnet polarity.

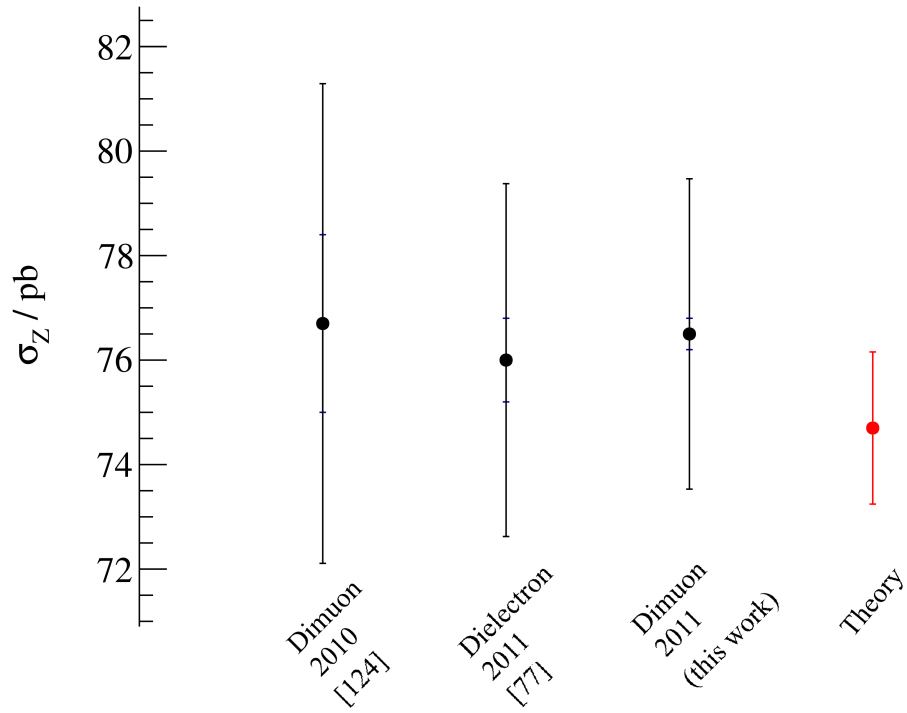


Figure 6.12: The inclusive Z production cross-section is shown for events where the Z boson decays directly to a dimuon final state (based on work in this thesis and Ref. [5]) and for events where the Z boson decays directly to a dielectron final state [6] in black. For the measurements the inner error bars show the statistical uncertainties, whilst the outer error bars show the total uncertainty. These results are compared to a theoretical prediction from FEWZ [67] evaluated at $\mathcal{O}(\alpha_s^2)$ using the MSTW08 NNLO PDF parametrisation [57], taken from Ref. [5], shown in red.

6.5.3 Closure tests

An uncertainty was applied to account for the differences between the unfolded and truth distributions when an independent simulation sample was studied. An uncertainty was also assigned to account for the (small) changes when ε_{jet} was re-determined using this independent sample. In both cases the assigned uncertainty was about 0.1%. These two stages together form a closure test; when the ‘jet’ analysis is applied to the independent simulation sample closure is seen at the level of about 0.2%.

In addition, the (un-normalised) differential cross-sections can be calculated. If the overall (un-binned) cross-section failed to account for an important variation of an efficiency with one of the variables the differential cross-sections consider, then the sum of all the bins in the relevant differential cross-section would not be equal to the overall cross-section. This is not the case. We see excellent agreement in the total cross-section, regardless of whether it is measured as a function of a differential variable or calculated globally¹⁵.

6.6 Results

The Z+jet cross-section and the cross-section ratio $\sigma(\text{Z+jet})/\sigma(\text{Z})$ are measured at the Born level (in QED). For the $p_{\text{T}}^{\text{jet}} > 20$ GeV threshold the results are

$$\sigma(\text{Z+jet}) = 6.3 \pm 0.1 \text{ (stat.)} \pm 0.5 \text{ (syst.)} \pm 0.2 \text{ (lumi.) pb,}$$

$$\frac{\sigma(\text{Z+jet})}{\sigma(\text{Z})} = 0.083 \pm 0.001 \text{ (stat.)} \pm 0.007 \text{ (syst.)},$$

and for the $p_{\text{T}}^{\text{jet}} > 10$ GeV threshold,

$$\sigma(\text{Z+jet}) = 16.0 \pm 0.2 \text{ (stat.)} \pm 1.2 \text{ (syst.)} \pm 0.6 \text{ (lumi.) pb,}$$

$$\frac{\sigma(\text{Z+jet})}{\sigma(\text{Z})} = 0.209 \pm 0.002 \text{ (stat.)} \pm 0.015 \text{ (syst.)},$$

where the first uncertainty is statistical, the second is systematic and the third is the uncertainty due to the luminosity determination. For each of the differential distributions

¹⁵This is of course a necessity; if this was not observed, we would have to investigate the problem, and potentially bin the analysis in another variable.

the cross-section and uncertainty in each bin are shown graphically below, and are tabulated in the appendix.

The measured cross-sections are compared to theoretical predictions for the Z+jet process evaluated at $\mathcal{O}(\alpha_s^2)$ using POWHEG [7–11] and matched to PYTHIA 6.4 [12] (using the Perugia 0 tune [173]) for shower development and hadronisation. Jets are created out of all stable particles in the final state that are not produced by the decay of the Z boson. These predictions are computed with the renormalisation scale and factorisation scales set to the nominal value of the vector boson transverse momentum. These are the default settings within POWHEG [11] for predictions of Z+jet production at $\mathcal{O}(\alpha_s^2)$. These theoretical predictions are computed for three different NLO PDF parametrisations: MSTW08 [57], CTEQ10 [64] and NNPDF 2.3 [65].

For the ratio $\sigma(Z + \text{jet})/\sigma(Z)$, the Z+jet and inclusive Z production cross-sections are evaluated separately. The Z+jet production cross-section is computed at $\mathcal{O}(\alpha_s^2)$ as above. The inclusive Z boson production cross-section is computed at $\mathcal{O}(\alpha_s)$, also using POWHEG and PYTHIA [7], with the nominal renormalisation and factorisation scales set to the POWHEG default for inclusive Z boson production (they are both set to the Z boson mass) [7]. These cross-sections, and the ratio, are evaluated separately for the MSTW08, CTEQ10 and NNPDF 2.3 PDF parametrisations.

We also wish to consider the effect of different order calculations. The $\mathcal{O}(\alpha_s)$ predictions of inclusive Z boson production include jet production at leading order. Therefore, to see the effect of higher orders in pQCD on the Z+jet cross-section, theoretical predictions for the ratio are also computed by taking the ratio between the Z+jet and inclusive Z cross-sections where both are evaluated¹⁶ at $\mathcal{O}(\alpha_s)$. This is calculated using the MSTW08 PDF parametrisation. Theoretical predictions of Z+jet production at $\mathcal{O}(\alpha_s)$ using the MSTW08 PDF parametrisation are also computed for the differential distributions. These predictions are all evaluated using the same settings described above for the calculation of the inclusive Z boson production cross-section.

In addition, we calculate a parton level prediction for the ratio $\sigma(Z + \text{jet})/\sigma(Z)$ using FEWZ [67]. In order to find this ratio the Z+jet cross-section is computed at $\mathcal{O}(\alpha_s^2)$ and the inclusive Z boson production cross-section¹⁷ is computed at $\mathcal{O}(\alpha_s)$. The predictions of

¹⁶These predictions correspond to the same inclusive simulation, where, in order to find the Z+jet cross-section, as opposed to the inclusive Z boson cross-section, we require a jet in the final state in the fiducial acceptance.

¹⁷We could have chosen to calculate the inclusive Z boson production cross-section at $\mathcal{O}(\alpha_s^2)$ using FEWZ. However, we choose to determine the cross-section at $\mathcal{O}(\alpha_s)$ to enable easier comparison with other predictions.

the numerator and the denominator both use the same MSTW08 PDF parametrisation, and the renormalisation scale and factorisation scales are set to the nominal value of the vector boson mass. These settings correspond to the default for FEWZ [67]. The parton level prediction of the ratio neglects effects from hadronisation and the underlying event, and so comparisons with the results and the other predictions are indicative of the size of these effects.

Uncertainties on all predictions are calculated by repeating the calculations with the renormalisation and factorisation scales simultaneously varied by a factor of two about their nominal values. The spread in predictions from the different PDF parametrisations is indicative of the PDF uncertainty. The matching of the PDF orders and the orders of the relevant calculation was made following Refs. [174, 175].

The cross-section ratios are compared in Fig. 6.13 to the Standard Model theoretical predictions discussed above. The results for the differential cross-sections, uncorrected for final state radiation from the muons, are presented in Figs. 6.14–6.19. For all cases reasonable agreement is seen between the Standard Model calculations and the data. The various trends in the distributions are similar to those discussed in detail in Section 5.2.3, and so we will not repeat all the points made there in detail.

The agreement of the measured cross-section ratio with the theoretical predictions is better for the 20 GeV threshold than the 10 GeV threshold. This is not surprising: pQCD is more valid for large values of the jet transverse momentum, and is typically considered reliable above about 10–15 GeV [136]. The smaller uncertainties associated with the $\mathcal{O}(\alpha_s)$ predictions than the $\mathcal{O}(\alpha_s^2)$ predictions are due to the scales used being larger in the $\mathcal{O}(\alpha_s)$ predictions (scale variations tend to decrease with increasing scale), and because a larger number of the jets are produced in the parton shower. The scales used in the parton shower stage of the simulation are not varied, so the $\mathcal{O}(\alpha_s)$ predictions, which rely more on the parton shower, appear to have a lower uncertainty.

We start our discussion of the differential distributions by noting that as the jet p_T threshold increases, both the jet pseudorapidity distribution (see Figure 6.15) and the Z boson rapidity distribution (see Figure 6.16) shift to favour more central production. Consequently the value of Δy (see Figure 6.19) becomes more strongly peaked at zero. This behaviour is due to the kinematics of such events since more energetic parton collisions tend to lead to jet production in the central region as it is easier to generate p_T in more central collisions (as was shown in Figure 4.1). There are two other important effects as the jet p_T threshold increases. Because of the correlations between the jet p_T

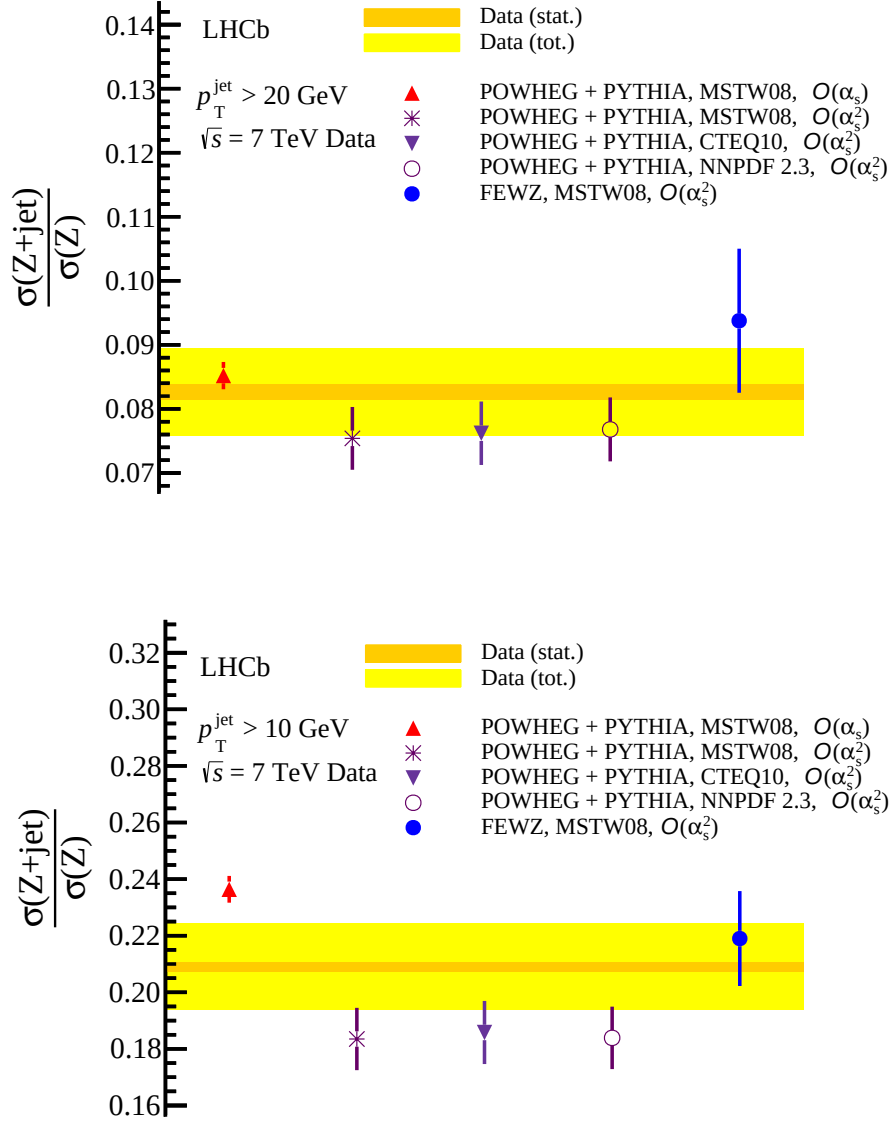


Figure 6.13: Ratio of the Z+jet cross-section to the inclusive Z cross-section, for (top) $p_T^{jet} > 20$ GeV and (bottom) $p_T^{jet} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). The points correspond to different theoretical predictions with the error bars indicating their uncertainties as described in the main text. These results are corrected for FSR from the final state muons from the Z boson decay.

