Measurement of particle multiplicity and energy flow in pp collisions at 13 TeV with the LHCb detector

University of Liverpool

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by

Kārlis Dreimanis

Supervisor: Prof. Tara Shears

Oliver Lodge Laboratory
Department of Physics
University of Liverpool

August 2017

Abstract

Measurement of particle multiplicity and energy flow in pp collisions at 13 TeV with the LHCb detector

The LHC is the world's largest and highest-energy particle collider. The LHCb experiment is one of four main experiments at the LHC. In July/August 2015, at the beginning of Run-II of the collider, the LHCb detector collected no-bias data during the so-called 'Early Measurements' low intensity data taking run. The analysis described in this thesis, the measurement of particle multiplicity and energy flow in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, uses this dataset. This analysis is aimed at improving our understanding of the hadron-hadron interaction process in full and, more specifically, our knowledge of the contributions of the underlying event and multi-parton interactions to the observable final states.

The data sample is split into the charged and neutral components and then further segregated into four event classes - inclusive minimum-bias, hard scattering, diffractive enriched and non-diffractive enriched. The measurement is carried out over a 2D $(e \times \eta)$ space within the fiducial acceptance of $2.0 < p < 1000.0$ GeV/c and $2.0 < \eta < 5.0$ and a full detector unfolding is performed. The results are presented as 1D multiplicity and energy flow distributions as a function of η and compared to four sets of theoretical predictions - Pythia 8 LHCb, Pythia 8 Monash 2013, EPOS LHC and Sibyll 2.1.

It is found that the LHCb tune of the Pythia 8.212 generator is able to describe the data the best, whilst the worst performing set of theoretical predictions is found to be EPOS LHC. Overall, the description of the inclusive minimum-bias and non-diffractive enriched event classes by Pythia 8.212 LHCb for both the multiplicity and energy flow is found to match the data well, while the theoretical predictions overestimate the charged component's multiplicity for the diffractive enriched event class and underestimate the neutral multiplicity for the neutral hard scattering event class.

Kārlis Dreimanis

University of Liverpool

Acknowledgements

I would like to start by thanking the entire Department of Physics at the University of Liverpool. The eight years of study here have been truly fulfilling and for that I am grateful to everyone with whom I studied for my undergraduate degree, to every lecturer who taught me along the way and to every undergraduate student I had the pleasure to teach during my Ph.D.

I must extend my deepest gratitude to my supervisor, Tara Shears. Despite her endless knowledge and experience and despite the lack of both of my own, her unparalleled positivity and encouragement helped me every step along the way. I always knew I could rely on her advice, help and guidance, no matter the date or the time of the day.

I also wish to thank Sergey Burdin and Tim Jones. Without their supervision during the last year of my masters degree, I doubt I would have had been given the opportunity to work towards this Ph.D. In addition, I would like to thank the Science and Technology Facilities Council for the provision of the funding, which allowed me to carry out my research.

The number of individuals who have played a huge part in making the last four years the unforgettable experience they have been is beyond count. I'd like to thank everyone in the QEE, HLT and VELO groups, especially Will Barter, Phil Ilten, Sascha Stahl, Mika Vesterinen, Conor Fitzpatrick and Kurt Rinnert. The amount of help I have received from them has been utterly invaluable. I wish to particularly thank Karol Hennessy for the vast amount and variety of support I have received from him over these four years.

My thanks go to everyone whom I have met and befriended over the time in Liverpool and at CERN; to the #CL - John Anders, Calum Macdonald and Jaime Norman - for the fact that only about 5% of any conversation we have ever had could actually be put down in these acknowledgements without being redacted considerably. The amount of hysterical laughter over ridiculous jokes we shared is only surpassed by the amount of beer, wine and other "encouraging" beverages we consumed; always "get two". I also would like to thank Carl Gwilliam, for all of his help with tackling pesky coding issues and only **nearly** running me over... twice.

Thanks to Espen Bowen and Jack Roberts, and the entire "Le Box" futsal team, for taking me in and making me feel "pretty in pink". Also thanks to the entire CERN

football mailing list and everyone with whom I shared the pitch, both in St Genis and Peron.

I also want to thank each and every colleague, too many to list, whom I shared the 4th floor Ph.D office with over these years. Although the faces have changed considerably since my first day, the discussions have remained as intellectually stimulating and the trips to the pub have stayed as entertaining, if blurry, as ever.

Furthermore, I must especially thank Stephen Farry. The amount of support he has given me over these four years is beyond immeasurable. Although I am sure he would say "You can't blame me for that!", he has been instrumental in developing every aspect of me as a researcher and a physicist.

To end with, I wish to thank everyone back home in Latvia and to my friends around the world. In particular I want to say thanks to Toms Skujiņš, for giving me so much to root for when I myself have felt a bit aimless, and to Edgars Zauls and Justs Zariņš, for always being there at the other end of the screen on "The Lounge".

Finally, I want to thank my family. To my parents, Ilze and Māris, for their endless parental sacrifice and effort, which has enabled me to be where I am today. To my grandparents, Ruta, Jānis, Ilga and Austris, who have aided immeasurably in stimulating my interest in the wider world and, subsequently, science. To my sister, Elīna, for being the most annoying and infuriating sibling in the entire world, in other words, for being the perfect "big" sister. Moreover, thanks to her and her partner, Ivo, for the three indescribably lovely little human beings they have brought into this world - Jēkabs, Rūdolfs and Emīlija. The stream of pictures and videos of their shenanigans have at times been quite literally the last and only thing that kept me going.

Thank you all

Declaration

This thesis is the result of my own work, except where a specific reference to the work of others is made. This thesis has not been submitted for any other qualification to this or any other university.

Kārlis Dreimanis

I would like to dedicate this work to my family

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER
ONE

Introduction

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!', but 'That's funny ...'"

- Isaac Asimov

High-energy particle physics is a field of science dedicated to the exploration and understanding of the most fundamental constituents of the physical and natural world along with the forces and laws that govern their behaviour.

The theoretical base of the field of particle physics is described with the use of quantum field theories, which, instead of viewing fundamental particles as rigid objects with definite size, shape and volume, treat them instead as excitations in the quantum fields. The current best quantum field theory framework, used to characterise particle physics, is the Standard Model. This theory is used to consistently describe the known fundamental particles and three of the four fundamental forces - electromagnetism, nuclear strong and nuclear weak force.

Although the Standard Model has been proven to be extremely robust and possesses an unparalleled predictive power, it is known to be incomplete. The most obvious omission within this framework is the lack of a description of the fourth fundamental force - gravity. However, besides this glaring shortcoming, the Standard Model does not encompass several other phenomena known to science. Both dark matter and dark energy remain unaccounted for in this framework and some fundamental properties of the known constituents of the Standard Model, such as the masses of the quarks, are still input as free parameters with no concrete explanation of the origin of their values.

In all fields of science, as long as theorists have developed a theoretical description of various aspects of nature, so have the experimentalists worked on confirming or

disproving the theories put forth. Experimental high-energy physics aims to test the predictions made by the Standard Model, as well as search for physics outside the realm of its boundaries, conveniently labelled as New Physics. In order to do so, large, powerful machines - particle colliders - are used. In such machines, subatomic particles are smashed together at high energies in an attempt to mimic the extremely hot and dense conditions of the early Universe, when many of the fundamental particles, which cannot be ordinarily observed in nature today, first came into existence.

The Large Hadron Collider (LHC) is the world's most powerful particle collider to date, providing proton-proton collisions at a centre-of-mass (c.o.m.) energy, $\sqrt{s}$, of up to $\sqrt{s} = 13$ tera-electronvolts (TeV). The collider is complemented by large particle detectors, which are used to explore the collision products. In the most simplistic terms, these detectors can be thought of as the world's most complex cameras, capturing incredibly short passages of time by observing many short-lived fundamental constituents of nature and documenting their properties.

The measurement of particle multiplicity and energy flow described in this thesis is performed using data collected by the Large Hadron Collider beauty (LHCb) experiment, one of the four major experiments at the LHC, during the summer of 2015. This measurement is aimed at improving our understanding of particle interaction processes in full, especially at low energy transfer scale, where perturbative calculations diverge and theoretical description remains a challenge.

This thesis is organised as follows. In chapter 2 the theoretical framework of high-energy particle physics is outlined, providing the fundamental grounds to the study presented in this thesis. In chapter 3 the experimental environment - the LHC and the LHCb experiment - is described in detail. The method and tools of turning the high-energy collisions provided by the LHC and recorded by LHCb into data streams that can be used in physics analysis is discussed in chapter 4. Chapter 5 is used to present the measurement of particle multiplicity and energy flow in pp collisions at 13 TeV. Finally, a brief discussion and summary is given in chapter 6.

In addition to the outline of the theoretical grounding and the experimental environment and the in-depth discussion of the measurement described in this thesis, certain sections, such as section 3.2.8 and section 4.1, are discussed in thorough detail. This is because the author has worked extensively on and dedicated a large portion of the time allowance of his post-graduate degree to the matter discussed in these sections.

CHAPTER
TWO

Theoretical overview

"Truth resists simplicity"

- John Green

This chapter is dedicated to the discussion of the theoretical grounding for modern high-energy particle physics. The first part of the chapter introduces the leading theory of particle physics - the Standard Model - in a complete, if brief, manner. Section 2.1.1 gives an overview of the framework, sections 2.1.2 to 2.1.5 describe the constituent theories of the Standard Model and section 2.1.6 summarises its shortcomings. The second part of this chapter focuses on the theoretical description of particle scattering processes. Section 2.2.1 briefly introduces scattering theory and sections 2.2.2 to 2.2.7 discuss the approaches to the calculation of the hard and soft scattering component's contributions to an observable final state. Finally, this chapter concludes with section 2.2.8, an introduction to the Monte-Carlo generators used for obtaining theoretical predictions for high-energy physics experiments.

2.1 The Standard Model

2.1.1 Overview

The Standard Model (SM) of Particle Physics [1–7] is a theoretical framework built over the 20^{th} century to describe and explain the existence and properties of fundamental particles and the forces that govern their behaviour. It contains two distinctive groups of fundamental particles - fermions and bosons - and encompasses the description of three of the four fundamental forces of the Universe - electromagnetic, weak and strong[1]. The

[1]The SM does not include a theory of gravity as further discussed in section 2.1.6.

Generation	Particle	Mass, MeV/c^2	Spin	Charge
I.	u	$2.2^{+0.6}_{-0.4}$	1/2	+2/3
	d	$4.7^{+0.5}_{-0.4}$	1/2	-1/3
	e	0.5109989461±0.0000000031	1/2	-1
	ν_e	< 0.00028	1/2	0
II.	c	1270±30	1/2	+2/3
	s	96^{+8}_{-4}	1/2	-1/3
	μ	105.6583745±0.0000024	1/2	-1
	ν_μ	< 0.00028	1/2	0
III.	t	173210±510±710	1/2	+2/3
	b	4180^{+40}_{-30}	1/2	-1/3
	τ	1776.86±0.12	1/2	-1
	ν_τ	< 0.00028	1/2	0

Table 2.1: Summary of the basic properties of the fundamental fermions in the Standard Model. Masses given are calculated using the $\overline{\text{MS}}$ renormalization scheme [12]. The charge is given in units of an electron charge e. Values taken from [13]. Neutrino masses taken from [14].

derivations in this section closely follow those outlined in Refs. [8–11].