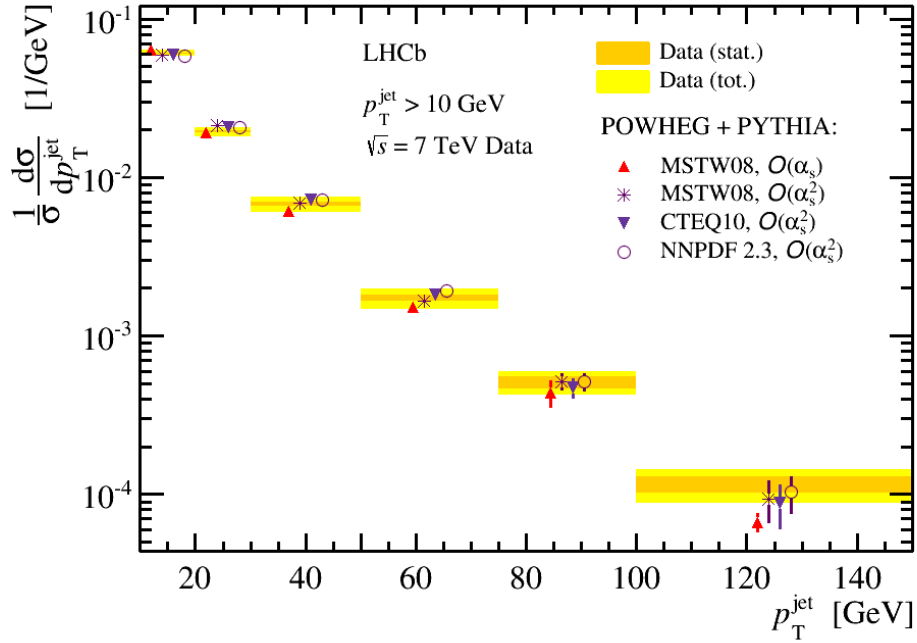


Figure 6.14: Cross-section for Z+jet production, differential in the leading jet p_T , for $p_T^{\text{jet}} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). The points correspond to different theoretical predictions with the error bars indicating their uncertainties as described in the main text. Predictions are displaced horizontally for presentation. These results are not corrected for FSR from the final state muons from the Z boson decay.

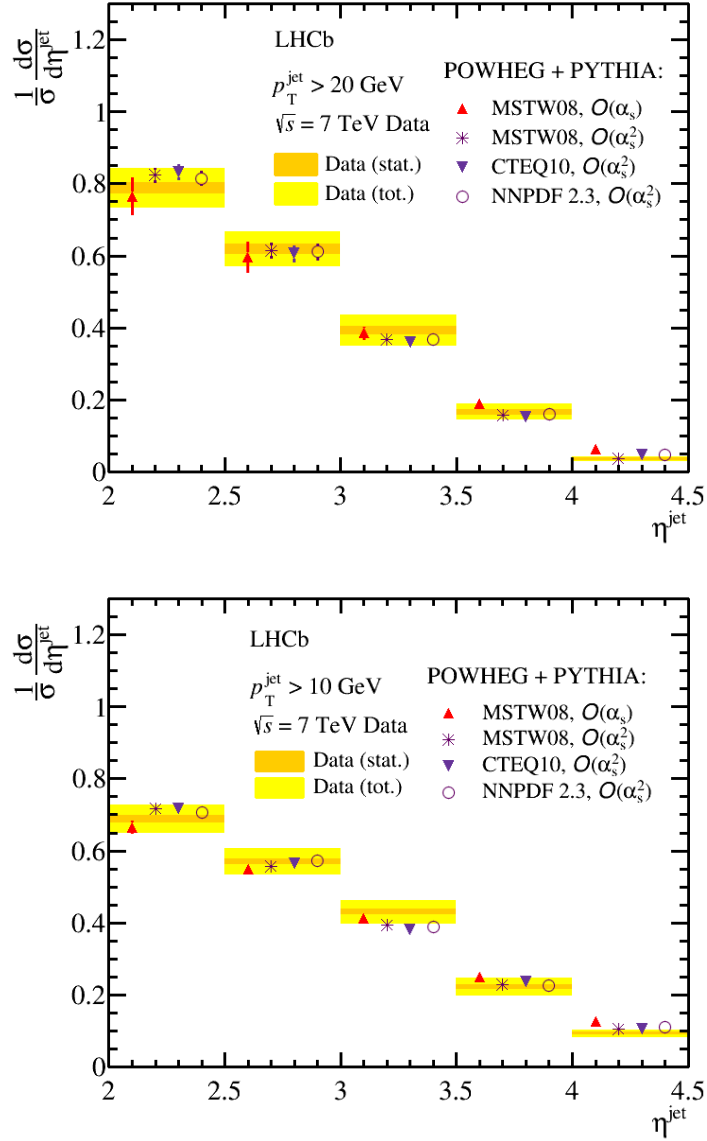


Figure 6.15: Cross-section for Z+jet production, differential in the leading jet pseudorapidity, for (top) $p_T^{\text{jet}} > 20$ GeV and (bottom) $p_T^{\text{jet}} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). Superimposed are predictions as described in Fig. 6.14.

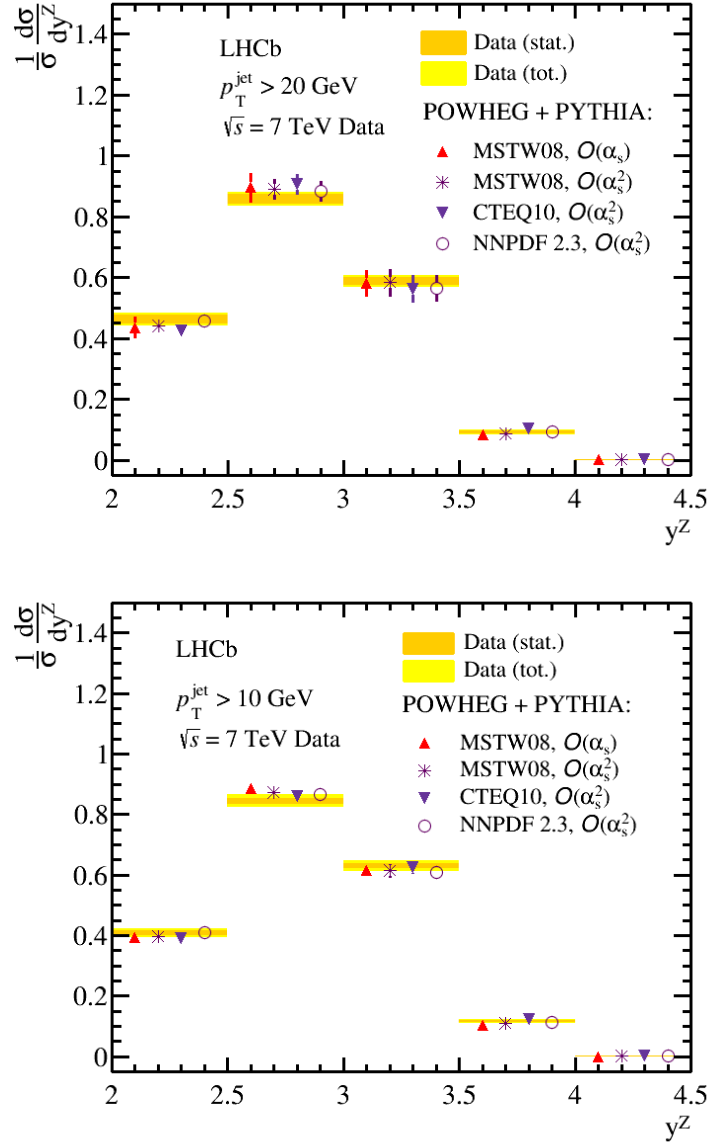


Figure 6.16: Cross-section for Z+jet production, differential in the Z boson rapidity, y^Z , for (top) $p_T^{\text{jet}} > 20$ GeV and (bottom) $p_T^{\text{jet}} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). Superimposed are predictions as described in Fig. 6.14.

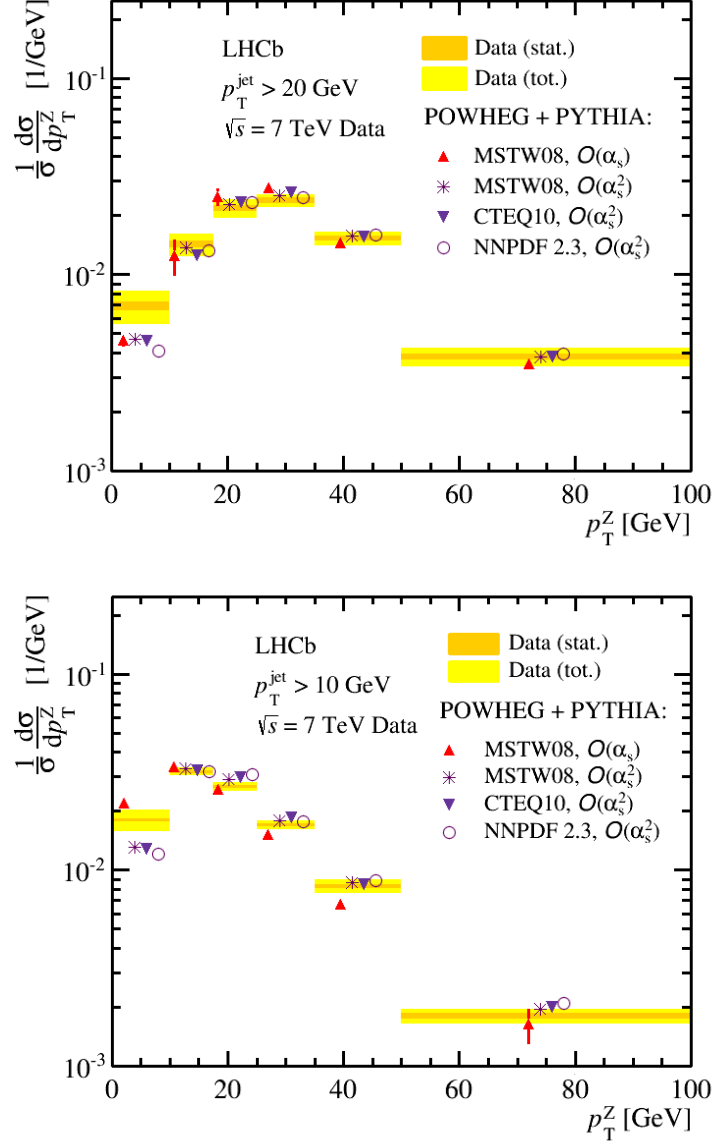


Figure 6.17: Cross-section for Z+jet production, differential in the Z boson transverse momentum, for (top) $p_T^{\text{jet}} > 20$ GeV and (bottom) $p_T^{\text{jet}} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). Superimposed are predictions as described in Fig. 6.14.