Fermions are particles with an intrinsic angular momentum, spin-$\frac{1}{2}$ and, within the constraints of the SM, form the matter in the Universe. It is important to note that all fundamental fermions in the SM have this value of spin, while the definition of a fermion extends to encompass all particles with spin-$\frac{n}{2}$, where n is an odd-numbered integer. The fermions are then further split into two categories - quarks and leptons. There are three charged leptons - e (electron), μ (muon) and τ (tauon[2]) - and three corresponding neutral leptons - ν_e (electron neutrino), ν_μ (muon neutrino) and ν_τ (tau neutrino). The charged leptons all have a fundamental unit charge[3] of -1 and with their respective neutrinos are classed in three generations or 'flavours', with each generation's charged lepton being more massive. Similarly to leptons the six quarks are arranged in three generations of progressively increasing mass. Unlike with leptons, however, all six quarks are charged and each generation contains an up-type and a down-type quark having respective fractional charges of +2/3 and -1/3. The three up-type quarks are denoted as u (up), c (charm) and t (top) with the down-type quarks being d (down), s (strange) and b (beauty). The properties of the twelve fundamental fermions are summarised in Table 2.1.

While fermions are the particles that make up matter, bosons can be thought of as the particles which mediate the interactions of all fundamental particles and forces,

[2]It is customary to refer to this particle after its designated Greek letter - 'tau'.
[3]A fundamental unit charge refers to the electric charge of an electron of 1.602×10^{-19}C.

Field	Particle	Mass, GeV/c^2	Spin	Charge
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^0	125.09±0.21±0.11	0	0

Table 2.2: Summary of the basic properties of the fundamental boson particles in the Standard Model. The charge is given in units of an electron charge e. Values taken from [13].

hence the commonly used collective term for these constituents of the SM - the force carriers. Moreover, unlike fermions, bosons have an integer spin. There are six types of bosons in the SM - the photon, the gluon and the $W^{\pm}$, Z and Higgs bosons. The photon, γ, is the mediator of the electromagnetic interaction, eight gluons, g, are responsible for the strong interaction, the $W^{\pm}$ and Z bosons are the carriers of the weak force and the Higgs, H^0, governs the particle interaction with the Higgs field. The γ and g are massless and have no electric charge, the Z and H^0 are neutral, but massive, and the $W^{\pm}$ are both massive and have a unit electric charge of ± 1 respectively. The properties of the six types of fundamental bosons are summarised in Table 2.2.

The SM is constructed as a Quantum Field Theory (QFT) based on the gauge group denoted by $SU(3)_C \bigotimes SU(2)_L \bigotimes U(1)_Y$. Here, the various members describe the different symmetry groups combined in this theoretical framework. $SU(3)_C$ represents the symmetry group which describes the colour interaction of the quarks and gluons, Quantum Chromodynamics, and $SU(2)_L \bigotimes U(1)_Y$ denotes the symmetry group of the unified electroweak interaction.

2.1.2 Quantum electrodynamics

Quantum electrodynamics (QED) is a QFT which describes the electromagnetic interaction in particle physics. It is an Abelian[4] theory invariant under local $U(1)$ transformations. The derivations in this and the following two sections closely follow Refs. [8] and [9].

QED starts with the Dirac Lagrangian describing a particle with spin $s = \frac{1}{2}$ and mass m,

$$\mathcal{L}_{Dirac} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi, \tag{2.1}$$

[4]Abelian gauge theory, also referred to as a 'commutative gauge theory', is one for which the order of the application of operators does not alter the outcome of the operations.

where ψ is the wavefunction of the particle and γ^μ are the Dirac gamma matrices [1] with $\mu = \{0, 1, 2, 3\}$. This is invariant under the global gauge transformation

$$\psi(x) \to \psi'(x) = e^{i\alpha}\psi(x), \quad \bar{\psi}(x) \to \bar{\psi}'(x) = e^{i\alpha}\bar{\psi}(x), \tag{2.2}$$

where $i\alpha$ is a phase independent of spacetime position. The invariance, however, does not immediately hold locally, where the transformation, $e^{i\alpha(x)}$, depends on the chosen point in space

$$\psi(x) \to \psi'(x) = e^{i\alpha(x)}\psi(x). \tag{2.3}$$

The Lagrangian given in Equation 2.1 is not a satisfactory description of nature where such a Lagrangian should be independent of the space-time coordinates. This can be rectified by introducing a covariant derivative, D_μ, which is constructed such that, under a gauge transformation $D_\mu\psi(x)$, it would transform in the exact same way as $\psi(x)$ itself, meaning

$$D_\mu\psi(x) \to D'_\mu\psi'(x) = e^{i\alpha(x)}(D_\mu\psi(x)). \tag{2.4}$$

This can be achieved by choosing

$$D_\mu \equiv \partial_\mu + ieA_\mu, \tag{2.5}$$

where e is the electric charge and the gauge field A_μ transforms as

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{e}\partial_\mu\alpha(x). \tag{2.6}$$

Now, contracting $\not{D} = \gamma^\mu D_\mu$, the QED Lagrangian can be expressed as

$$\mathcal{L}_{QED} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.7}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.8}$$

is the electromagnetic field tensor and describes the propagation of the gauge field, A_μ, which corresponds to the vector boson field of the propagator mediating QED, ie. the photon. This expression is invariant under local gauge transformations given by Equations 2.3 and 2.6. It should be pointed out that A_μ was introduced with no associated mass term as doing so would have broken the invariance under the local

transformations. This gives rise to a need for a massless mediator implying that the photon must have zero mass, which is consistent with the experimental observations.

The interaction strength of QED in natural units[5] at the leading order (LO), so called *tree-level*, is given by

$$\alpha = \frac{e^2}{4\pi}. \tag{2.9}$$

QED calculations at this order clearly give finite results, however it is not so at higher orders such as next-to-leading order (NLO), so called *loop-level*. At the NLO and higher orders an infinite number of radiative corrections can be introduced via loops, forcing the calculations to diverge.

This can be rectified, due to QED being renormalizable. The in-depth description of renormalization, found in Refs. [8] and [9], is beyond the scope of this thesis; in broad terms it is possible, through this process, for all the loop contributions to be summed and divergences removed. This, in turn, requires the coupling strength to vary with the energy scale of the renormalization, μ. This is known as the 'running' of the coupling constant defined by the β-function

$$\beta(\alpha) = \mu^2 \frac{\partial \alpha}{\partial \mu^2}, \tag{2.10}$$

giving

$$\beta_{QED}(\alpha) = \frac{1}{3\pi}\alpha^2. \tag{2.11}$$

The positive sign of this function means that, in QED, α increases with an increase of the energy scale. This means, for example, that it takes progressively more energy to bring two carriers of like electric charges closer together and, conversely, the interaction strength between such charges decreases as the distance between them is increased.

Finally, the predictive power of this theory should be discussed. For example, the anomaly of the magnetic moment of an electron, a_e, the deviation from the LO prediction of zero, has been calculated to the tenth order [15] giving

$$a_e = 159652181.78(6)(4)(3)(0.77) \times 10^{-12}, \tag{2.12}$$

[5]Natural units are used in order to simplify complicated expressions by setting various constants to unity. In this case $\varepsilon_0 = \hbar = c = 1$, where ε_0 is permittivity of free space, $\hbar$ is the reduced Planck's constant and c is the speed of light.

and obtained experimentally [16] to be

$$a_e = 159652180.73(0.28) \times 10^{-12}, \tag{2.13}$$

with the difference between the two values of $1.06(0.82) \times 10^{-12}$. Such predictive power makes QED the most precisely tested theory of nature.

2.1.3 Quantum chromodynamics

Quantum Chromodynamics (QCD) is a QFT which describes the strong interaction between quarks and gluons, collectively known as partons. It is a non-Abelian gauge theory based on and invariant under the $SU(3)$ gauge group and its transformations. In the case of QCD, the charge around which the symmetry is built is the colour charge, thus the symmetry group used is called $SU(3)_C$. The colour fields - red (R), green (G) and blue (B) - form a field triplet given by

$$\psi(x) = \begin{pmatrix} \psi_R(x) \\ \psi_G(x) \\ \psi_B(x) \end{pmatrix}. \tag{2.14}$$

The Lagrangian of QCD is constructed similarly to QED and is given by

$$\mathcal{L}_{QCD} = \bar{\psi}_i(i\not{D}_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu}, \tag{2.15}$$

where the covariant derivative is

$$D^\mu_{ij} \equiv \delta_{ij}\partial_\mu + ig_s G^{a\mu}T^a_{ij}, \tag{2.16}$$

the kinetic or field propagating term is expressed as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \tag{2.17}$$

and the summation over all three field components is implied. This Lagrangian is invariant under the local transformation

$$\psi(x) \to \psi'(x) = e^{iT^a\alpha_a(x)}\psi(x). \tag{2.18}$$

The f^{abc} denotes the structure constants, which are related to the generators of the

$SU(3)$ group, T^a, following Lie algebra

$$[T^a, T^b] = if^{abc}T^c. \tag{2.19}$$

The number of generators of a symmetry group is given by N^2-1, where N is the number of fields in the theory. Therefore $SU(3)_C$, where $N_C = 3$, has $N_C^2 - 1 = 8$ generators known as the Gell-Mann matrices, which correspond to the number of gluons.

The coupling strength of QCD, α_s, takes a similar form to that of QED,

$$\alpha_s = \frac{g_s^2}{4\pi}, \tag{2.20}$$

and hence diverges at higher orders like its QED counterpart. As QCD is renormalizable, these divergences coming from loop level particle exchanges can be summed and subtracted. This implies the 'running' of the strong coupling constant with the variation in the renormalization scale. For QCD the β-function is given by

$$\beta_{QCD}(\alpha_S) = -\frac{11N_C - 2n_f}{12\pi}\alpha_S^2, \tag{2.21}$$

where N_C is again the number of colour fields ($N_C = 3$) and n_f is the number of quark flavours ($n_f = 6$). Thus the β-function in QCD is

$$\beta_{QCD}(\alpha_S) = -\frac{21}{12\pi}\alpha_S^2. \tag{2.22}$$

The negative value of the β-function implies that the strong coupling becomes weaker with increasing energy scale. This stems from the property of the gluon, the force carrier of the strong interaction, of having a colour charge itself, meaning that, unlike photons in QED, gluons can self-interact. Moreover, this leads to two important properties of QCD - confinement and asymptotic freedom.

Confinement arises from the strength of the strong coupling constant at low energies and leads to quarks being locked in colourless bound states called hadrons. These are either baryons, formed from three quarks of different colour (ie. RGB), or mesons, formed from a quark-antiquark pair (e.g. $G\bar{G}$). At these energy scales the perturbative QCD (pQCD) methods of describing the interactions cannot be used and non-perturbative approaches must be employed. It is due to confinement that quarks are never found in isolation in nature.

Asymptotic freedom is a property of the strong interaction, which manifests itself

at large energy scales where the coupling is weak. In this high-energy regime, partons within the hadrons become so energetic that they can be treated as free particles and pQCD calculations become valid.

2.1.4 Electroweak theory

The electroweak theory is the unification of the weak and electromagnetic interactions into a single theory. It describes the two interactions by introducing Dirac fields for left-handed and right-handed field components separately as left-handed doublets and right-handed singlets in the form of

$$\begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, l_R, u_R, d'_R, \tag{2.23}$$

where l and ν_l, u and d' represent a charged lepton and its corresponding neutrino, an up-type quark and a down-type quark's weak eigenstate respectively. The 'handedness' refers to a particle's helicity - the component of a particle's spin in its direction of motion. It is important to point out here that, under this theory, there are only left-handed neutrinos and right-handed anti-neutrinos. The electroweak Lagrangian consists of multiple components

$$\mathcal{L}_{EW} = \mathcal{L}_{fermions} + \mathcal{L}_{bosons} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{2.24}$$

and is required to be invariant under the transformations described by the gauge group $SU(2)_L \bigotimes U(1)_Y$, where the subscript Y refers to the weak hypercharge[6] [8].

The first component,

$$\mathcal{L}_{fermions} = \bar{\psi}_L \not{D}^L \psi_L + \bar{\psi}_R \not{D}^R \psi_R, \tag{2.25}$$

describes the interaction of the electroweak vector boson fields with the fermion fields from Equation 2.23 and covariant derivatives given by

$$D^L_\mu \equiv \partial_\mu - i g_W \frac{\sigma^i}{2} W^i_\mu - i g_B Y B_\mu \quad \text{and} \quad D^R_\mu \equiv \partial_\mu - i g_B Y B_\mu, \tag{2.26}$$

where g_W and g_B are the fermion coupling strengths to the W^i fields, describing the weak interaction, and B field, describing the electromagnetic interaction under the weak

[6]The weak hypercharge is the generator of the $U(1)_Y$ gauge group. In the electroweak theory it is associated with the B field [8].

hypercharge, respectively. Here and through this section the summation across the three generations of fermions is again implied. As required, this part is invariant under local $SU(2)_L \bigotimes U(1)_Y$ transforms, however it requires for four gauge bosons - W^1, W^2, W^3 and B - to be massless.

The second part of the $\mathcal{L}_{EW}$,

$$\mathcal{L}_{bosons} = -\frac{1}{4} W^i_{\mu\nu} W^i \mu\nu - \frac{1}{4} B^i_{\mu\nu} B^i \mu\nu, \tag{2.27}$$

characterises the field strength. The latter constituent is equivalent to the field strength tensor of QED given by Equation 2.8 whilst the former is given by

$$W^i_{\mu\nu} = \partial\nu W^i_\mu - \partial\mu W^i_\nu - g_W \varepsilon^{ijk} W^j_\mu W^k_\nu, \tag{2.28}$$

where ε^{ijk} are the structure constants satisfying the Lie algebra

$$[\frac{1}{2}\tau^i, \frac{1}{2}\tau^j] = i\varepsilon^{ijk}\tau^k, \tag{2.29}$$

with τ standing for the $N^2 - 1 = 3$ generators of the $SU(2)$ group, the Pauli matrices

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.30}$$

The four gauge fields discussed above relate to the four physical fields - $W^+_\mu, W^-_\mu, Z_\mu, A_\mu$ - through

$$W^1_\mu = \frac{1}{\sqrt{2}}(W^+_\mu + W^-_\mu,) \tag{2.31}$$

$$W^2_\mu = \frac{1}{\sqrt{2}}(W^+_\mu \pm W^-_\mu), \tag{2.32}$$

$$W^3_\mu = cos\,\theta_W Z_\mu + sin\,\theta_W A_\mu, \tag{2.33}$$

$$B_\mu = cos\,\theta_W A_\mu - sin\,\theta_W Z_\mu, \tag{2.34}$$

where θ_W is the weak mixing angle. These relations between the massless fields of the electroweak theory and the observable physical gauge boson fields occur under spontaneous symmetry breaking (SSB), which will be discussed in the following section.

2.1.5 Spontaneous symmetry breaking and the Higgs mechanism

So far in the discussion of the SM the masses of different particles have been ignored. This has not been an issue whilst describing QED and QCD, for which the respective gauge bosons - γ and g - are massless in nature. The problem of mass has arisen only partially as the matter particles, the fermions, upon which QED and QED fields act are massive. However, addition of these masses into the SM by hand, while unsatisfactory, does not break the theoretical framework. This ignorance can no longer be tolerated at this stage as the force carriers of the weak interaction - $W^{\pm}$ and Z - are known to have mass. In this case, adding the masses of the gauge bosons by hand would cause the gauge invariance to be broken. This is rectified elegantly by the introduction of the concept of SSB to the theory along with a new scalar field - the Higgs field.

The final two parts of the electroweak Lagrangian - $\mathcal{L}_{Higgs}$ and $\mathcal{L}_{Yukawa}$ - encompass these concepts to provide terms that can be associated with giving mass to the electroweak vector bosons and the particles of matter. To describe the former, a complex scalar Higgs field doublet, Φ, is introduced as

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.35}$$

with the associated scalar potential

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \tag{2.36}$$

This leads to $\mathcal{L}_{Higgs}$ given by

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger (D_\mu \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2. \tag{2.37}$$

Setting $\lambda < 0$ and $\mu^2 > 0$, leads to a trivial minimum of the potential of this Lagrangian at 0. Choosing $\mu^2 < 0$, however, it acquires an infinite number of degenerate minima given by

$$(\Phi^\dagger \Phi)_{min} = \frac{\mu^2}{2\lambda}. \tag{2.38}$$

In terms of real scalar fields, these minima can be expressed as

$$\frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{\mu^2}{2\lambda}. \tag{2.39}$$

Choosing the vacuum expectation value, v, of three of the components to be $\langle 0|\phi_1|0\rangle = \langle 0|\phi_2|0\rangle = \langle 0|\phi_4|0\rangle = 0$, a non-zero v can be assigned to ϕ_3

$$\langle 0|\phi_3|0\rangle = v^2 = \frac{\mu^2}{\lambda}. \tag{2.40}$$

Combining Equations 2.39 and 2.40, yields an infinite number of field minima given by

$$|\Phi| = \frac{v}{\sqrt{2}}. \tag{2.41}$$

This is the concept of SSB, where an initially symmetric system is spontaneously broken by a choice of a specific minimum of the potential.

Although the explicit derivation is beyond the scope of this thesis (see Refs. [8,9] for more detail), $\mathcal{L}_{Higgs}$ can be dissected further by expanding Φ around the chosen vacuum via setting $\phi_3 = H + v$ giving

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ H + v \end{pmatrix}. \tag{2.42}$$

This expansion and a unitary transform of the field Φ yields three significant results. Firstly, it gives an expression relating the mass of the Higgs boson to the value of μ via

$$M_H = \sqrt{-2\mu^2}. \tag{2.43}$$

Secondly, it allows three of the four physical vector bosons - W^+, W^- and Z - to gain mass and provides a relation between these masses given by

$$\frac{M_W^{\pm}}{M_Z} = cos\,\theta_W. \tag{2.44}$$

Finally, this expansion also allows for the fourth electroweak vector boson, γ, to remain massless, as required to satisfy QED. This is known as the electroweak Higgs mechanism.

The Higgs mechanism can be equally used to construct the final part of $\mathcal{L}_{EW}$ which gives rise to fermion masses in the SM in a natural and elegant way,

$$\mathcal{L}_{Yukawa} = -g_s \bar{\psi} \Phi \psi, \tag{2.45}$$

where the coupling constant, g_s, is related to the fermion mass, m_f, via

$$g_s = \frac{2m_f}{v}. \tag{2.46}$$

The fermion masses are determined experimentally and are input into the SM as free parameters.

2.1.6 Performance and the shortcomings of the Standard Model

An example of the predictive power of the SM has already been introduced in section 2.1.2 with the discussion of the anomaly of the electron's magnetic moment. This, however, is only one example in a vast array of predictions that can be made by the SM. These predictions culminated on the of 4^{th} of July 2012 with the announcement of the discovery of the Higgs boson [17, 18] - the last predicted but unseen constituent of the SM at that point. For the majority of the last half-century the experimental high-energy physics community has searched for signatures of the so-called Beyond the Standard Model (BSM) physics. These searches have, so far, been unsuccessful. Nevertheless, despite its tremendous predicting power, the SM is known to be incomplete and has some notable shortcomings.

Matter-antimatter asymmetry

The current theoretical description of the creation of the Universe requires an equal amount of matter and antimatter to have been created at its birth. At its current state, however, the Universe is matter dominated, meaning that there must exist a violation of the $\mathcal{CP}$ symmetry between the two, where $\mathcal{C}$ and $\mathcal{P}$ stand for charge and parity conjugate respectively. The former implies that the antimatter particle has the opposite charge of its matter counterpart, whereas the latter is a mirror image symmetry between the two. This imbalance is only partially accounted for in the SM via $\mathcal{CP}$ violation (CPV) in the quark sector. CPV was first discovered in decays of neutral kaons, where a decay forbidden in a case of no CPV was observed [19]. This observation implied that the weak eigenstate of a quark, q', is not equivalent to its mass eigenstate, q, ie. quarks have a non-zero probability to oscillate between different flavours in the electroweak decay process. This probability of various oscillations is given by the Cabibbo-Kobayashi-Maskawa

(CKM) matrix [20],

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.47}$$

where the magnitude of the matrix elements can be reduced to a combination of four variables, three mixing angles - θ_{12}, θ_{23} and θ_{13} - and a complex phase - δ. The magnitude of this complex phase is the degree of the $\mathcal{CP}$ violation and thus is a measure of the contribution to the matter-antimatter asymmetry from the quark sector of the SM. The current experimental observations imply that the magnitude of this phase, and thus the SM, can account for just a fraction of the matter-antimatter asymmetry evident in nature.

The current magnitudes of the elements in the CKM matrix obtained by a combination of world leading experimental results are given below [13].

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} 0.97417 \pm 0.00021 & 0.2248 \pm 0.0006 & (4.09 \pm 0.39) \times 10^{-3} \\ 0.220 \pm 0.005 & 0.995 \pm 0.016 & (40.5 \pm 1.5) \times 10^{-3} \\ (8.2 \pm 0.6) \times 10^{-3} & (40.0 \pm 2.7) \times 10^{-3} & 1.009 \times 0.031 \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.48}$$

Neutrino masses

Within the framework of the SM, the neutrinos are assumed to be massless. If this assumption was true no mixing between neutrino flavours would be possible. The discovery of neutrino oscillations [21], i.e. mixing of neutrino mass eigenstates, is evidence of neutrinos having mass [22]. These masses can be introduced into an extended SM through either adding a right-handed $SU(2)$ field singlet - ν_l - to the set already given in section 2.1.4 or producing the mass term via combination of the current left-handed neutrino field with its complex conjugate. The former approach leads to the neutrinos being referred to as Dirac, whereas the latter results in them being Majorana [23]. The latest indirect [14] and direct [24] measurements have put the upper constraints on the combined mass of the three flavours of neutrino, m_ν, to $m_\nu < 0.28$ eV and $m_\nu^2 < 5.6^2$ MeV^2, respectively.

The non-zero mass of the neutrinos leads to the mixing of their weak and mass eigenstates. The Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [21, 25] describes this mixing in a way similar to that of CKM matrix in the quark sector. Moreover, via their mixing, the neutrino sector could yield a further explanation of the matter-antimatter asymmetry.

Gravity and cosmological observations

Although the SM is the current best theory of particle physics, it does not account for further major observable phenomena. Firstly, it does not account for arguably the most evident force of nature - gravity. The absence of gravity is due to the fact that SM is a QFT, whereas no renormalizable quantized description of gravity currently exists, it being described by General Relativity [26], which is a continuum theory. This omission of gravity is tolerable only because gravity is weak in comparison to the other fundamental forces.

Furthermore, the SM only accounts for the baryonic matter, while it is known from cosmological observations that most of the mass of the Universe is contained within so-called dark matter [27, 28]. Finally, it is evident from cosmological studies that the Universe is expanding at an accelerated rate, a phenomenon currently explained via an introduction of a new source of energy - dark energy [29], also absent from the SM.

2.2 Hard and soft scattering

In order to be a valid description of nature a theory has to provide parameters and predictions that are experimentally testable. As mentioned earlier in section 2.1.2, for example, QED is a theory which has an immense predictive power that has been extensively tested.

Most tests of particle physics theory have used data gathered by high-energy scattering experiments where particle beams are accelerated and collided either with a static target or each other. The ability to translate theoretical predictions to physical observables, such as the cross-section, σ, of various processes in these scattering experiments is crucial in this approach. The components of these predictions are discussed in the following sections.