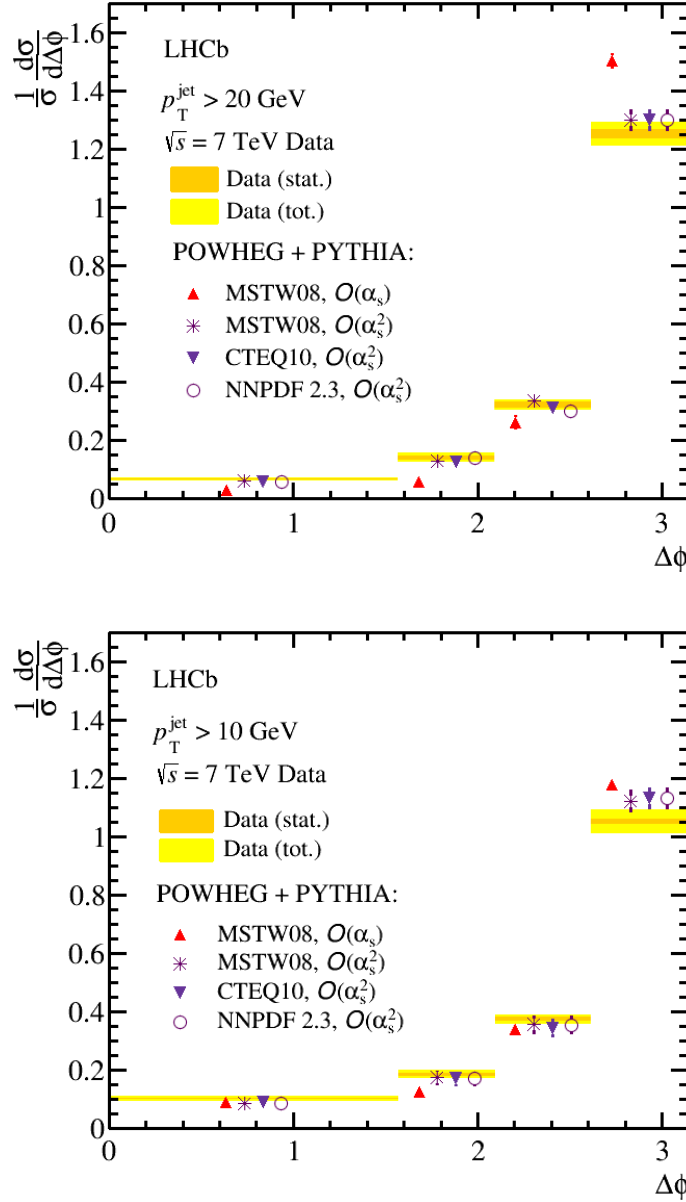


Figure 6.18: Cross-section for Z+jet production, differential in the modulus of the difference in ϕ between the Z boson and the leading jet, for (top) $p_T^{\text{jet}} > 20$ GeV and (bottom) $p_T^{\text{jet}} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). Superimposed are predictions as described in Fig. 6.14.

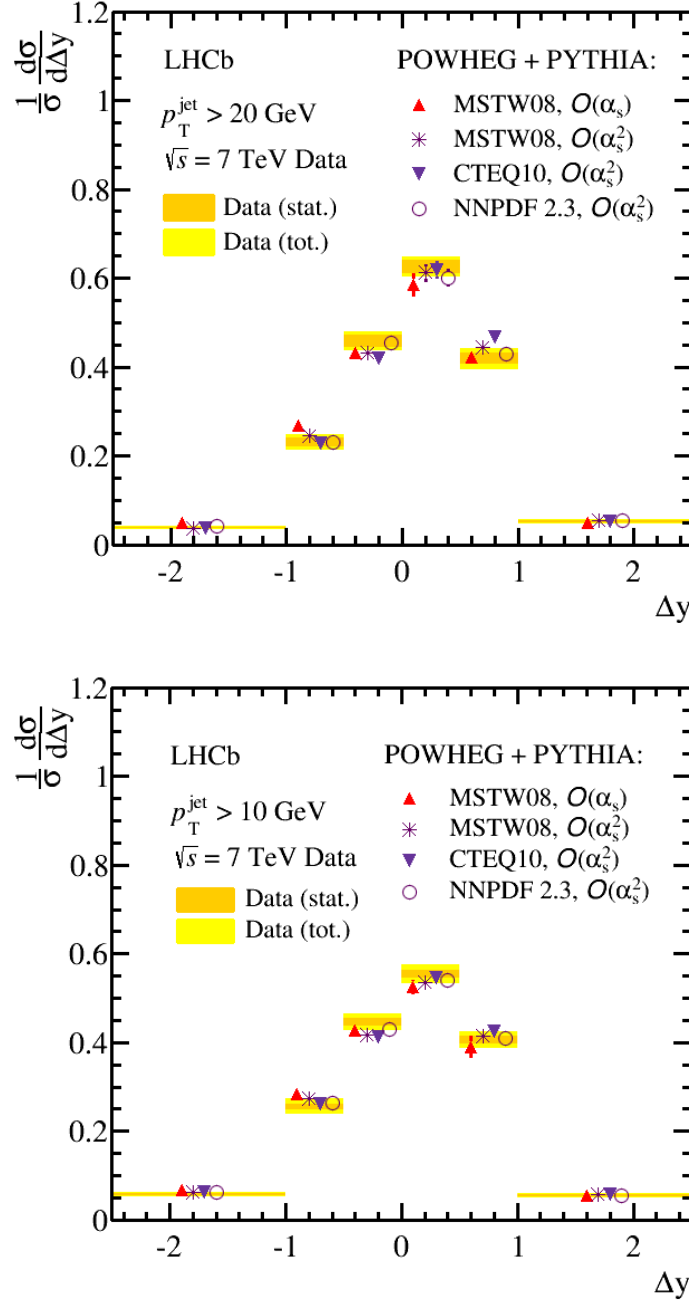


Figure 6.19: Cross-section for Z+jet production, differential in the difference in rapidity between the Z boson and the leading jet, for (top) $p_T^{\text{jet}} > 20$ GeV and (bottom) $p_T^{\text{jet}} > 10$ GeV. The bands show the LHCb measurement (with the inner band showing the statistical uncertainty and the outer band showing the total uncertainty). Superimposed are predictions as described in Fig. 6.14.

and the Z boson p_T , the peak of the Z boson p_T distribution (see Figure 6.17) shifts to higher values. In addition, the Z boson and the leading jet become more back-to-back. As the p_T of the leading jet increases it is less likely that another jet is produced in the event with a similar p_T [136]. Consequently, there is less other significant activity in the events which pass the higher jet p_T threshold, and the distribution of the difference in azimuthal angle between the leading jet and the Z boson (see Figure 6.18) favours more production of back-to-back events in order to ensure that the total momentum in the transverse plane balances. In all these differences between the 10 and 20 GeV thresholds the theoretical predictions match the data well.

However, we also note that the differential distribution measurements, like the ratios, show slightly better agreement between data and theory above the 20 GeV threshold than above the 10 GeV threshold. This is probably because soft, non-perturbative, effects become more important at low p_T [136]. Such effects are modelled within the parton shower approximation, but a degree of Monte Carlo tuning is expected to be needed to describe the data. Different tunes can have a noticeable effect on the jet p_T and correlated distributions where the parton showers are important (typically low p_T) [140, 152].

The potential importance of accurately modelling soft QCD behaviour can also be seen in the Z boson p_T spectrum at low p_T , for both jet transverse momentum thresholds. In this region no prediction completely reproduces the measured distribution. Similar behaviour was seen by the CMS experiment for the inclusive Z boson p_T distribution [152]. Non-perturbative physics dominates predictions in this region of phase-space [152], so such behaviour is not surprising. Indeed, further tuning of the parameters that govern non-perturbative behaviour in Monte Carlo based simulation of proton-proton interactions is expected to provide better agreement between data and the theoretical predictions [152].

Irrespective of the p_T threshold, the theoretical predictions tend to agree better with the measured differential cross-sections when the hard process is evaluated at $\mathcal{O}(\alpha_s^2)$. The leading jet p_T distribution, shown in Figure 6.14, clearly exhibits this behaviour: the $\mathcal{O}(\alpha_s)$ prediction tends to underestimate the number of high p_T jet events. The Z boson transverse momentum distribution behaves similarly, with the higher order calculations predicting more events with higher Z boson p_T , matching the data better. Since, the jet p_T and the Z boson p_T are highly correlated, this behaviour is expected. The difference in azimuthal angle between the Z boson and the leading jet shows similar behaviour. Parton level predictions of the hard process at $\mathcal{O}(\alpha_s)$ require the parton and the Z boson to be produced back-to-back. The theoretical principles underlying parton shower based predictions are only accurate in the limit where a parton splits collinearly; parton shower

models favour collinear splitting. It is for this reason that the predictions of the $\Delta\phi$ distribution at $\mathcal{O}(\alpha_s)$ fail to reproduce the data, even when matched to a parton shower as the parton shower fails to predict enough events where the azimuthal angle between the Z boson and the jet is small. At higher orders, more partons are produced in the hard process, and non-trivial angles between the Z boson and the partons are generated more easily. Consequently higher order predictions of the $\Delta\phi$ distribution contain more events where the Z boson and the jet are not produced back-to-back, and so the higher order predictions better match the data.

Unlike the distributions discussed above, the leading jet η , Z boson rapidity, and the difference between the rapidity of the Z boson and the leading jet show little dependence upon the perturbative order of the hard process predictions. This is because these distributions tend to measure how forward the event is, which tends to depend more on the momenta of the initial state partons than the evolution of the final state, since the conservation of momentum means that the initial state largely determines the rapidity. Therefore, such measurements are dependent on the PDFs of the colliding partons, probing their values of Bjorken- x ; we recall the approximate formula: $x_{1,2} \simeq \frac{M}{\sqrt{s}}e^{\pm y}$. Clearly, if the assumptions made in the DGLAP equations break down then the PDFs will not be extrapolated correctly from previously measured constraints: in such a case the data and theoretical predictions of the rapidity based observables are expected to disagree. Likewise, differences between the different PDF parametrisations, if present, are expected in these distributions, as these variables, unlike the jet p_T , are more dependent on the PDFs.

All the different PDF parametrisations studied agree with the data, confirming our understanding of the inner structure of the proton. However, there are hints of tension between the different PDF sets in the Δy distribution. This tension between PDF parametrisations highlights the sensitivity of the Δy distribution to the PDF parametrisation used. The NNPDF 2.3 parametrisation has been tuned on LHC data, unlike the other parametrisations, so it is therefore perhaps unsurprising that such tension is visible. Whilst the current uncertainties on the measurement mean that the data agree with all the parametrisations considered, the experimental uncertainties on future measurements are expected to decrease. Measurements of this distribution in the coming years are therefore expected to provide one way to test the performance of the different PDF parametrisations.

Regardless, it is clear that the measured cross-sections shown here agree very well with the Standard Model predictions. Whilst this measurement of Z+jet production

at LHCb does not provide an obvious channel to search for new physical effects and processes, it remains important to test the Standard Model in regions where it is expected to provide accurate predictions. In that sense, this measurement is a continuation of the achievements of the initial LHC measurements, both at the general purpose detectors and at LHCb, in that it provides another confirmation of the great success of the Standard Model in describing physics at high GeV and low TeV energy scales. The measurement also confirms our current understanding of the proton's inner structure, as the distributions potentially sensitive to PDFs show good agreement between data and theoretical predictions, for a range of PDF parametrisations.

6.7 Summary

We have discussed the measurement of the Z+jet production cross-section at LHCb, based on events where the Z boson decays to two muons. We have considered the effect of the jet transverse momentum resolution on the data, correcting for this and for detection efficiencies. We have also considered the systematic uncertainties associated with the measurement. The dominant uncertainty is the jet energy scale. A variety of sources contribute to this uncertainty. We discussed some of the sources earlier; in this chapter we focused on estimating the level to which the energy scale, determined from simulation, agrees with the data, by comparing the p_T balance in back-to-back Z+jet events.

The total Z+jet cross-section was measured for two jet transverse momenta thresholds, 20 and 10 GeV. Differential cross-sections were also measured as a function of kinematic variables associated with the Z and the jet. The measured cross-sections agree reasonably well with theoretical predictions based on fixed order matrix element calculations that are matched with parton showers. The predictions at $\mathcal{O}(\alpha_s^2)$ show better agreement with the p_T and $\Delta\phi$ distributions than predictions at $\mathcal{O}(\alpha_s)$, as these distributions are sensitive to higher order effects. The agreement of the data with these theoretical predictions provides another confirmation of pQCD at LHC energy scales, and also confirms our current understanding of the inner structure of the proton.

We now turn our attention to future related measurements at LHCb, and their potential impact.

Chapter 7

Related measurements for the coming years

“Predictions can be very difficult—especially [those] about the future.”

— Neils Bohr

The measurement of the Z+jet production cross-section in $\sqrt{s} = 7$ TeV data recorded by the LHCb experiment in 2011 opens up many other exciting possibilities. In the coming pages we will discuss some of these options. We begin by considering the impact and potential of repeating the measurement using data recorded in 2012 (at the higher centre-of-mass energy of 8 TeV). We then turn our attention to other related precision measurements that will provide important tests of the Standard Model, and extend the studies presented here. The Z+jet analysis presented earlier is important for such studies, as the good agreement with the Standard Model predictions means that we are confident in the performance of the LHCb jet reconstruction algorithms.

7.1 Z+jet measurements using data recorded in 2012

The LHCb collaboration is yet to publish any measurements of electroweak boson production using data recorded in 2012, and the jet reconstruction for this data-taking period is still to be finalised¹. However, it is instructive to consider this dataset and

¹In this section, seeing as the jet reconstruction is still being tuned, we do not apply a correction factor for the jet p_T . However, since studies have shown that the correction factor is close to unity at low p_T , the effect of this choice is small.

potential measurements using it. We will discuss here the feasibility and attractions of a measurement of the Z+jet production cross-section using this data.

The jet reconstruction performance in simulated Z+jet events at $\sqrt{s} = 8$ TeV using the default LHCb 2012 reconstruction is shown in Figure 7.1. The jet p_T resolution is similar to the values quoted earlier (11-13%), and is again dependent on p_T and the number of pile-up interactions. Whilst the performance of the jet reconstruction algorithms has remained broadly unchanged with respect to the data recorded in 2011, the knowledge of the detector alignment has improved significantly, especially at low pseudorapidities. The probability distribution associated with the track fit is shown in Figure 7.2. This distribution is noticeably flatter than the same distribution in the reconstruction of the 2011 dataset (see Figure 5.25 for comparison). This means that the tracking efficiency for muons associated with a Z boson decay is expected to increase by about 3-5% per muon, as the χ^2 requirement rejects fewer true events (with the most pronounced effect at low pseudorapidities). The global event cut efficiency is expected to decrease however², as the number of pile-up events is greater in the 2012 dataset, and the overall underlying event activity is greater in $\sqrt{s} = 8$ TeV collisions than $\sqrt{s} = 7$ TeV collisions.

The increase in the centre-of-mass energy in 2012 relative to 2011 means that the expected Z+jet cross-section is higher: results evaluated by the author at $\mathcal{O}(\alpha_s^2)$ using FEWZ [67] predict that the cross-section for Z+jet production in the LHCb fiducial acceptance defined previously is 9.7 pb (for $p_T^{\text{jet}} > 20$ GeV). This is an increase of about 40% on the predicted $\sqrt{s} = 7$ TeV Z+jet cross-section above the same p_T threshold. This increase in cross-section comes from two related effects. The higher centre-of-mass energy of the proton-proton collisions increases the overall inclusive Z cross-section by about 25%. In addition, the increased energy of the collisions also increases the fraction of events above any jet p_T threshold (relative to the total number of Z events).

The 2012 dataset recorded at LHCb corresponds to an integrated luminosity of about 2.1 fb^{-1} , roughly two times the amount of data recorded in 2011. This also increases the number of events seen and the statistical precision of any potential measurement.

The same selection requirements outlined in Section 5.2.2 can be applied to the 2012 dataset to determine the number of Z boson events and the number of Z+jet events recorded by the LHCb experiment at $\sqrt{s} = 8$ TeV. In the 2012 dataset there are roughly

²Preliminary studies by other members of the LHCb collaboration suggest that this efficiency drops by about 3% relative to the 2011 dataset [176].

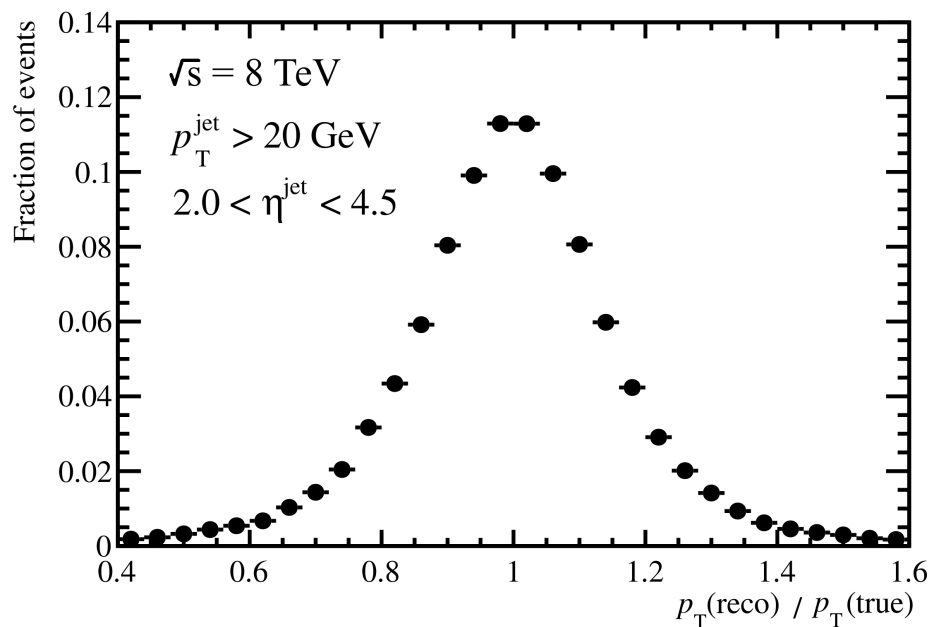


Figure 7.1: The jet p_T resolution in simulation using the LHCb reconstruction for data recorded in 2012, for $p_T^{\text{jet}} > 20$ GeV. The matching between reconstructed and truth jets is performed using the $R - p_T$ algorithm discussed earlier.

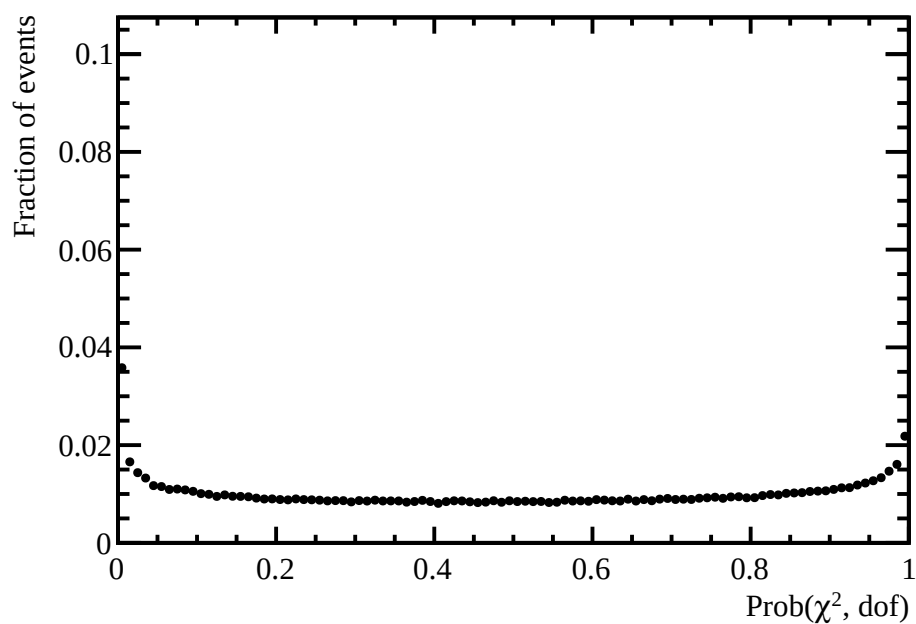


Figure 7.2: The quality of the track fit in the 2012 dataset, for muons selected as potentially being produced in Z boson decays.

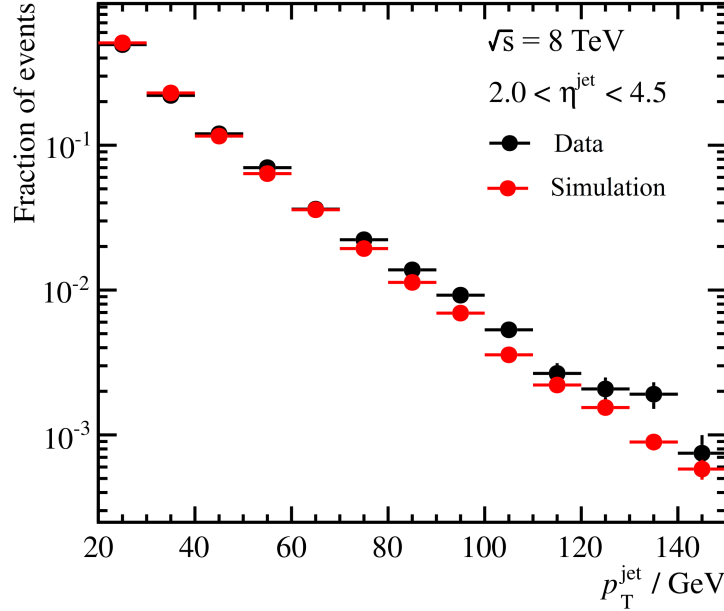


Figure 7.3: The leading jet p_T distribution in data (black) and simulation (red), for Z+jet events produced in $\sqrt{s} = 8 \text{ TeV}$ pp collisions. No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

three times the number of events in the 2011 dataset: about 145 thousand Z boson events are reconstructed, with roughly thirty thousand events containing a jet with p_T above the 10 GeV threshold and about twelve thousand events containing a jet above the 20 GeV threshold. We briefly consider here the events selected in the 2012 dataset with $p_T^{\text{jet}} > 20 \text{ GeV}$. A sample of Z+jet events is simulated using the same conditions as discussed in Section 5.2.1, except that the sample discussed here corresponds to the higher centre-of-mass energy recorded in 2012. The jet transverse momentum distribution and the Z boson rapidity distribution are shown in Figures 7.3 and 7.4 respectively for both the 2012 data and simulation. Relatively good agreement is seen; comparisons of data and simulation show similar features to those outlined in the discussion in Section 5.2.3, so we shall not discuss these distributions further here.