2.2.1 S-matrix

A scattering of two particles with initial momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ resulting in a production of n particles with momenta $\mathbf{k}_1$,...,$\mathbf{k}_n$ is shown in Figure 2.1. The assumption made here is that the timescale, t, at which the scattering takes place, is extremely short compared to the global time where particles evolve independently [8]. The initial and final states at large timescales can be viewed as free fields with the opaque bubble in Figure 2.1 containing all the information about the scattering process. Expressing all the initial

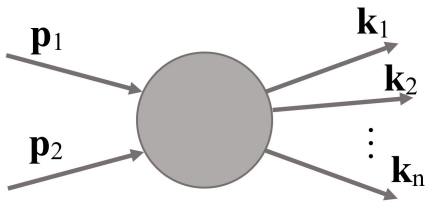

Figure 2.1: Scattering process of two particles with momenta $\mathbf{p}_1$ and $\mathbf{p}_2$, producing n final state particles with momenta $\mathbf{k}_1$,$\mathbf{k}_2$,...,$\mathbf{k}_n$.

states as $|\phi_i\rangle$ and all the final states as $|\phi_f\rangle$, the probability of the process in this bubble taking place can be expressed via a unitary scattering matrix S [10], components of which are given by

$$S_{fi} \equiv \langle \phi_f | \phi_i \rangle. \tag{2.49}$$

In the trivial case of a free field, ie. particle coming in, no scattering taking place and particle going out, S simply becomes an identity matrix and the relation holds. In a non-trivial case, where scattering does take place, S is expressed as

$$S = 1 + iT, \tag{2.50}$$

where the T holds all the information of the non-trivial case. Returning to the example in Figure 2.1, if the incoming particles and outgoing particles are denoted as states p_i and k_f , the matrix S can be expressed as

$$S_{fi} = \langle k_f | 1 + iT | p_i \rangle = \delta^2_{fi} + i2\pi^4 \delta^4(k_f - p_i)\mathcal{M}, \tag{2.51}$$

where $\mathcal{M}$ is the invariant amplitude [11] related to the cross-section of a process through

$$\frac{d\sigma}{d\Omega} = \frac{|\mathcal{M}|^2}{64\pi^2 E^2_{CM}}, \tag{2.52}$$

where $d\Omega$ is a fraction of a solid angle and E_{CM} is the centre-of-mass energy of the scattering process in question.

2.2.2 Feynman diagrams and the perturbative approach

Calculation of the invariant amplitudes of complex scattering processes is a daunting and time consuming task. There exists, however, a shorthand approach introduced by Richard Feynman - Feynman diagrams [30]. Expressed in the form of these diagrams the most basic non-trivial process - e^+e^- scattering - is given in Figure 2.2. Each constituent

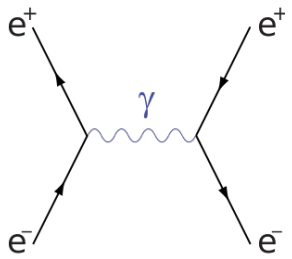

Figure 2.2: A diagram of a leading order electron-positron scattering [31].

of the diagram holds a physical and mathematical meaning. In momentum space, external lines, internal lines, or propagators, and vertices are assigned a mathematical expression given by Equations 2.53 to 2.55.

$$\text{External line:} \qquad \frac{i}{p_k^2 + m^2 + i\epsilon} \tag{2.53}$$

$$\text{Propagator:} \qquad \int \frac{d^4 k_j}{(2\pi)^4} \frac{i}{k_j^2 - m^2 + i\epsilon} \tag{2.54}$$

$$\text{Vertex:} \qquad -\frac{i\lambda}{4!}(2\pi)^4 \delta^4 \left(\sum momenta \right) \tag{2.55}$$

Here, λ is the order, to which the calculations are performed, ϵ is a mathematical tool in the form of a small offset in the complex momentum space, which ensures the convergence of the calculations, and p_k and k_j are the momenta of the scattering particles.

At LO, as in Figure 2.2, the benefit from this shorthand is superficial. It does become beneficial at NLO, however, where this LO diagram is complemented by six new diagrams where each of the four leptons can radiate a photon as well as exchange a photon between them before and after the scattering process. The divergence in the number of Feynman diagrams with increasing order of calculation is problematic. This implies an essentially infinite number of Feynman diagrams being present for even the simplest of processes with increasing energy scale, where increasingly higher order of calculations becomes relevant. The approximation of the solution to a set order of calculations is the perturbation theory [8]. Using this approach, the contributions to a physical process can be added up as a set of expanded series with a cut-off at a chosen energy scale. This does not manifest itself as a problem when dealing with QED, where the coupling constant grows in strength with increasing energy. For QCD, the strengthening of the strong coupling constant with decreasing energy leads to perturbative calculations becoming impossible in the low-energy regime where the coupling constant is large.

2.2.3 Factorisation theorem and parton distribution functions

The electron-positron annihilation discussed earlier is a scattering of two fundamental point-like objects. This point-like nature eases the calculation of observables in the scattering experiments. Hadrons are composite, which severely complicates these calculations. Consider, for example, a Drell-Yan process [32] as shown in Figure 2.3. It

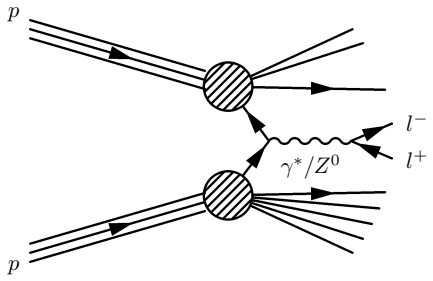

Figure 2.3: A Feynman diagram depicting the Drell-Yan process $pp \rightarrow \gamma^*/Z \rightarrow l^+l^-$ [33].

is obvious that, in comparison with the process in Figure 2.2, this situation is far less straightforward. Besides the hard scattering process which produces a pair of oppositely charged leptons, there are other contributions to the total final state for which the calculations need to account. Furthermore, it is evident that the incident protons did not participate in the creation of the di-lepton final state in their entirety, but rather only through a single parton each, with the rest of the constituents remaining as spectators. Thus, it is not possible to use the experimentally well known momentum of the whole proton in cross-section calculations. The solution to this problem comes in the form of the factorisation theorem and parton distribution functions (PDFs) [9]. The factorisation theorem can be expressed in the form of Equation 2.56,

$$\sigma_{(P_1P_2\rightarrow X)} = \int dx_1 dx_2 \sigma_{(p_1p_2\rightarrow X)}\ f_{p_1/P_1}(x_{p_1}, Q^2)\ f_{p_2/P_2}(x_{p_2}, Q^2), \tag{2.56}$$

where P_1 and P_2 are the momenta of the incident protons, p_1 and p_2 are the momenta of the partons of the respective protons related through $p_{1,2} = x_{1,2}P_{1,2}$ and $f_{p_{1,2}/P_{1,2}}(x_{p_{1,2}}, Q^2)$ are the PDFs of the given partons as a function of their fractional momenta, $x_{1,2}$, and the energy scale of the interaction, Q^2. The value of the hard scattering component, $\sigma_{p_1p_2\rightarrow X}$, can be calculated via a perturbative approach whereas PDFs, which are probability density functions for a parton to carry a certain fractional momentum at the energy scale considered, cannot be calculated using the same method. Examples of parton PDFs from the CTEQ6M [34] PDF set for two different values of

Q^2 are given in Figure 2.4.

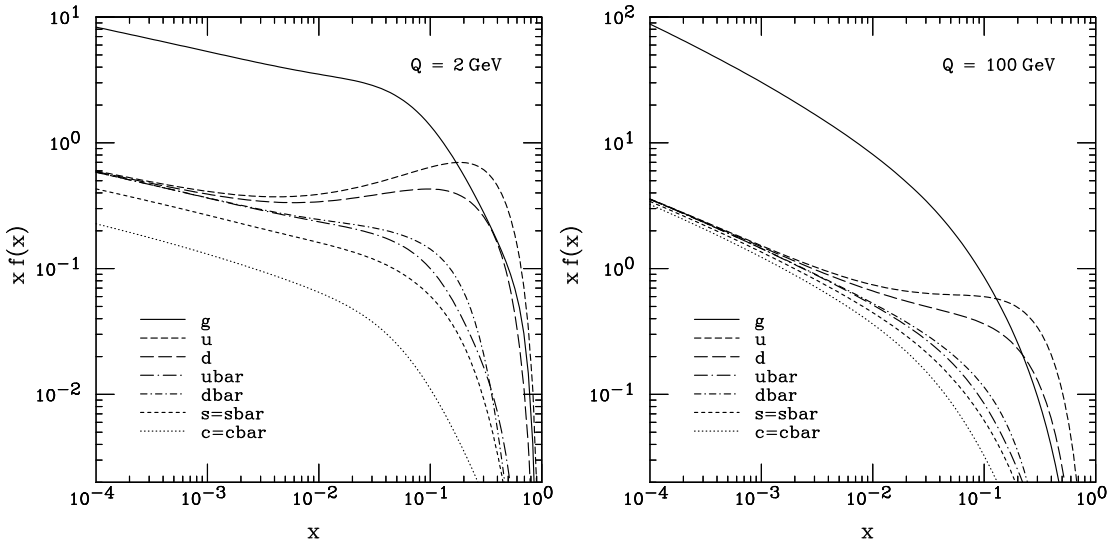

Figure 2.4: Example parton distribution functions of various partons from the CTEQ6M PDF set for Q^2 of 2 GeV (left) and 100 GeV (right) [34].

PDFs are parametrizations fitted to experimental data, which means that they are determined at certain discrete energy scales of high-energy physics experiments. They can, however, be evolved to other energies using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [35–37]. Alternatively PDFs can be obtained from lattice QCD [38] calculations, though these currently yield substantial uncertainties.

Most of the discussion above has so far concentrated on the physical processes that can be described through perturbative calculations and thus yield exact results up to corresponding energy scales. A full description of a proton-proton collision, however, must include further semi-perturbative and non-perturbative processes. These, as illustrated in Figure 2.5, include parton showering, hadronisation and the underlying event, which are discussed in the following sections.

2.2.4 Parton showering model

The number of calculations that need to be performed to provide exact predictions for the observables from a hard scattering processes increase approximately factorially with the order to which the prediction is being calculated [9]. Yet terms from orders higher than currently available NNLO[7] predictions can give non-negligible contributions to the observable final state. A parton showering model is used to account for these contributions by treating them as radiative emissions by the partons participating in the

[7]Equivalently to NLO in section 2.1.2, NNLO stands for next-to-next-to-leading order.