The increase in the number of events in the 2012 dataset means that the statistical uncertainty on any Z+jet measurement is expected to be about 1.7 times smaller than that on the same 2011 measurement. In addition, the dominant systematic uncertainty on the 2011 measurement was based on the degree to which the jet energy scale (determined using simulation) could be validated on data. This validation was also statistically limited, and, as such, this uncertainty is also expected to decrease within the 2012 dataset, potentially by roughly the same factor (providing this uncertainty remains

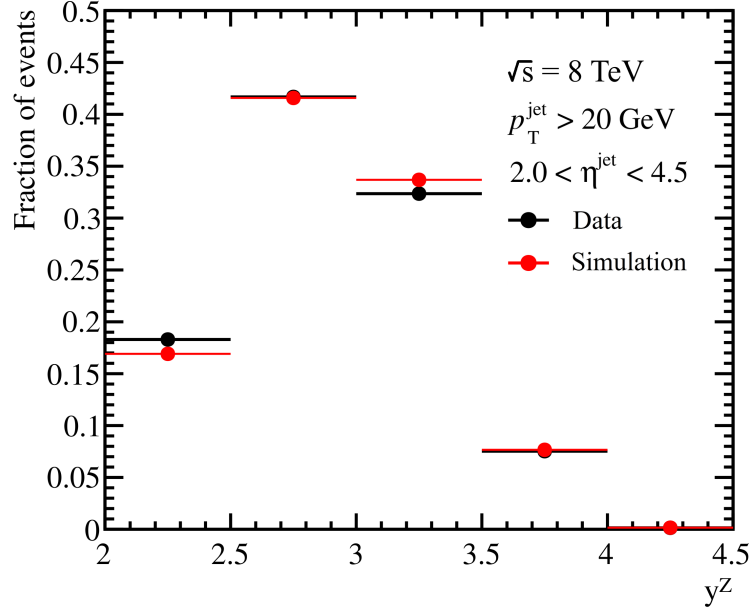


Figure 7.4: The Z boson rapidity distribution in data (black) and simulation (red), for Z+jet events produced in $\sqrt{s} = 8$ TeV pp collisions. No corrections for detection inefficiencies or resolution effects have been made. The uncertainties shown are statistical.

statistically limited), yielding a total jet energy scale uncertainty of about 4%. The uncertainty associated with the jet identification requirements will need to be studied further for any measurement based on the 2012 dataset, as its level within the 2011 analysis is comparable to this expected jet energy scale uncertainty in the 2012 dataset. However, it is not unrealistic to also expect that this uncertainty could decrease in any future analysis, given that the systematic uncertainty currently assigned is conservative, and that there are more events available to test potential discrepancies between data and simulation in the 2012 sample than the 2011 sample. The current uncertainty on the recorded integrated luminosity in the 2012 data taking period is about 1.2% [99, 100]. This is significantly lower than the 3.5% luminosity uncertainty on the 2011 dataset; the jet energy scale is still expected to provide the dominant uncertainty in future measurements. Considering these different contributions, measurements of $\frac{\sigma(\text{Z+jet})}{\sigma(\text{Z})}$ at the 5% level should be feasible. Such measurements will enable LHCb to further probe PDF physics at low Bjorken- x .

A direct comparison of the reconstructed (detector-level) distributions in the 2011 dataset with the same distributions in the 2012 dataset is not valid as various efficiencies have changed with the improved reconstruction available in the 2012 dataset (for example

the variations in the track reconstruction efficiency with the muon pseudorapidity are much smaller in the 2012 dataset [156]). We therefore turn to theoretical predictions to briefly examine the physics sensitivity of a $\sqrt{s} = 8$ TeV measurement relative to the $\sqrt{s} = 7$ TeV measurement. We compare predictions evaluated at $\mathcal{O}(\alpha_s^2)$ within the LHCb Z+jet acceptance using FEWZ [67] and the MSTW08 PDF parametrisation for the two different centre-of-mass energies. These predictions are shown as a function of the Z boson rapidity in Figure 7.5. In addition to the increased cross-section at the higher centre-of-mass energy the events produced are also noticeably more forward: the region probed shifts to moderately higher rapidities as the centre-of-mass energy increases. This means that a potential 8 TeV measurement probes slightly different physics than the 7 TeV measurement: the presence of more events at higher rapidities and the higher centre-of-mass energy means that LHCb is able to probe lower values of Bjorken- x (as can be seen through the relation $x_{1,2} \simeq \frac{M}{\sqrt{s}} e^{\pm y}$). Such measurements will therefore be an important contribution to LHCb's measurement of production cross-sections in the forward region in the coming years.

7.2 Other measurements at LHCb

With the increased number of events available for analysis in the 2012 dataset, LHCb has increased sensitivity to other variables. The 2012 dataset contains roughly seven hundred events containing at least two jets with p_T above the 20 GeV threshold. This is enough events for LHCb to perform a statistically sensitive measurement of the jet multiplicity distribution. In addition, differential measurements can be made as function of correlations between the kinematic variables associated with these jets. Such measurements would enable detailed probes of pQCD, and have already been made by other LHC experiments in different kinematic regions [145, 147].

There are ongoing efforts [177] to extend the LHCb jet reconstruction programme by developing jet flavour-tagging algorithms that will allow LHCb measurements to determine whether a jet is heavy-flavour initiated (and if so, whether the jet was initiated by a b or a c quark). Preliminary studies of these algorithms suggest that they can provide a light-parton initiated jet rejection rate of 99.5% for a b-tagging efficiency of 50% [177]. For the same efficiency the charm quark initiated jet rejection rate is about 98% [177]. This tagging is being used in an ongoing analysis of Z+b jet production [178], which probes heavy flavour PDFs, building on the work presented in this thesis. However,

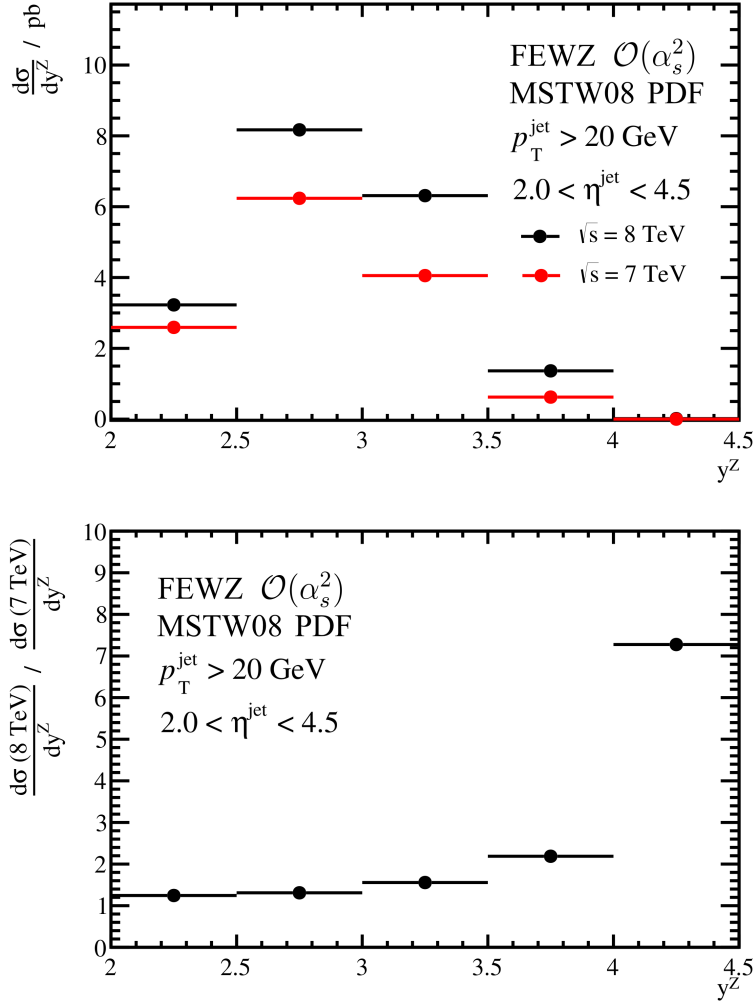


Figure 7.5: The predicted Z boson rapidity distributions for different centre-of-mass energies. The top plot shows the differential cross-sections for Z+jet production as a function of the Z rapidity for the two collision energies, whilst the bottom plot then shows the ratio of these differential cross-sections. Uncertainties are not shown on these plots, as the scale uncertainties between the 7 TeV and 8 TeV cross-section predictions are highly correlated.

the cross-section for the production of Z bosons in association with heavy flavour particles in the LHCb acceptance is small [178, 179].

Regardless of such future studies of the Z+jet process, the Z+jet analysis presented here and published in Ref. [4] has provided an important benchmark result for other jet based analyses at LHCb, as the agreement between the Z+jet results and Standard Model predictions provides confidence in the standard LHCb jet reconstruction. We therefore consider here one more future jet measurement. LHCb has made measurements of inclusive W boson production in events where the W boson decays to a muon and a

neutrino [5]. A measurement of the W+jet production cross-section using such W boson decays would be complementary to similar ATLAS [180] and CMS [146] measurements. However, it should be noted that such a measurement is challenging for LHCb, as LHCb is not a hermetic detector and does not measure the total missing transverse energy in an event. This means that LHCb is not able to determine a reliable estimate of the direction and momentum of the neutrino in the transverse plane, and so LHCb is not able to use transverse mass requirements to select W boson events. Instead, a high p_T muon is required to indicate the presence of a W boson [5]. However, with no transverse mass requirement possible, background levels in the LHCb selection are higher than at ATLAS and CMS. It should be noted though that the physics interest in such a measurement is high: a measurement at LHCb has the potential to probe lower values of Bjorken- x than measurements at ATLAS and CMS. A measurement of the total W+jet (the sum of contributions from positive and negative W bosons) production cross-section would also be a pathway to a measurement of the ratio of the W+jet and Z+jet production cross-sections. A lot of the experimental systematic uncertainties (like the luminosity and jet energy scale) are expected to (largely) cancel in such a ratio, enabling a precision test of the electroweak and pQCD sectors of the Standard Model at the level of about 1%. Such measurements would be amongst the most precise at the LHC.

Historically, PDF fitting groups have relied on a few very precise measurements at colliders as inputs: typically measurements of inclusive (massive) vector boson production cross-sections, their ratios, and measurements of inclusive jet production cross-sections are combined with measurements of deep inelastic scattering to create a dataset which can be used to fit PDFs [140]. Measurements of the production cross-section of massive vector bosons in association with jets are not typically used as an input to these fits as theoretical predictions of these measurements are usually reasonably sensitive to different choices (in, for example, the choice of method used to match the parton shower to predictions of the hard process or from tunable parameters within the parton shower) [140]. Additionally, there has been little impetus experimentally for such measurements to be used, as their impact is usually reasonably similar to that of the (typically more precise) inclusive measurements [181]. The experimental uncertainties are usually correlated between such measurements as well. Measurements of massive vector boson and associated jet production are therefore typically used to probe the behaviour of PDFs (without being used to constrain them) or are used as benchmark measurements of Standard Model processes, validating the experimental performance. However, a recent paper by Malik and Watt (see Ref. [140]) suggests that the related measurements of inclusive W and Z boson production at high p_T can provide very important constraints on PDFs and

should be considered by PDF fitting groups³. Theoretical predictions are well understood when the p_T of the vector boson is larger than its mass as fixed order calculations are known to provide accurate results; matching methods, soft gluon resummation and other theoretical choices associated with calculations at lower p_T play a negligible role [140] here. Such high p_T inclusive measurements can therefore provide important constraints on the gluon PDF [140]. With its different kinematic coverage to the general purpose detectors, LHCb can make complementary measurements to ATLAS and CMS, adding additional power to constrain PDFs. Such measurements are therefore of great interest in the coming years.

³These measurements are also an important way to constrain potential backgrounds to new physics signals [140].

Chapter 8

Conclusions

“I’d take the awe of understanding over the awe of ignorance any day.”

— Douglas Adams

The initial run of the Large Hadron Collider finished in early 2013. The measurements made using the first data have written a new chapter in the history of particle physics: the overwhelming success of the Standard Model at low TeV energy scales. First conceived in the 1970s, no collider experiment has yet produced results which place this theory in serious jeopardy. Within the initial LHC run the ATLAS and CMS experiments have discovered a particle with properties similar to the postulated Higgs boson, and no LHC experiment has yet published significant evidence for supersymmetry or other theories describing physics beyond the Standard Model.

The LHCb detector at the LHC was initially designed to study heavy flavour decays, in order to probe $\mathcal{CP}$ violation within the CKM matrix (and consequently theories of baryogenesis that explain the observed matter/anti-matter asymmetry in the universe). The LHCb experiment was also designed to search for new physics in very rare decays of particles containing b quarks. Since the majority of B hadrons are produced in the forward (and backward) region the LHCb detector has a limited angular coverage ($2 < \eta < 4.5$). However, the LHCb detector is able to make complementary measurements to the general purpose detectors in studies of electroweak boson production. These studies tend to probe pQCD, and are sensitive to parton distribution functions at low Bjorken- x . The thesis therefore began by discussing the theoretical considerations that underlie these different areas. This was followed by a detailed discussion of the LHCb detector. We also

discussed the triggering and selection of events at LHCb, given the author's involvement in this work.

We have considered the reconstruction of jets at LHCb. Measurements of jet production provide a test of pQCD, but are also often a key signature in events described by physics beyond the standard model. The ability to reconstruct jets is therefore an important extension to LHCb's initial mission. Given the forward coverage of the detector, LHCb tends to be sensitive to jets with $p_T < 100$ GeV. The LHCb experiment uses a particle flow algorithm to reconstruct jets, using as much information as possible from the tracking systems, and only relying on the calorimeters to reconstruct neutral particles. The jet p_T resolution is typically 11-13% (depending upon the jet p_T , and the definition used for the resolution). This performance is comparable to that of the general purpose detectors for jets with similar p_T .