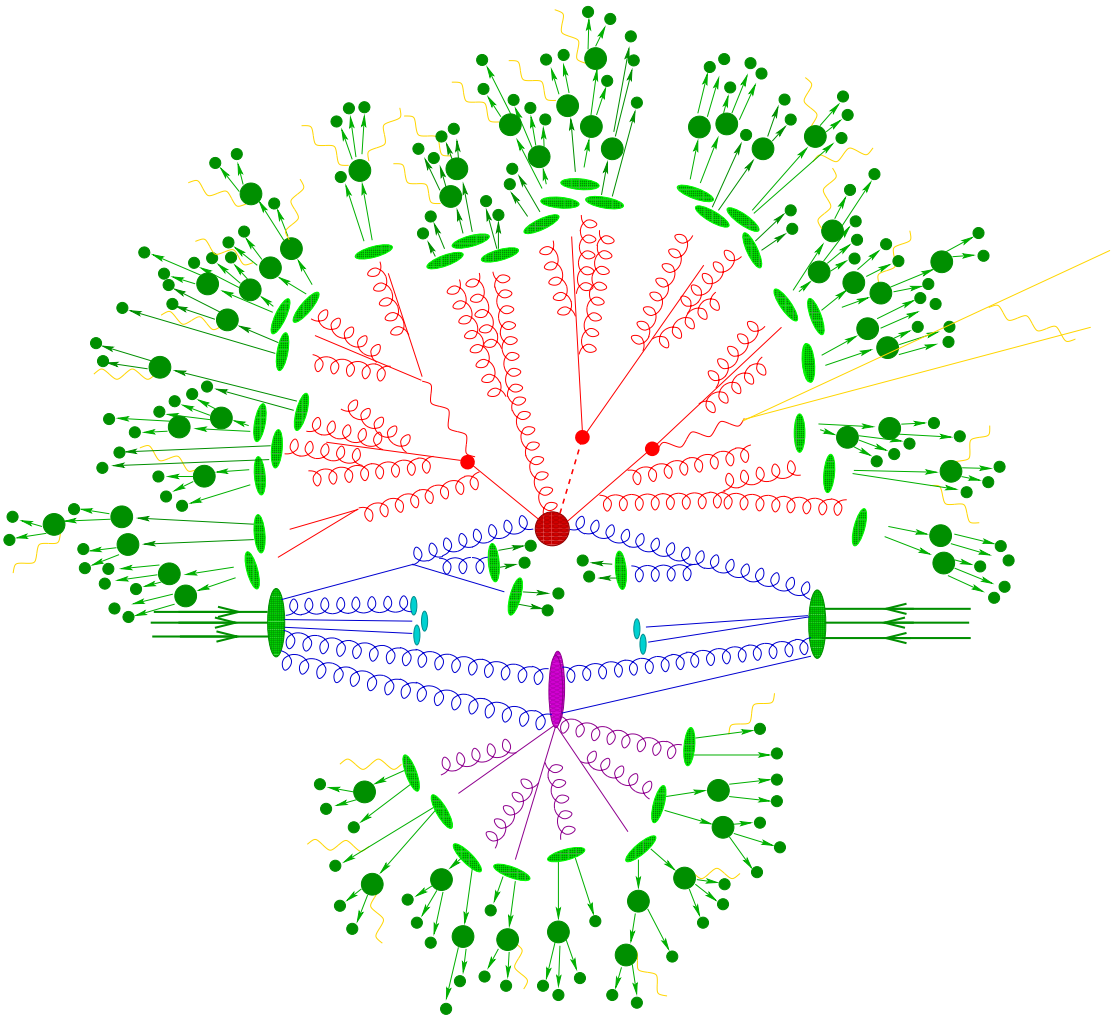

Figure 2.5: A depiction of a proton-proton collision process. The incoming particles are shown in green in the middle of the image. The dark blue lines leading to the large red spot show the partons of the incoming protons that take part in the collision, while the light blue spots represent spectator partons after proton disassociation. The large red spot shows the main hard scattering event with further red lines showing immediate collision product decay and parton showering. The lighter green spots show the hadronisation process and the dark green spots depict the hadron decay process. The purple interaction depicts a secondary interaction and in combination with the light blue disassociated partons form the underlying event [39].

hard scattering event before and after the event has taken place, known as initial-state (ISR) and final-state (FSR) gluon radiation.

ISR (FSR) is treated as a splitting of the incoming (outgoing) parton, a, into a new parton, b, and a less energetic version of itself, a', which carry a fraction of the initial momentum each - $p_b = zp_a$ and $p_{a'} = (1-z)p_a$. The splitting is governed by the Altarelli-Parisi splitting functions [37]. This splitting can proceed to an arbitrary soft or collinear limit, where the emission of an extra parton becomes unresolvable, ie. $z \to 0$, due to the momentum $p_b \to 0$ or the angle between the two $cos\theta_{ba'} \to 0$. This cut-off, known as the infra-red limit, is a cut on the momentum-transfer squared, t, often chosen to be $t = 1$ GeV^2 [9] for transverse momentum, p_T, ordered showers.

The evolution between different emissions of partons is accompanied by the evolution of Q^2 of the parton showering process which, in a way similar to DGLAP evolution of PDFs, is governed by the Sudakov form-factors [40]. The radiative contributions are ordered, in the evolution of the Q^2 of the parton splitting, from high values down to the chosen cut-off. This process is evolved forwards toward the hard scattering event in the case of ISR and backwards for FSR.

2.2.5 Hadronisation

The description of the hard scattering via the perturbative approach, dissociation of the colliding protons and the addition of the parton shower model, leaves the remnants of the event shown in Figure 2.5 in a state with an abundance of free coloured quarks and gluons. As mentioned in section 2.1.3, such a state cannot be observed in nature. Thus any true description of the full hadron collision event must combine these free coloured partons into colourless bound states. This is achieved by non-perturbative hadronisation models.

String fragmentation

One hadronisation model is the string fragmentation (SF) model [41], also known as the Lund[8] model. According to this model, after the parton showering is finished at t, the neighbouring asymptotically free quarks and antiquarks pair up forming two ends of a string. The string connecting them is a self-interacting colour field which has collapsed into a narrow tube [42]. As the partons move apart as shown in Figure 2.6 and lose their energy to the colour field these flux tubes are stretched until breaking via

[8] After Lund University, near Malmö, Sweden, where this model was first developed.

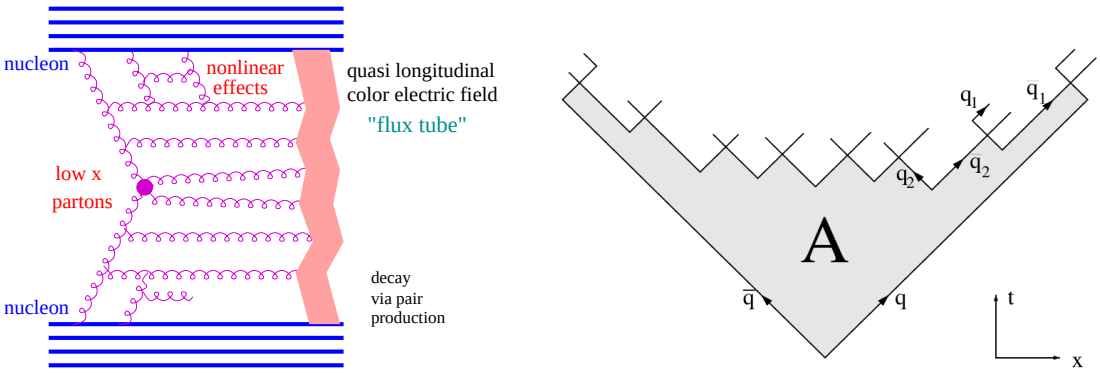

Figure 2.6: Left: a graphic depiction of a colour flux tube in the string fragmentation model with the gluon induced kinks evident [43]. Right: String fragmentation via $q\bar{q}$ pair production induced by the separation of the quark-antiquark pair at the ends of the string [44].

a spontaneous creation of $q\bar{q}$ pairs, eventually leading to the total energy of the colour field being incrementally transformed into colourless hadrons. The hard free gluons in this framework are introduced as kinks in the flux tubes with each kink proportional to the energy of the incoming gluon. This treatment has the benefit of producing a well modelled angular distribution of hadrons [9].

Cluster fragmentation

An alternative hadronisation description is provided by the cluster fragmentation (CF) model [45]. The CF model relies on colour pre-confinement [46]. This phenomenon implies that the hard process can be calculated until the appearance of pre-confined colour-singlet quark and gluon clusters of finite mass. This cluster mass, m_c is dependent on the parton shower cut-off scale, t, as $m_c \approx \mathcal{O}(\sqrt{t})$ [42]. The post-showering gluons split into $q\bar{q}$ pairs which then merge with the neighbouring quarks into colour singlet clusters. At $t = 1$ GeV^2 the values of m_c are typically around a few GeV/c^2 and the clusters can be treated as a superposition of meson resonances [9].

It is clear that either hadronisation model depends on the chosen value of the arbitrary cut-off scale, t, of the parton showering stage.

2.2.6 Underlying event

In addition to the main hard scattering event and the parton showering of ISR and FSR, there is a significant contribution from the underlying event (UE). The UE consists of the interactions between the beam remnants and the multi-parton interactions [47] (MPIs), the latter being the dominant contribution.

Multi-parton interactions

In an idealised circumstance, the hadron-hadron collision would contain a single hard scattering event involving a single parton from each colliding beam. Instead of the single scattering, a typical hadron collision involves multiple parton interactions - MPIs. These occur between other valence quarks in the hadron participating in the hard scattering event or gluons or quarks from the sea[9].

The perturbative calculations can, to a degree, account for MPIs with large transverse momentum, p_T, but contributions from MPIs also reach into the non-perturbative, soft regime. The term "soft" here simply implies that the momentum transfer in these additional interactions is not large enough for the pQCD approaches to be valid.

Furthermore, the higher the c.o.m. energy of a hadron collider, the greater the probability of MPIs occurring, as the constituents of the quark-gluon sea become more energetic. In addition, the soft regime enhancements from MPIs become increasingly prominent as the fractional momentum of one of the partons involved - given by the Björken-x variable - tends to zero [47].

One of the more developed theories describing contributions from MPIs at this time is Regge-Gribov theory, though no universally coherent description of both the hard and soft regime interactions currently exists. Approximative models, such as the dual-parton model, are currently used by the Monte-Carlo generators; the latter will be introduced in section 2.2.8.

MPIs are of significant motivational importance to the physics analysis described in this thesis and thus will be further discussed in chapter 5.

2.2.7 Regge-Gribov theory

The Regge-Gribov field theory [48–50] (referred to hereafter as "Regge theory") is a non-perturbative theory which describes the parton interactions at the soft limit (also known as the Regge limit). This theory was first introduced before the quark model was fully developed and therefore is a theory of hadrons.

Regge theory is based on treating the angular momentum in the scattering amplitude as a complex variable [42] connecting the behaviour of the scattering matrix to singularities in this complex space known as the Regge poles [51]. For a simple two

[9]Quarks that nominally form the hadron are known as the valence quarks. A proton, for example, is made up of the *uud* valence quarks. Any hadron, however, also contains a plethora of low-energy constituents known as the quark-gluon sea.

particle scattering,

$$p(p_1) + p(p_2) \rightarrow p(p_3) + p(p_4), \tag{2.57}$$

where the c.o.m. energy squared, s, and momentum transfer squared, t, are

$$s = (p_1 + p_2)^2 \quad \text{and} \quad t = (p_1 - p_3)^2, \tag{2.58}$$

the non-trivial scattering amplitude, T, can be expressed via the Regge theory as [52]

$$T(s,t) \propto s^{\alpha(t)}, \tag{2.59}$$

where $\alpha(t)$ denotes the position of a Regge pole in the complex space, also known as a Regge trajectory [52].

In the framework of Regge theory all hadrons can be described as Regge trajectories. The two leading trajectories are the Reggeon, $\alpha_{\mathbb{R}}$, and the Pomeron, $\alpha_{\mathbb{P}}$[10], with the latter being a subset of the former. A Reggeon depicts an exchange of a real hadronic state with definite quantum numbers (e.g. $\alpha_{\mathbb{R}} \in \rho$, ω, etc.) [42].

The Pomeron is an exchange of a colourless multi-gluon state with the quantum numbers of the vacuum. As shown in Figure 2.7, at the lowest order a $\alpha_{\mathbb{P}}$ consists of two QCD gluons.

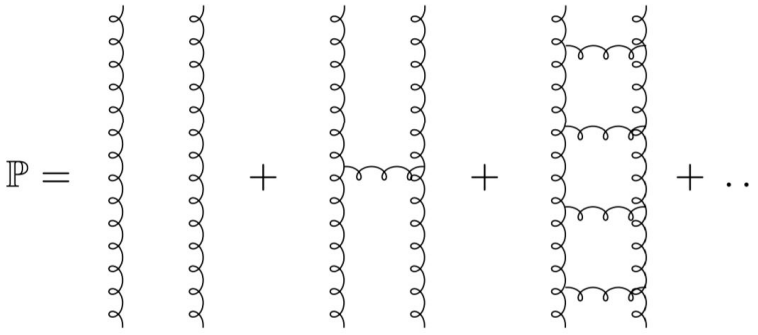

Figure 2.7: The soft Pomeron in terms of QCD gluons [52].

2.2.8 Monte-Carlo Simulation

To tackle the vast array of necessary calculations, various dedicated generators of theoretical predictions, known as Monte-Carlo (MC) generators, have been created. An overview of three such generators is given in the sections below.

[10] When referring to a Pomeron as a colourless state exchanged in an interaction, rather than a Regge trajectory, the symbol $\mathbb{P}$ is often used.

Pythia

PYTHIA [53, 54] is a general purpose MC generator used by the majority of high-energy collider experiments. The development of this tool began in 1978 and has continued ever since. Originally the code was implemented using `Fortran77` culminating in PYTHIA 6; modern computing approaches have prompted the redesign of the code to `C++` and the creation of the latest version - PYTHIA 8. Processes in PYTHIA are calculated to LO and the parton showering is implemented in either a $p_\perp$ or Q^2 ordered manner. The hadronisation process is modelled using the SF approach.

EPOS

EPOS [55] is an extensive air shower (EAS), also known as cosmic-ray (CR), generator. It was developed and is extensively used to describe the interactions between ultra-high-energy[11] cosmic rays and the atmosphere. The latest stable version of this generator is EPOS 1.99. It models the hard scattering as a two-part parton ladder joining the interacting hadrons with a hard and soft component, with the latter characterised using the Regge approach. Similarly to PYTHIA, EPOS uses the SF model to describe the hadronisation process. EPOS is unique in its characterisation of the collective flow, the collective behaviour of the dense medium in a hadron collision, during hadronisation [43], which is split into a corona and the core. Part of each hadronising string is treated as in other generators (corona like) and part of it hadronised in the core structure where collective effects are prominent. The core is distinguished from the corona via a critical string density parameter, ρ_0, above which the string behaviour is core-like. The core decay follows the methodology outlined in Ref. [57].

Sibyll

SIBYLL 2.1 [58], like EPOS, is a MC generator designed for modelling EAS. The interactions are based on the dual-parton model (DPM) with aspects of the Regge theory implemented for the description of the soft component. The hard component is simulated using the minijet model [59]. An in-depth discussion on the DPM and the minijet model can be found in Refs. [60] and [61] respectively. Like with the previous two generators, SIBYLL also utilizes the Lund string fragmentation model to describe the hadronisation process.

[11] Typically, when referring to ultra-high-energy cosmic rays, energy of $E > 10^{17}$ eV is implied [56].

All three generators are used for theoretical predictions for the physics analysis presented in this thesis. Additional details on the specific tunes and the PDF sets used will be elaborated upon in section 5.3.

CHAPTER

THREE

The LHC and the LHCb Experiment

"If the only tool you have is a hammer, you tend to see every problem as a nail"

- Abraham Maslow

In this chapter the experimental setup of the Large Hadron Collider and the Large Hadron Collider beauty experiment is discussed. In section 3.1 the accelerator complex of the LHC is described from the production of protons via the ionisation of hydrogen gas and the very first stage of acceleration to the delivery of the particle beams to the experiments. In section 3.2 a detailed depiction of the LHCb experiment and its constituent sub-detectors is provided. Finally, in section 3.3, the performance of both the accelerator and the experiment along with the highlights of the physics output to date is briefly discussed.

3.1 The LHC

The LHC [62] is the world's largest and highest-energy particle accelerator located at the European Organisation for Nuclear Research, CERN, near Geneva, Switzerland. It is housed in a 27 kilometre long, near-circular tunnel under the Franco-Swiss border originally constructed for the use of the Large Electron Positron (LEP) collider. An overview of the accelerator is shown in Figure 3.1. The LHC is a two-ring superconducting synchrotron of two counter-rotating beams designed to accelerate tightly packed bunches of protons up to a design energy of 7 TeV per beam.

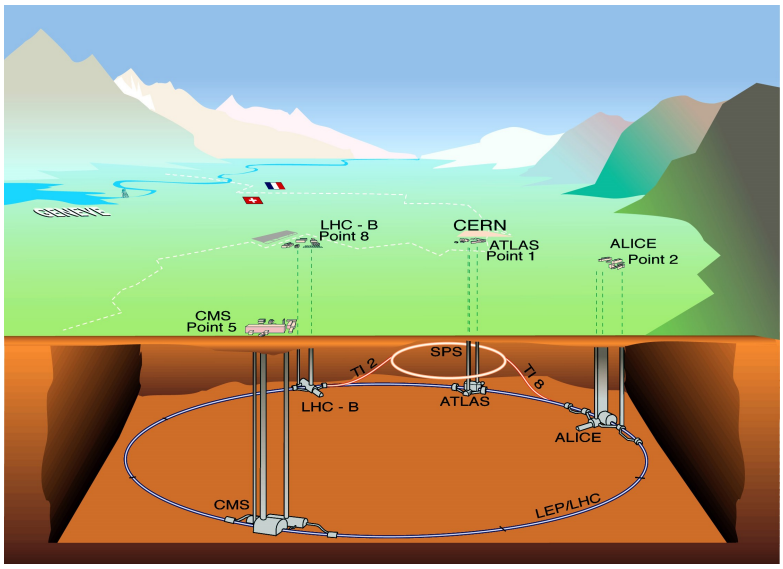

Figure 3.1: An overview of the LHC and the four main experiments [63]. It lies 45 to 170 metres underground [62] between Lac Léman (left) and the Jura mountains (right).

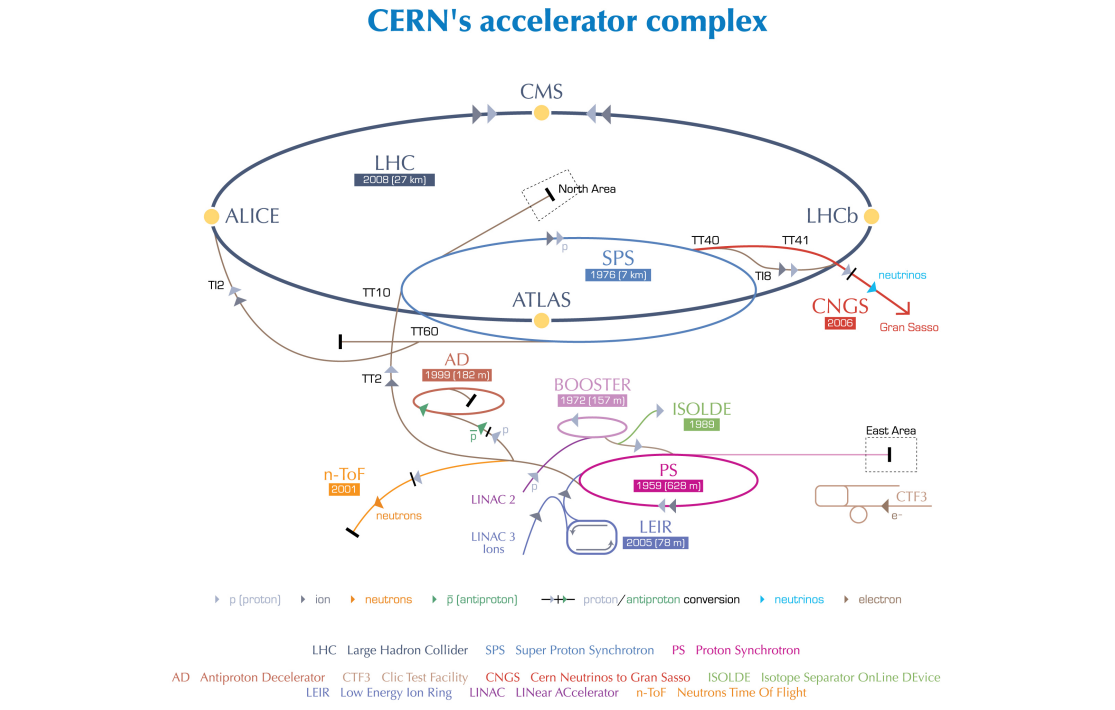

Figure 3.2: The accelerator complex at CERN. Among other accelerator facilities, the whole LHC injection chain, from LINAC 2 to the main ring, can be seen [64].

3.1.1 The accelerator and collider complex

The LHC ring alone cannot be used to accelerate protons from rest to the LHC design energy. A multi-stage accelerator complex, shown in Figure 3.2, is used for this process. The delivery of the high-energy proton beams starts with bottles of hydrogen gas. The gas is ionised and the protons are fed into the first stage of the acceleration process - LINAC 2[1] - where they are accelerated to an energy of 50 MeV [65]. From there the protons are accelerated by the Proton Synchrotron Booster (PSB) and Proton Synchrotron (PS) to an energy of 25 GeV. During this process the proton bunches are structured into bunch trains. This bunch train structure allows for a faster and more stable filling procedure with respect to the injection of individual bunches. From the PS the proton bunches are injected into the Super Proton Synchrotron (SPS), where they are accelerated to 450 GeV and transferred to the LHC [62].

The acceleration of protons throughout all of these stages is performed using radio-frequency (RF) cavities. In these cavities an electromagnetic wave is introduced in phase with the particles traversing them. The electromagnetic field transfers energy to particles every time they pass through such a cavity. In a linear accelerator the RF cavities are placed along the waveguide with increasing spacing between them such as to keep the electromagnetic oscillations in phase with the accelerating particles. This means that the accelerating power of a LINAC is predominantly limited by its length. In order to accelerate particles to higher energies, synchrotrons - circular accelerators - are used. In a synchrotron the particles are guided in a closed, near-circular path by powerful dipole magnets and traverse the same RF cavities multiple times gaining energy with each revolution. The oscillation frequency of the electromagnetic wave in the RF cavities is adjusted after each revolution to keep it in phase with the particles being accelerated. The energy that can be reached with a synchrotron is limited by the emission of the synchrotron radiation, which is emitted as charged particles are accelerated around a circular path. The energy loss per revolution, ΔE, due to synchrotron radiation is given by

$$\Delta E = \frac{4\pi q^2 \beta^2 \gamma^4}{3\varepsilon_0 r}, \tag{3.1}$$

where q is the charge of the particle, r is the bending radius, $\beta = v_p/c$, v_p being the velocity of the particle and c being the speed of light, $\gamma = 1/\sqrt{1-\beta^2}$ and ε_0 is the permittivity of free space [66]. In the relativistic approximation $\beta \approx 1$ and

[1] LINAC stands for Linear Accelerator. There are four LINACs in the CERN accelerator complex.

$\gamma \approx E/m$. Therefore the limiting factors for the maximum energy attainable by a synchrotron accelerator are the mass of the particles to be accelerated and the radius of the machine itself. The first factor can be suppressed by using massive particles, such as protons, whereas the second factor translates into a need for powerful bending magnets, building of ever larger accelerating machines or the combination of both. The LHC ring consists of eight straight sections containing the four experimental caverns, RF cavities and collimation and beam dump facilities. The latter two are used for narrowing of the particle beams and for the discarding of degraded beams once their use has expired, respectively. Furthermore, there are eight arced sections containing the dipole magnets [62]. There are 1232 dipole magnets delivering a magnetic field of up to 8.33 T. In addition there are quadrupole and higher order magnets for focusing the particle beams. To deliver the required magnetic field the LHC magnets are liquid-helium cooled to 1.9 K and superconducting.