The thesis then documented a measurement of the production cross-section of Z+jet events within the LHCb fiducial acceptance above two jet p_T thresholds: 10 and 20 GeV. This measurement was also published in Ref. [4], and was the first published LHCb measurement to include jets. The kinematic coverage of LHCb for this measurement is unique at the LHC: both the Z boson and at least one jet must be produced in the forward region. This analysis provides a test of pQCD and probes parton distribution functions at low Bjorken- x . The measured Z+jet cross-section, and the ratio of this cross-section with respect to the inclusive Z cross-section, for $p_T^{\text{jet}} > 20$ GeV are,

$$\begin{aligned}\sigma(\text{Z+jet}) &= 6.3 \pm 0.1 \text{ (stat.)} \pm 0.5 \text{ (syst.)} \pm 0.2 \text{ (lumi.) pb,} \\ \frac{\sigma(\text{Z+jet})}{\sigma(\text{Z})} &= 0.083 \pm 0.001 \text{ (stat.)} \pm 0.007 \text{ (syst.)},\end{aligned}$$

and the corresponding measurements for $p_T^{\text{jet}} > 10$ GeV are

$$\begin{aligned}\sigma(\text{Z+jet}) &= 16.0 \pm 0.2 \text{ (stat.)} \pm 1.2 \text{ (syst.)} \pm 0.6 \text{ (lumi.) pb,} \\ \frac{\sigma(\text{Z+jet})}{\sigma(\text{Z})} &= 0.209 \pm 0.002 \text{ (stat.)} \pm 0.015 \text{ (syst.)}.\end{aligned}$$

These results and the measured differential distributions show good agreement with Standard Model theoretical predictions based on matrix element calculations evaluated at $\mathcal{O}(\alpha_s^2)$ using POWHEG [7–11] that are matched to PYTHIA [12] for simulation of the parton shower and hadronisation. This agreement was seen for three different PDF

parametrisations, confirming our current understanding of the inner structure of the proton, and providing another confirmation of pQCD at LHC energy scales.

The agreement of the LHCb Z+jet measurement with the Standard Model predictions can also be viewed as providing confidence in the performance of the LHCb jet reconstruction algorithms. We therefore discussed potential future measurements with jets at LHCb. We considered both the impact of repeating the measurement with the 2012 dataset and other related measurements (like that of the W+jet production cross-section), and ratios with respect to the Z+jet production cross-section. We also considered other similar (inclusive) measurements: Ref. [140] suggests that the related measurement of the inclusive vector boson p_T distribution (at large p_T) may provide new inputs which can be used to constrain PDFs. Such measurements will probe the inner structure of the proton and test the Standard Model.

The coming years in particle physics will see a host of interesting measurements. The Standard Model will continue to be probed at higher energies (the next LHC run is expected to have a centre-of-mass energy of 13 TeV). These measurements will allow precision tests of both the electroweak sector and pQCD. Measurements sensitive to parton distribution functions will provide related tests of the DGLAP equations. There will also be direct searches for new particles, and indirect searches for new effects. Together, these measurements will test our knowledge of the universe at a new energy frontier. The Standard Model has provided extremely accurate predictions, but we know that it is not an ultimate theory, as it has no answer to the question of the nature of dark matter, and it provides no quantum theory of gravity. The coming years may not answer these mysteries, but the principle of naturalness will be placed under the microscope: just how fine-tuned is nature? With few hints of possible new effects and mysteries, we must ask the fundamental question: what's next?

Colophon

This thesis was made in $\text{\LaTeX} 2_{\epsilon}$ using the “hepthesis” class [182]. Feynman diagrams have been produced using Axodraw [183].

Bibliography

- [1] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716**, 1 (2012), 1207.7214.
- [2] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716**, 30 (2012), 1207.7235.
- [3] Daya-Bay collaboration, F. An *et al.*, *Observation of electron-antineutrino disappearance at Daya Bay*, Phys.Rev.Lett. **108**, 171803 (2012), 1203.1669.
- [4] LHCb collaboration, R. Aaij *et al.*, *Study of forward $Z + \text{jet}$ production in pp collisions at $\sqrt{s} = 7 \text{ TeV}$* , JHEP **1401**, 033 (2014), 1310.8197.
- [5] LHCb collaboration, R. Aaij *et al.*, *Inclusive W and Z production in the forward region at $\sqrt{s} = 7 \text{ TeV}$* , JHEP **1206**, 058 (2012), 1204.1620.
- [6] LHCb collaboration, R. Aaij *et al.*, *Measurement of the cross-section for $Z \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7 \text{ TeV}$* , JHEP **1302**, 106 (2013), 1212.4620.
- [7] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO vector-boson production matched with shower in POWHEG*, JHEP **0807**, 060 (2008), 0805.4802.
- [8] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411**, 040 (2004), hep-ph/0409146.
- [9] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711**, 070 (2007), 0709.2092.
- [10] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **1006**, 043 (2010), 1002.2581.
- [11] S. Alioli, P. Nason, C. Oleari, and E. Re, *Vector boson plus one jet production in POWHEG*, JHEP **1101**, 095 (2011), 1009.5594.
- [12] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and manual*, JHEP **0605**, 026 (2006), hep-ph/0603175.

- [13] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579 (1961).
- [14] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967).
- [15] G. 't Hooft and M. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl.Phys. **B44**, 189 (1972).
- [16] A. Salam, *Weak and electromagnetic interactions*, in *Proceedings of the eighth Nobel symposium*, edited by N. Svartholm, p. 367, (1968).
- [17] M. A. Thomson, *Modern Particle Physics*, Cambridge University Press (2013).
- [18] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley (1995).
- [19] F. Mandl and G. Shaw, *Quantum Field Theory*, Wiley (2009).
- [20] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*, Cambridge University Press, 1996.
- [21] R. Batley, *Lectures on Gauge Field Theory*, Cambridge University (2010).
- [22] ATLAS collaboration, CMS collaboration, F. Santanastasio, *Exotic Phenomena Searches at Hadron Colliders*, p. 365 (2013), 1301.2521.
- [23] M. Dine, *Supersymmetry and String Theory: Beyond the Standard Model*, Cambridge University Press (2007).
- [24] J. Beringer *et al.* (Particle Data Group), *Review of Particle Physics, 2012-2013. Review of Particle Properties*, Phys. Rev. D **86**, 010001 (2012).
- [25] R. Feynman, *QED: The Strange Theory of Light and Matter*, Penguin Press Science (1985).
- [26] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-Order QED Contribution to the Electron $g-2$ and an Improved Value of the Fine Structure Constant*, Phys.Rev.Lett. **109**, 111807 (2012), 1205.5368.
- [27] D. Hanneke, S. Fogwell, and G. Gabrielse, *New Measurement of the Electron Magnetic Moment and the Fine Structure Constant*, Phys. Rev. Lett. **100**, 120801 (2008).
- [28] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964).
- [29] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12**, 132 (1964).
- [30] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964).

- [31] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964).
- [32] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145**, 1156 (1966).
- [33] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155**, 1554 (1967).
- [34] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys.Rev.Lett. **10**, 531 (1963).
- [35] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog.Theor.Phys. **49**, 652 (1973).
- [36] LHCb collaboration, H. Dijkstra *et al.*, *LHCb Letter of Intent*, 1995.
- [37] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys.Rev.Lett. **30**, 1343 (1973).
- [38] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys.Rev.Lett. **30**, 1346 (1973).
- [39] G. Ladinsky and C. Yuan, *The Nonperturbative regime in QCD resummation for gauge boson production at hadron colliders*, Phys.Rev. **D50**, 4239 (1994), hep-ph/9311341.
- [40] C. Balazs, J.-w. Qiu, and C. Yuan, *Effects of QCD resummation on distributions of leptons from the decay of electroweak vector bosons*, Phys.Lett. **B355**, 548 (1995), hep-ph/9505203.
- [41] C. Balazs and C. Yuan, *Soft gluon effects on lepton pairs at hadron colliders*, Phys.Rev. **D56**, 5558 (1997), hep-ph/9704258.
- [42] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Comm. **178**, 852 (2008), 0710.3820.
- [43] M. Bahr *et al.*, *Herwig++ Physics and Manual*, Eur.Phys.J. **C58**, 639 (2008), 0803.0883.
- [44] S. Frixione, F. Stoeckli, P. Torrielli, B. R. Webber, and C. D. White, *The MC@NLO 4.0 Event Generator*, (2010), 1010.0819.
- [45] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton Fragmentation and String Dynamics*, Phys.Rept. **97**, 31 (1983).
- [46] A. Kupco, *Cluster hadronization in HERWIG 5.9*, in *Proc. of the workshop on Monte Carlo Generators for HERA Physics 1998-1999* (1998), hep-ph/9906412.
- [47] G. P. Salam, *Towards Jetography*, Eur.Phys.J. **C67**, 637 (2010), 0906.1833.

- [48] J. E. Huth *et al.*, *Toward a standardization of jet definitions*, in *Proc. of the 1990 Summer Study on High Energy Physics, Research Directions for the Decade, Snowmass, Colorado* (1990).
- [49] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Longitudinally invariant k_t clustering algorithms for hadron hadron collisions*, Nucl.Phys. **B406**, 187 (1993).
- [50] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys.Rev. **D48**, 3160 (1993), hep-ph/9305266.
- [51] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better jet clustering algorithms*, JHEP **9708**, 001 (1997), hep-ph/9707323.
- [52] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- k_t jet clustering algorithm*, JHEP **0804**, 063 (2008), 0802.1189.
- [53] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys.Lett. **B641**, 57 (2006), hep-ph/0512210.
- [54] M. Dasgupta, L. Magnea, and G. P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, JHEP **0802**, 055 (2008), 0712.3014.
- [55] G. P. Salam, *Jet clustering tools*, Talk to MC4LHC readiness workshop (2010), <http://indico.cern.ch/event/74601/other-view?view=standard>.
- [56] J. D. Bjorken, *Current Algebra at Small Distances*, in *Proceedings of the International School of Physics Enrico Fermi Course XLI*, edited by J. Steinberger, pp. 55–81, (1968).
- [57] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63**, 189 (2009), 0901.0002.
- [58] R. D. Ball *et al.*, *Parton distributions with QED corrections*, Nucl.Phys. **B877**, 290 (2013), 1308.0598.
- [59] I. Hinchliffe and A. Kwiatkowski, *Parton model sum rules*, Ann.Rev.Nucl.Part.Sci. **46**, 609 (1996), hep-ph/9604210.
- [60] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl.Phys. **B126**, 298 (1977).
- [61] 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 (1977).
- [62] V. Gribov and L. Lipatov, *Deep inelastic $e p$ scattering in perturbation theory*, Sov.J.Nucl.Phys. **15**, 438 (1972).
- [63] R. Thorne, A. Martin, W. Stirling, and G. Watt, *Status of MRST/MSTW PDF sets*, (2009), 0907.2387.

- [64] H.-L. Lai *et al.*, *New parton distributions for collider physics*, Phys.Rev. **D82**, 074024 (2010), 1007.2241.
- [65] R. D. Ball *et al.*, *Parton distributions with LHC data*, Nucl.Phys. **B867**, 244 (2013), 1207.1303.
- [66] R. Thorne, A. Martin, W. Stirling, and G. Watt, *Parton distributions and QCD at LHCb*, in *Proc. of XVI Int. Workshop on Deep-Inelastic Scattering and Related Topics* (2008), 0808.1847.
- [67] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, *FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order*, Comput.Phys.Commun. **182**, 2388 (2011), 1011.3540.
- [68] J. M. Campbell and R. Ellis, *MCFM for the Tevatron and the LHC*, Nucl.Phys.Proc.Suppl. **205-206**, 10 (2010), 1007.3492.
- [69] F. Landry, R. Brock, P. M. Nadolsky, and C. Yuan, *Tevatron Run-1 Z boson data and Collins-Soper-Sterman resummation formalism*, Phys.Rev. **D67**, 073016 (2003), hep-ph/0212159.
- [70] S. Berge, P. M. Nadolsky, F. Olness, and C.-P. Yuan, *Transverse momentum resummation at small x for the Tevatron and CERN LHC*, Phys.Rev. **D72**, 033015 (2005), hep-ph/0410375.
- [71] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307**, 001 (2003), hep-ph/0206293.
- [72] F. Caravaglios, M. L. Mangano, M. Moretti, and R. Pittau, *A New approach to multijet calculations in hadron collisions*, Nucl.Phys. **B539**, 215 (1999), hep-ph/9807570.
- [73] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, JHEP **0902**, 007 (2009), 0811.4622.
- [74] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur.Phys.J. **C45**, 97 (2006), hep-ph/0506026.
- [75] G. Corcella *et al.*, *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **0101**, 010 (2001), hep-ph/0011363.
- [76] G. Corcella *et al.*, *HERWIG 6.5 release note*, Online documentation (2002), hep-ph/0210213.

- [77] S. Agostinelli *et al.*, *Geant4 a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).
- [78] J. Allison *et al.*, *Geant4 developments and applications*, Nuclear Science, IEEE Transactions on **53**, 270 (2006).
- [79] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3**, S08001 (2008).
- [80] LHCb collaboration, A. A. Alves Jr *et al.*, *The LHCb Detector at the LHC*, Journal of Instrumentation **3**, S08005 (2008).
- [81] LHC Study Group, T. S. Pettersson and P. Lefvre, *The Large Hadron Collider: conceptual design*, CERN Report No. CERN-AC-95-05 LHC, 1995 (unpublished).
- [82] C. Hill *et al.*, *Tests of the CERN proton linac performance for LHC - type beams*, eConf **C000821**, TUD17 (2000), physics/0008102.
- [83] C. E. Hill, A. M. Lombardi, E. Tanke, and M. Vretenar, *Present performance of the CERN proton linac*, p. 4 p (1998).
- [84] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3**, S08003 (2008).
- [85] CMS collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3**, S08004 (2008).
- [86] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3**, S08002 (2008).
- [87] M. L. Mangano, *Two lectures on heavy quark production in hadronic collisions*, (1997), hep-ph/9711337.
- [88] LHCb RICH group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73**, 2431. 25 p (2012).
- [89] O. Deschamps, *LHCb Calorimeters - Commissioning & performance*, Talk to 12th Topical Seminar on Innovative Particle and Radiation Detectors (2010), www.bo.infn.it/sminiato/sm10/presentations/s13/o-deschamps.pdf.
- [90] A. A. A. Jr *et al.*, *Performance of the LHCb muon system*, Journal of Instrumentation **8**, P02022 (2013).
- [91] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8**, P04022 (2013), 1211.3055.
- [92] J. Albrecht, V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, (2013), 1310.8544.

- [93] G. Haefeli *et al.*, *The LHCb DAQ interface board TELL1*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **560**, 494 (2006).
- [94] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering **82**, 35 (1960).
- [95] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl.Instrum.Meth. **A262**, 444 (1987).
- [96] B. Jost, *The LHCb DAQ system*, in *Nuclear Science Symposium Conference Record, 2000 IEEE*, pp. 3.26.1–3.26.6, (2000).
- [97] LHCb collaboration, R. Aaij *et al.*, *Absolute luminosity measurements with the LHCb detector at the LHC*, JINST **7**, P01010 (2012), 1110.2866.
- [98] P. Koppenburg *et al.*, *Nu, Mu and Pile-Up: LHCb definitions of what we see and what we don't*, <https://twiki.cern.ch/twiki/bin/view/LHCb/NuMuPileUp>.
- [99] J. Panman, *Private communication*.
- [100] J. Panman, *Status of luminosity calibrations*, Talk to the 69th LHCb Collaboration Week (2013),
<https://indico.cern.ch/event/266735/>.
- [101] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, Ann. Math. Statist. **22**, 79 (1951).
- [102] M. Needham, *Clone Track Identification using the Kullback-Liebler Distance*, LHCb-2008-002, LHCb Public Documentation (2008).
- [103] J. van Tilburg, *Track Reconstruction in LHCb*, Talk to the LHCb UK Student Group (2013),
<http://indico.cern.ch/event/280053/contribution/1>.
- [104] M. Needham, *Performance of the LHCb Track Reconstruction Software*, LHCb-2007-144, LHCb Public Documentation (2007).
- [105] H. Terrier and I. Belyaev, *Particle identification with LHCb calorimeters*, LHCb-2003-092, LHCb Public Documentation (2003).
- [106] C. Jones, *Private communication*.
- [107] C. Jones, *ANN PID Retuning for Reco14 Data*, Talk to Physics Performance Meeting (2013),
<https://indico.cern.ch/event/226062/contribution/1>.
- [108] M. Clemencic *et al.*, *Recent developments in the LHCb software framework Gaudi*, J.Phys.Conf.Ser. **219**, 042006 (2010).

- [109] M. Cattaneo, *Boole and Brunel: Status and Plans*, Talk at 46th LHCb Week (2007),
<https://indico.cern.ch/event/20202/>.
- [110] F. Stagni *et al.*, *LHCbDirac: distributed computing in LHCb*, Journal of Physics: Conference Series **396**, 032104 (2012).
- [111] U. Egede, *Ganga and distributed analysis in LHCb*, Talk to CHEP (2006),
<http://indico.cern.ch/event/048/session/9/contribution/317>.
- [112] T. Shears and R. McNulty, *EOLAS: online documentation*,
<https://twiki.cern.ch/twiki/bin/view/LHCbPhysics/Eolas>.
- [113] T. Ruf, *PANORAMIX: online documentation*,
<https://twiki.cern.ch/twiki/bin/view/LHCb/PanoRamixWiki>.
- [114] M. Clemencic *et al.*, *The LHCb Simulation Application, Gauss: Design, Evolution and Experience*, Journal of Physics: Conference Series **331**, 032023 (2011).
- [115] I. Beiyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, in *Nuclear Science Symposium Conference Record (NSS/MIC)*, pp.1155-1161 (2010).
- [116] D. J. Lange *et al.*, *The EvtGen Event Generator Package*, in *Proceedings of CHEP* (1998).
- [117] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462**, 152 (2001).
- [118] BaBar collaboration, B. Aubert *et al.*, *The BaBar detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **479**, 1 (2002).
- [119] M. Cattaneo, *LHCb Digitisation Simulation*, Talk at LHCb Advanced Simulation Workshop (2012),
<https://indico.cern.ch/event/175918/>.
- [120] R. Nandakumar, *LHCb on the Grid*,
www.gridpp.ac.uk/gridpp20/gridpp20lhcb.ppt.
- [121] W. Barter *et al.*, *Jet reconstruction and performance at LHCb*, LHCb-INT-2012-015, LHCb Internal Documentation (2012).
- [122] V. Salustino Guimares, M. Rangel, V. Coco, and C. Potterat, *Energy-Momentum ratio of charged particles at LHCb*, LHCb-INT-2013-034, LHCb Internal Documentation (2013).
- [123] M. Santana Rangel, R. Ticse Torres, V. Coco, and C. Potterat, *Jet energy studies for RunI simulation*, LHCb-INT-2013-036, LHCb Internal Documentation (2013).

- [124] W. Barter, A. Bursche, and D. R. Ward, *Study of forward Z +jet production in pp collisions at $\sqrt{s} = 7$ TeV*, LHCb-ANA-2013-056, LHCb Internal Documentation (2013).
- [125] V. Coco *et al.*, *Jet Reconstruction at LHCb: Online Documentation*, <https://twiki.cern.ch/twiki/bin/view/LHCbPhysics/JetReco>.
- [126] V. Salustino Guimares, M. Rangel, V. Coco, and C. Potterat, *Energy-Momentum ratio of charged particles at LHCb*, LHCb-INT-2013-034, LHCb Internal Documentation (2013).
- [127] V. Coco, *Private communication*.
- [128] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur.Phys.J. **C72**, 1896 (2012), 1111.6097.
- [129] M. Green *et al.*, *Electron-Positron physics at the Z*, IOP (1998).
- [130] O. Francisco, *Jet-ID at LHCb*, Talk to Tools for Jet Reconstruction at LHCb Working Group (2012), <https://indico.cern.ch/event/231232/contribution/1>.
- [131] A. Höcker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT**, 040 (2007), physics/0703039.
- [132] M. Rangel, *Private communication*.
- [133] ATLAS collaboration, *Determination of the ATLAS jet energy measurement uncertainty using photon-jet events in proton-proton collisions at $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2011-031, ATLAS-COM-CONF-2011-011 (2011).
- [134] ATLAS collaboration, G. Aad *et al.*, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur.Phys.J. **C73**, 2306 (2013), 1210.6210.
- [135] The CMS collaboration, S. Chatrchyan *et al.*, *Determination of jet energy calibration and transverse momentum resolution in CMS*, Journal of Instrumentation **6**, P11002 (2011).
- [136] W. J. Stirling, *Private communication*.
- [137] H1 and ZEUS collaborations, F. Aaron *et al.*, *Combined Measurement and QCD Analysis of the Inclusive $e^\pm p$ Scattering Cross Sections at HERA*, JHEP **1001**, 109 (2010), 0911.0884.
- [138] LHCb collaboration, R. Aaij *et al.*, *A study of the Z production cross-section in pp collisions at $\sqrt{s} = 7$ TeV using tau final states*, JHEP **1301**, 111 (2013), 1210.6289.
- [139] LHCb collaboration, R. Aaij *et al.*, *Graphical comparison of the LHCb measurements of W and Z boson production with ATLAS and CMS*, LHCb-CONF-2013-005 (2013).

- [140] S. A. Malik and G. Watt, *Ratios of W and Z cross sections at large boson p_T as a constraint on PDFs and background to new physics*, (2013), 1304.2424.
- [141] L. Gribov, E. Levin, and M. Ryskin, *Semihard processes in QCD*, Physics Reports **100**, 1 (1983).
- [142] P. B. Arnold and M. H. Reno, *The complete computation of high p_t W and Z production in second-order QCD*, Nucl.Phys. **B319**, 37 (1989).
- [143] W. Giele, E. N. Glover, and D. A. Kosower, *Higher order corrections to jet cross-sections in hadron colliders*, Nucl.Phys. **B403**, 633 (1993), hep-ph/9302225.
- [144] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: electroweak gauge boson rapidity distributions at NNLO*, Phys.Rev. **D69**, 094008 (2004), hep-ph/0312266.
- [145] ATLAS collaboration, G. Aad *et al.*, *Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **1307**, 032 (2013), 1304.7098.
- [146] CMS collaboration, S. Chatrchyan *et al.*, *Jet Production Rates in Association with W and Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV*, JHEP **1201**, 010 (2012), 1110.3226.
- [147] CMS collaboration, S. Chatrchyan *et al.*, *Event shapes and azimuthal correlations in $Z +$ jets events in pp collisions at $\sqrt{s} = 7$ TeV*, Phys.Lett. **B722**, 238 (2013), 1301.1646.
- [148] CMS collaboration, S. Chatrchyan *et al.*, *Rapidity distributions in exclusive $Z +$ jet and photon + jet events in pp collisions at $\sqrt{s}=7$ TeV*, (2013), 1310.3082.
- [149] P. M. Nadolsky *et al.*, *Implications of CTEQ global analysis for collider observables*, Phys. Rev. **D78**, 013004 (2008), 0802.0007.
- [150] B. Webber, *Private Communication*.
- [151] T. Shears, *Private communication*.
- [152] CMS collaboration, S. Chatrchyan *et al.*, *Measurement of the Rapidity and Transverse Momentum Distributions of Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV*, Phys.Rev. **D85**, 032002 (2012), 1110.4973.
- [153] R. Gauld and A. Nomerotski, *Top Quark Physics in the High Eta Range*, LHCb-INT-2012-010, LHCb Internal Documentation (2012).
- [154] H. Brown, *Private Communication*.
- [155] LHCb collaboration, R. Aaij *et al.*, *Measurement of the cross-section for $Z \rightarrow \mu\mu$ production with 1 fb^{-1} of pp collisions at $\sqrt{s}=7$ TeV*, LHCb-CONF-2013-007 (2013).