The LHC beam consists of tightly packed particle bunches containing $\mathcal{O}(10^{11})$ protons each. The number of bunches injected into the LHC and the spacing between them depends on the mode of operation of the machine. During nominal running, the LHC can accommodate up to 39 bunch trains each containing 72 bunches separated by 25 ns, leading to a maximum of 2808 proton bunches circulating the machine. The most common filling schemes are described in Ref. [67].

In addition to the proton beams, the LHC and the preceding accelerator complex can be used to accelerate and collide lead ions. In this case the ions start from a source of vaporised lead and are initially accelerated using the LINAC 3 and the Low Energy Ion Ring (LEIR) before progressing to the PS.

The two LHC beams are made to collide at four locations around the main ring within the caverns of the main LHC experiments - ALICE, ATLAS, CMS and LHCb. ATLAS [68] (A Toroidal LHC ApparatuS) and CMS [69] (Compact Muon Solenoid) experiments are general purpose detectors (GPDs). These have a near 4π coverage and have been designed for the search for new physics such as supersymmetry. ALICE [70] (A Large Ion Collider Experiment) focuses on heavy-ion physics topics such as the determination of the properties of quark-gluon plasma - nuclear matter at extreme values of energy density. LHCb [71] has been designed to perform heavy-flavour physics measurements, yet it can be used to explore a vastly diverse high-energy physics program. LHCb will be described in further detail in section 3.2.

3.2 The LHCb experiment

The LHCb experiment is a single-arm forward spectrometer located in an underground cavern at interaction point 8 (IP8) of the LHC. It is fully instrumented in the forward region covering the angular acceptance of approximately 10 mrad to 300 (250) mrad in the bending[2] (non-bending) plane [71].

The LHCb collaboration predominantly focuses on the study of heavy flavour physics and $\mathcal{CP}$ violation in the decays of particles containing beauty and charm quarks. At the LHC these heavy particles are predominantly produced in cones along the beam axis, which has influenced the unique design choices of the LHCb experiment [71]. A side view of the entire detector is shown in Figure 3.3.

3.2.1 Beampipe

The beampipe, shown in Figure 3.4, traverses the entire LHCb detector with the exception of the VELO (**Ve**rtex **Lo**cator) sub-detector, which surrounds the interaction point. It is 19 metres long and consists of four consecutive conical sections. The vacuum in the LHCb beampipe is isolated from the LHC beam vacuum with two sector valves situated at the cavern entrances. This allows for interventions at the detector independently of the LHC vacuum system [71].

The high pseudorapidity[3] area of interest of the LHCb detector means that particles of interest will often traverse the beampipe. Thus the minimization of the material budget, i.e. the transparency of the beampipe to the traversing particles, is of great importance. The first three sections, corresponding to the first 12 metres of the beampipe, are made of beryllium. This material was chosen due to its transparency to high-energy particles. However, it is also costly, fragile and highly toxic [72]. Therefore the last 7 metres of the beampipe, where the material budget is of lesser importance, are constructed of stainless steel. Other elements of the beampipe, such as the VELO exit window, flanges and bellows, are made of aluminium alloys [71].

[2]Throughout this thesis, when using the 3-dimensional coordinates in reference to the experimental setup, z denotes the direction along the beampipe and x and y denote the horizontal and vertical axes transverse to the direction of the LHC beams, respectively. Bending and non-bending planes are thus defined by $(x \times z)$ and $(y \times z)$ axes, respectively.

[3]Pseudorapidity, η, is defined as $\eta = -ln(tan(\theta/2))$, where θ is the polar angle with respect to the beampipe.

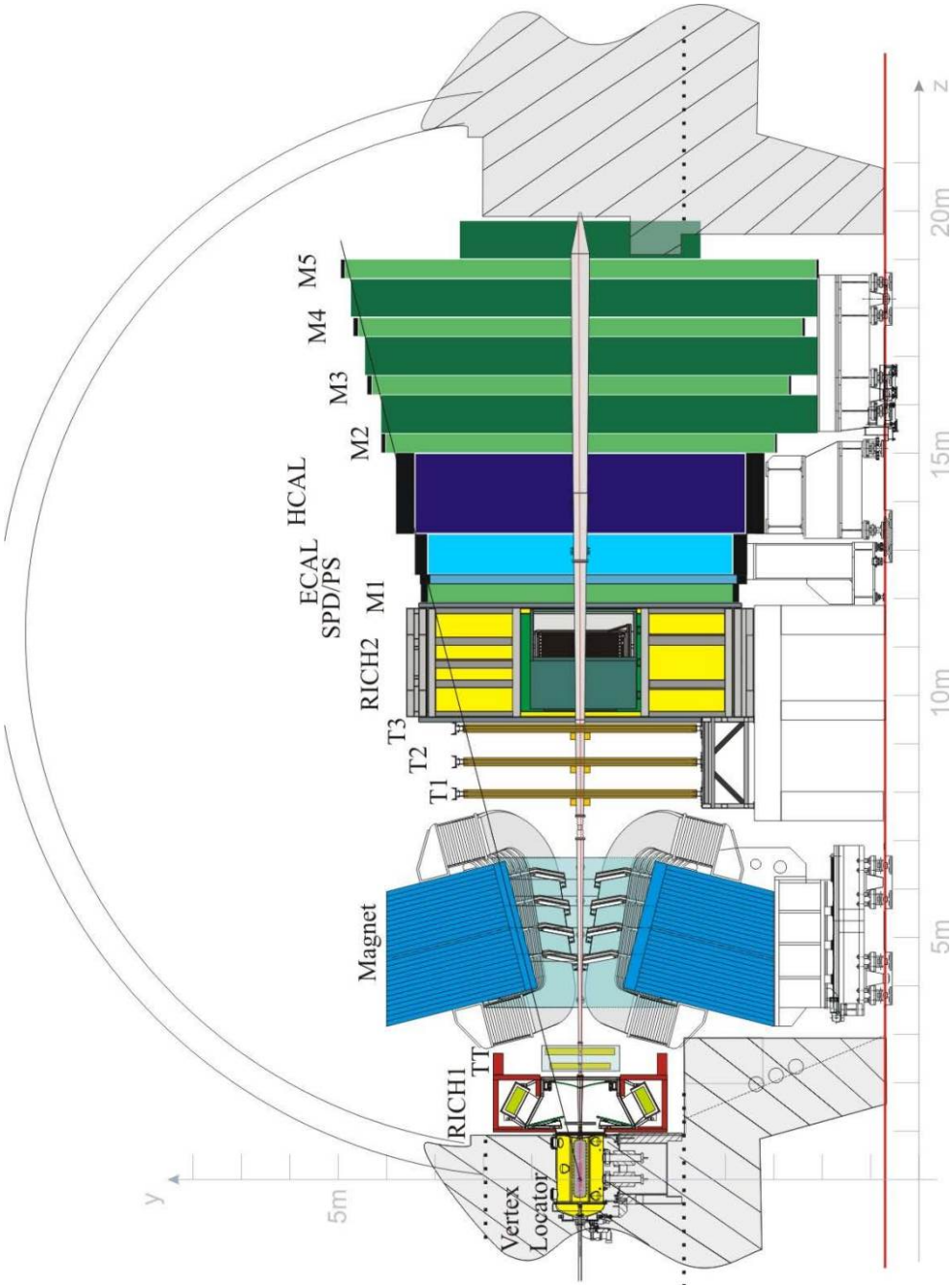

Figure 3.3: The side view of the LHCb detector [71]. The interaction point is located inside the Vertex Locator on the left of the image. In the text the assignment of 'downstream' refers to components further to the right and 'upstream' to components further to the left of any other part of the detector; *e.g.* RICH1 is located upstream and RICH2 downstream of the magnet.

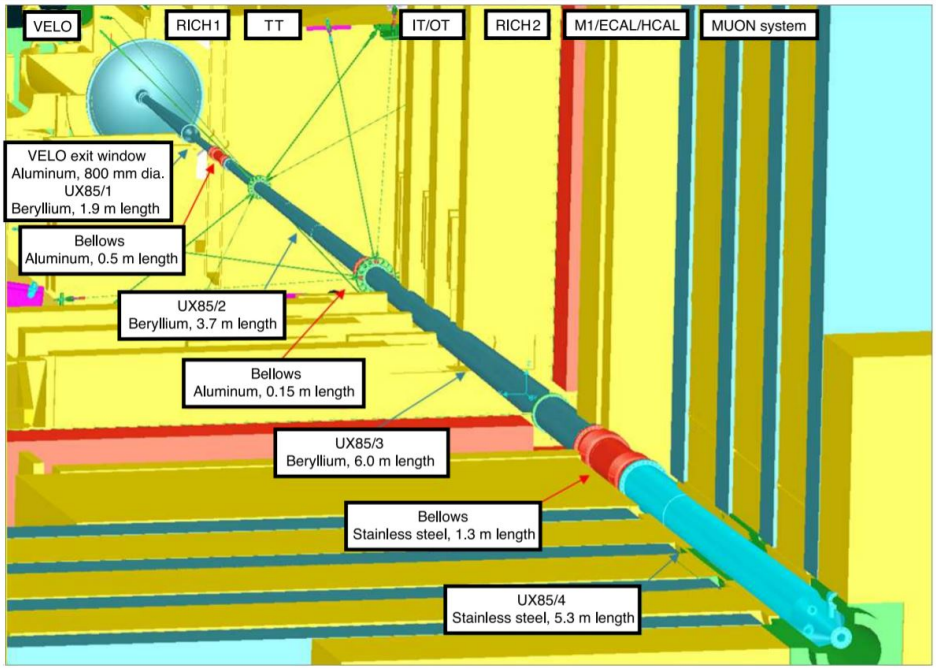

Figure 3.4: The beampipe traversing the LHCb experiment [71].

3.2.2 Vertex Locator

The VELO surrounds the interaction point at the LHCb experiment. It is a silicon-strip detector used to measure charged track coordinates near the interaction region with high precision. This enables the VELO to be used to identify, resolve and accurately determine the position of primary and displaced secondary vertices of particle decays. The latter are the decay vertices of prompt particles arising from the interaction point and are a distinctive feature of decays of heavy particles containing b and c quarks [73].

The VELO consists of two sides separated in the vertical plane, each containing 21 silicon modules. To assist with the alignment and to ensure the full coverage of the azimuthal acceptance the two halves are required to overlap. To facilitate this, modules on one side are shifted by 1.5 cm in the z direction with respect to the other side. The spacing between the VELO stations in the z direction is determined by the requirement that all tracks that originate from the interaction region and fall inside the LHCb acceptance should traverse at least three VELO stations. There are two further VELO stations upstream of the interaction region on each side of the detector. These make up the pile-up[4] veto system [71]. A schematic view of the layout of the VELO is shown in Figure 3.5.

Each VELO module is made of two back-to-back silicon-strip sensors. These provide

[4]Pile-up is the occurrence of more than one pp collision in a single beam interaction.