- [156] S. Farry, *Update on 2011 Efficiency Studies*, Talk to the LHCb Electroweak Working Group (2013),
<https://indico.cern.ch/event/253804/>.
- [157] J. Anderson *et al.*, *Collated Muon ID Efficiency Numbers*,
<https://twiki.cern.ch/twiki/bin/view/LHCbPhysics/MuonEff>.
- [158] G. J. Feldman and R. D. Cousins, *A Unified approach to the classical statistical analysis of small signals*, Phys.Rev. **D57**, 3873 (1998), physics/9711021.
- [159] R. McNulty, *Efficiencies for Exclusives*, Talk to the LHCb Electroweak Working Group (2013),
<https://indico.cern.ch/event/253804/>.
- [160] A. Jaeger, P. Seyfert, M. De Cian, J. van Tilburg, and S. Hansmann-Menzemer, *Measurement of the track finding efficiency*, LHCb-PUB-2011-025, LHCb Public Documentation (2011).
- [161] J. Anderson *et al.*, *Collated Muon ID Efficiency Numbers*,
<https://twiki.cern.ch/twiki/bin/view/LHCbPhysics/TrackingEff>.
- [162] D. R. Ward, *Measurement of the cross-section for $Z^0 \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV*, LHCb-ANA-2012-078, LHCb Internal Documentation (2012).
- [163] T. Bayes and R. Price, *An Essay towards solving a Problem in the Doctrine of Chances*, Phil.Trans. **53**, 370 (1763).
- [164] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, Nucl.Instrum.Meth. **A362**, 487 (1995).
- [165] A. Höcker and V. Kartvelishvili, *SVD approach to data unfolding*, Nucl.Instrum.Meth. **A372**, 469 (1996), hep-ph/9509307.
- [166] T. Adye, *Unfolding algorithms and tests using RooUnfold*, ArXiv e-prints (2011), 1105.1160.
- [167] M. Bahr *et al.*, *Herwig++ Physics and Manual*, Eur.Phys.J. **C58**, 639 (2008), 0803.0883.
- [168] K. Hamilton, *Private Communication*.
- [169] K. Hamilton and P. Richardson, *Simulation of QED radiation in particle decays using the YFS formalism*, JHEP **0607**, 010 (2006), hep-ph/0603034.
- [170] D. Yennie, S. C. Frautschi, and H. Suura, *The infrared divergence phenomena and high-energy processes*, Annals Phys. **13**, 379 (1961).
- [171] T. Shears, *QED final state radiation corrections for electroweak processes*, LHCb-INT-2012-022, LHCb Internal Documentation (2012).

- [172] K. Petridis and M. Williams, *Measurement of the Forward-Central Beauty-Quark Asymmetry at LHCb*, LHCb-ANA-2013-007, LHCb Internal Documentation (2014).
- [173] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys.Rev. **D82**, 074018 (2010), 1005.3457.
- [174] A. Sherstnev and R. Thorne, *Parton Distributions for LO Generators*, Eur.Phys.J. **C55**, 553 (2008), 0711.2473.
- [175] R. Thorne, *Private communication*.
- [176] D. R. Ward, *$Z \rightarrow ee$ at 8 TeV*, Talk to the LHCb Electroweak Working Group (2013),
<https://indico.cern.ch/event/285474/contribution/5>.
- [177] W. Barter, *Current and future jet measurements at LHCb*, Talk to the LHCb Implication Workshop (2013),
<https://indico.cern.ch/event/255380/session/4/contribution/35>.
- [178] M. Rangel *et al.*, *b -tagging in $Z+b$* , Talk to Tools for Jet Reconstruction at LHCb Working Group (2013),
<https://indico.cern.ch/event/288941/contribution/1>.
- [179] LHCb collaboration, R. Aaij *et al.*, *Observation of associated production of a Z boson with a D meson in the forward region*, (2014), 1401.3245.
- [180] ATLAS collaboration, G. Aad *et al.*, *Study of jets produced in association with a W boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys.Rev. **D85**, 092002 (2012), 1201.1276.
- [181] K. Müller, *Private communication*.
- [182] A. Buckley, *The hepthesis L^AT_EX class*.
- [183] J. Vermaseren, *Axodraw*.

List of Figures

2.1	Loops in QED require renormalisation in order to keep predictions finite	13
2.2	The difference between the Born level and the bare level.	20
2.3	The shape of jets in the same event that are clustered using different algorithms.	26
2.4	The sum of the squares of the expected differences in jet p_T between the parton and hadron levels.	28
2.5	PDFs at different values of x and Q^2	30
3.1	The CERN site	36
3.2	The CERN accelerator complex	37
3.3	Production angles of $b\bar{b}$ pairs at the LHC	39
3.4	Beam crossing angles at LHCb	40
3.5	The LHCb detector	41
3.6	The layout of the VELO	43
3.7	The RF foils in the VELO	44
3.8	The R and $\phi(\text{Phi})$ sensors in the VELO.	44
3.9	The position resolution of primary vertices at LHCb	46
3.10	The LHCb magnet	48
3.11	The magnetic field throughout the LHCb detector	48
3.12	The coverage of the IT and OT within the T1-3 stations	49

3.13	The design of the RICH 1 detector	52
3.14	The Cherenkov angles of the different radiators used in RICH 1 and RICH 2	53
3.15	The performance of the RICH sub-detectors in LHCb data	53
3.16	The design of the RICH 2 detector	54
3.17	Schematic to show the shower structure within the LHCb calorimeters . .	56
3.18	The design of the PRS, SPD and ECAL	57
3.19	The design of the HCAL	58
3.20	The design of the muon chambers	59
3.21	The design of each muon chamber	60
3.22	The LHCb trigger in 2012	62
3.23	The total integrated luminosity delivered to and recorded by the LHCb detector in 2010, 2011 and 2012	69
3.24	The efficiency with which the LHCb detector was able to collect the luminosity provided by the LHC machine	70
3.25	The average value of μ during 2011 fills at LHCb	70
3.26	The average pile-up in bunch crossings at LHCb in 2011 fills	71
3.27	Estimates of μ from different detector sub-systems in 2010 data	73
4.1	The relationships between the pseudorapidity, momentum and transverse momentum of jets	80
4.2	Expected energy deposits in the ECAL for given track momenta	83
4.3	The mean fraction of the jet's p_T carried by different types of particles within the jet, at truth level.	85
4.4	The mean fraction of the jet's p_T carried by different types of particles within the jet in simulation at reconstructed level.	86
4.5	The distributions of fake jets and 'good' jets in the mpt, mtf and n_{Pointing} distributions	88

4.6	The jet p_T correction factor as a function of the jet p_T , η , and the number of primary vertices	91
4.7	Typical values of the uncertainty in the jet energy correction factor . . .	93
4.8	The distribution of $\Delta R(\text{truth, reco})$	95
4.9	The jet direction resolution at LHCb, as a function of the jet p_T . The statistical uncertainties are too small to be visible on this plot.	96
4.10	The ability of LHCb to reconstruct the jet p_T	98
4.11	The resolution of the jet p_T at LHCb	99
5.1	The values of x and Q^2 probed by different experiments.	103
5.2	PDF uncertainties on standard model measurements of electroweak vector boson production as a function of rapidity.	104
5.3	Previous results in electroweak vector boson production at LHCb	105
5.4	The lowest order Feynman Diagrams for Z+jet production	107
5.5	The lowest order Feynman Diagram for Z production	107
5.6	A Z+jet event measured at LHCb	110
5.7	The leading jet transverse momentum distribution	113
5.8	The leading jet pseudorapidity distribution	114
5.9	The measured Z boson transverse momentum distribution	114
5.10	The measured Z rapidity distribution	115
5.11	The measured difference in rapidity between the Z boson and the leading jet in the event	115
5.12	The measured difference in azimuthal angle between jets and Z bosons .	116
5.13	The measured invariant mass distributions for inclusive Z and inclusive Z+jet candidate events, for $p_T^{\text{jet}} > 10$ GeV	119
5.14	The measured invariant mass distributions for inclusive Z and inclusive Z+jet candidate events, for $p_T^{\text{jet}} > 20$ GeV	120

5.15	The tag-and-probe method of measuring efficiencies	122
5.16	The muon identification efficiency, binned as a function of the jet multiplicity and the muon pseudorapidity	124
5.17	The muon identification efficiency, binned as a function of the muon pseudorapidity	125
5.18	The invariant mass distribution of the tag and probe pairs considered when determining the muon identification efficiency	126
5.19	The muon identification efficiency, binned in the muon ϕ	127
5.20	The muon identification efficiency, measured on simulation using two different methods	128
5.21	The invariant mass distribution of the tag and probe pairs considered when determining the track efficiency	130
5.22	The muon track reconstruction efficiency	131
5.23	The muon track reconstruction efficiency, binned in the muon ϕ	132
5.24	The muon track reconstruction efficiency, measured on simulation, using the tag-and-probe method, and from MC truth information.	133
5.25	The χ^2 probability distribution for the tracks used to determine the track reconstruction efficiency, and the muon tracks selected in the Z+jet analysis	134
5.26	The single muon trigger efficiency, binned as a function of the muon η	135
5.27	The single muon trigger efficiency, measured on simulation, using the tag-and-probe method, and from MC truth information.	137
5.28	The distribution of n_{SPD} for the two different trigger routes.	138
5.29	An example of the gamma function fit used to determine the number of candidates that would have had more than 900 SPD hits.	140
5.30	The distribution of the number of primary vertices in the reconstructed Z+jet events.	142
6.1	The response matrix for migrations in the jet p_{T}	149

6.2	The unfolded jet p_T distribution for the D'Agostini and SVD unfolding methods.	154
6.3	The unfolded jet p_T distribution, found using two different numbers of iterations within the D'Agostini method.	154
6.4	The jet reconstruction and selection efficiency as a function of the jet p_T	156
6.5	The jet reconstruction and selection efficiency as a function of the true jet pseudorapidity.	156
6.6	The reconstructed and true p_T^Z distributions in simulation.	159
6.7	The p_T balance of the selected back-to-back Z+1 jet events in simulation at truth level	164
6.8	The p_T balance of the selected back-to-back Z+1 jet events in data and simulation	165
6.9	The level of agreement between data and simulation measured using a χ^2 comparison as a function of β	166
6.10	The level of agreement between data and simulation measured by comparing their means as a function of β	167
6.11	The n_{Pointing} and mtf distributions in data and simulation.	172
6.12	The inclusive Z cross-section measured here compared to previous measurements and a theory prediction.	177
6.13	The ratio of the Z+jet production cross-section to the inclusive Z cross-section for the two jet p_T thresholds.	181
6.14	The Z+jet production cross-section, differential in the leading jet p_T . . .	182
6.15	The Z+jet production cross-section, differential in the leading jet pseudorapidity, for the two jet transverse momentum thresholds.	183
6.16	The Z+jet production cross-section, differential in the Z boson rapidity, for the two jet transverse momentum thresholds.	184
6.17	The Z+jet production cross-section, differential in the Z boson transverse momentum, for the two jet transverse momentum thresholds.	185

6.18	The Z+jet production cross-section, differential in the difference in azimuthal angle between the Z boson and the leading jet.	186
6.19	The Z+jet production cross-section, differential in the difference in rapidity between Z boson and the leading jet.	187
7.1	The jet p_T resolution in simulation using the LHCb reconstruction for data recorded in 2012	193
7.2	The quality of the track fit in the 2012 dataset	193
7.3	The leading jet p_T distribution in data recorded in 2012	194
7.4	The measured Z boson rapidity distribution in data recorded in 2012 . .	195
7.5	The predicted Z boson rapidity distributions for different centre-of-mass energies	197

List of Tables

2.1	The fundamental bosons in the Standard Model	11
2.2	The fundamental leptons in the Standard Model	11
2.3	The fundamental quarks in the Standard Model	12
3.1	Beam crossing angles at LHCb	40
3.2	Typical Level-0 trigger lines and their rates	64
3.3	Typical HLT1 trigger lines for electroweak production physics studies at LHCb and their rates	66
3.4	Typical HLT2 trigger lines for electroweak production physics studies at LHCb and their rates	67
3.5	Three typical stripping lines in the EW stream	77
5.1	Global event cut efficiencies, measured as a function of the number of primary vertices and the jet multiplicity	141
6.1	The systematic uncertainties associated with measuring the Z+jet production cross-section.	161
6.2	The values of β where we see best agreement between data and simulation	168
6.3	The agreement between data and simulation in the variables used to select jets.	174