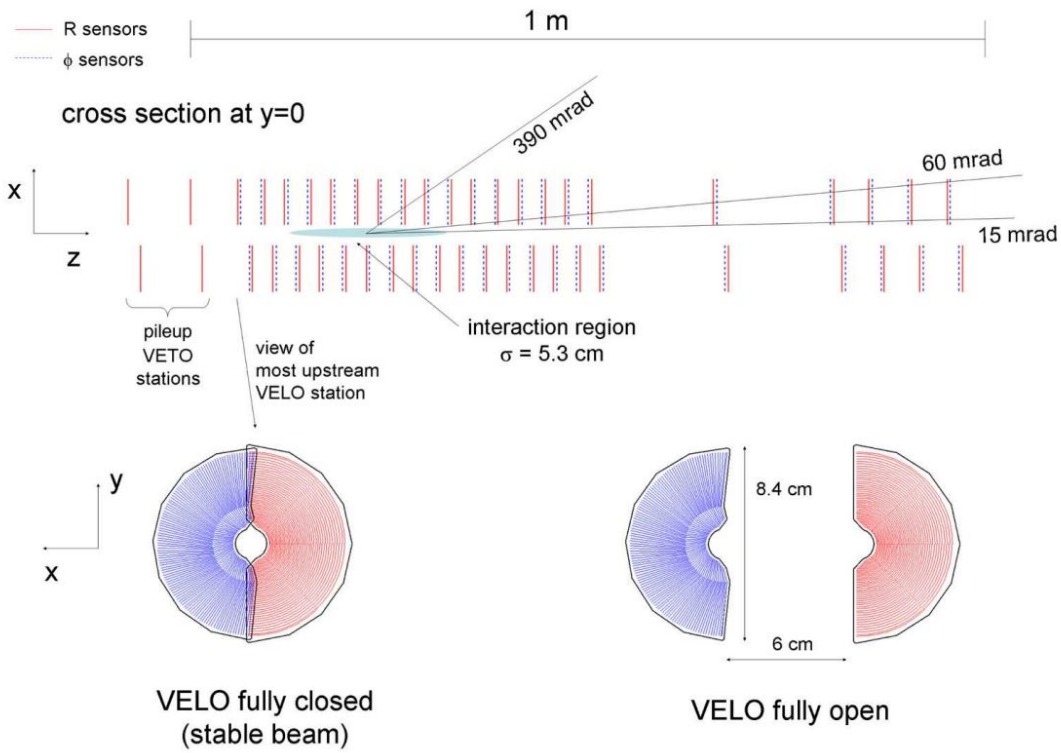

Figure 3.5: Top down view of the cross-section of the VELO silicon sensors in the fully closed position at $y = 0$ (top) and the close-up of the VELO sensors from the LHC beam view in fully closed (left) and fully open (right) positions [71].

the r (R-sensor) and ϕ (ϕ-sensor) coordinates of the tracks traversing them, where r is the radial position and ϕ the azimuthal angle centered around the beam axis.

The R-sensors are a set of concentric semi-circular silicon strips. To minimize the occupancy the module is separated into four 45° sectors. The strip pitch of the R-sensor ranges from 38 μm in the innermost region to 101.6 μm at the outer edge of the active area.

The ϕ-sensor consists of silicon strips arranged orthogonally to those of the R-sensor. It is divided into two regions as allowing the radial increase of the strip pitch across the whole of the module would result in an unacceptably broad pitch at the outer edge. In the inner region the strip pitch ranges from 38 μm at the inner edge of the active area to 78 μm at the radius of 17.25 mm. From there the strip pitch is halved to 39 μm and radially increases to 97 μm at the outer edge of the sensor [71]. The schematic layout of the two different VELO sensors is shown in Figure 3.6.

To shield the sensors from the RF pickup effects and to separate the VELO from the LHC vacuum, an aluminium cover, referred to as the RF-box, covers each half of the detector. To minimize the material budget and to allow for the overlap of the two

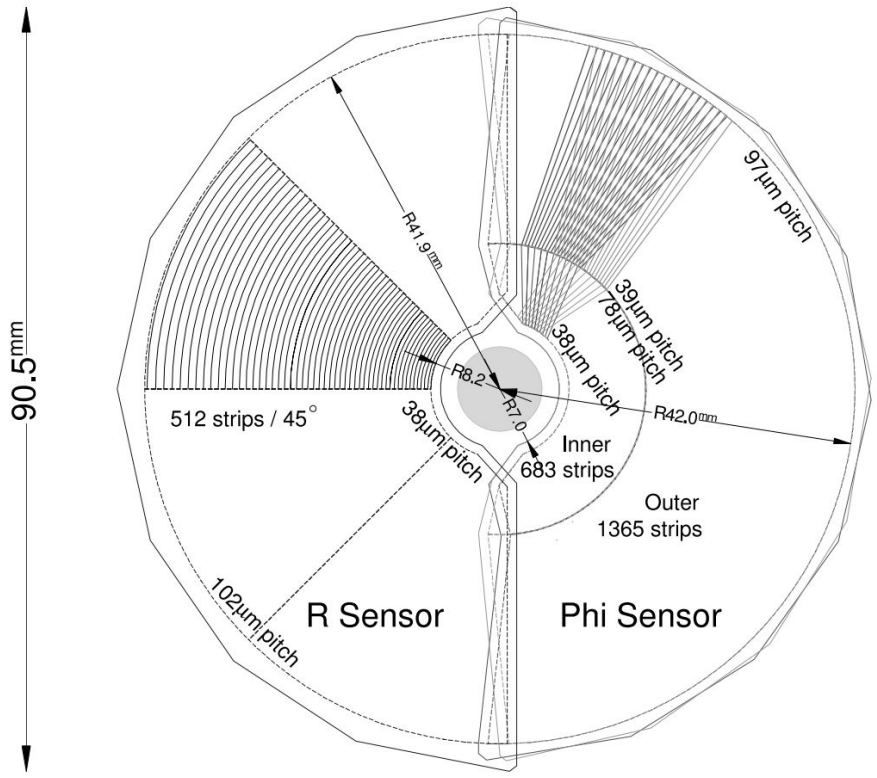

Figure 3.6: Schematic of the two VELO sensors in the closed position showing the r-ϕ geometry. Due to different bonding, the R-side sensors are slightly larger, however the active areas are of identical dimensions [71].

VELO halves, the sensor face of this box is a precision-engineered corrugated RF-foil. The LHC vacuum is required to be better than 10^{-8} mbar. The secondary vacuum inside the RF-box, due to outgassing of the detector components, is around 10^{-4} mbar. A safety system is present, which does not allow the pressure differential between the two to exceed 5 mbar in detector vacuum overpressure and 2 mbar in beam vacuum overpressure. This means that the difference between the pressure inside the RF-box and that outside the RF-box is not allowed to exceed negative 5 mbar and positive 2 mbar. The maximum pressure difference tolerance of the RF-box is 20 mbar before deformation and 50 mbar before failure. Figure 3.7 shows the module arrangement, support and the RF-box for one half of the detector [71].

The focus of the LHCb experiment on precision heavy-flavour physics measurements requires excellent vertexing capabilities close to the interaction region. Hadrons containing b quarks produced at the LHC energies will typically traverse a distance of around 1 cm before decaying. Optimally, the vertexing detector should be placed no farther from the beam axis than this distance. This is challenging as during beam injection and ramping phases of the LHC cycle all detectors are required to be no closer to the

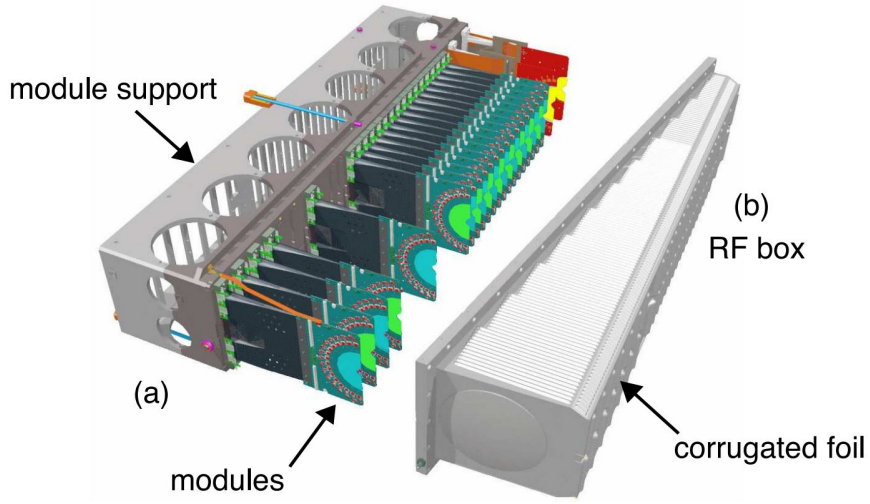

Figure 3.7: View of the VELO modules and their support structure (a) and the RF-box with the corrugated RF-foil (b) [71].

beam axis than 30 mm to allow for beam excursions. Therefore the VELO was designed to be made of two retractable halves which can be retracted to ±3 cm away from the beam axis, as shown in Figure 3.5, until stable beam conditions have been declared by the LHC. When fully closed, the inner edge of the active area of the VELO detector sits at the radius of 8.2 mm from the beam axis as shown in Figure 3.6 and is the closest-to-beam detector at the LHC. It should be pointed out that, to avoid damaging the detector, no beams can be injected in the LHC before LHCb has declared that the VELO is fully retracted [71].

The proximity of the VELO sensors to the LHC beam means that during operation these are exposed to an extremely high and non-uniform irradiation. For an annual accumulated luminosity of 2 fb^{-1}, the 1 MeV neutron equivalent flux at the inner region would reach 1.3×10^{14} n_{eq}/cm^2 while at the outer edge it would not exceed 5×10^{12} n_{eq}/cm^2. To keep the radiation damage as low as possible the irradiated VELO modules should be kept at a temperature below -5°C at all times. To achieve this the VELO modules have an integrated two-phase CO_2 cooling capable of dissipating 24 W per sensor. Although the VELO was designed to withstand these irradiation levels for a minimum of 3 years of nominal LHCb operation, a fully operational copy of the detector - VELO2 - was constructed and delivered as backup system [71].

3.2.3 Tracking system

Besides the VELO detector, the tracking system at the LHCb experiment consists of the Silicon Tracker (ST), the Outer Tracker (OT) and a dipole magnet. The former is

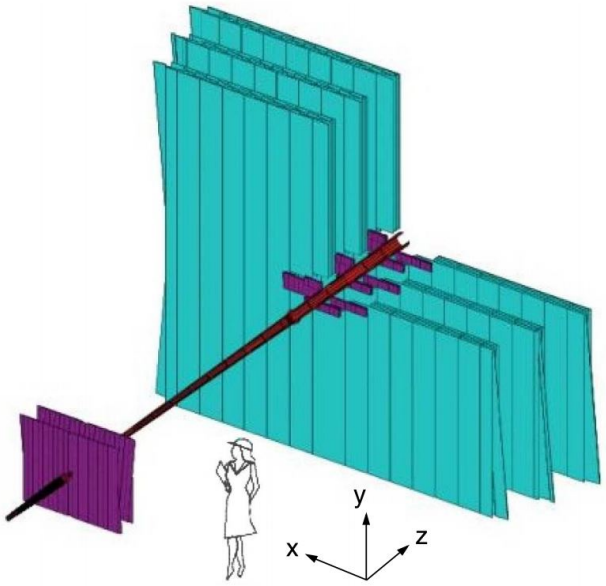

Figure 3.8: Layout of the LHCb tracking detectors. TT (purple) is located on the near side and the IT (purple) and OT (turquoise) are located on the far side of the image [71] with the beampipe traversing the centre of the system.

further split into Tracker Turicensis[5] (TT) and the Inner Tracker (IT). An overview of the tracking stations is shown in Figure 3.8 [71].

Magnet

In order to determine charged particle momenta, high-energy physics experiments use strong magnetic fields to bend their paths. The momentum information can be extracted from the radius of the curvature of the tracks in the presence of magnetic field.

The LHCb experiment employs a warm[6] saddle-shaped dipole magnet. The two magnetic coils, each weighting 54 tonnes, are made of pure Al-99.7 hollow conductor mounted on a 1500 tonne iron yoke. The schematic depiction of the LHCb magnet assembly is shown in Figure 3.9. The magnet produces an integrated field of 4 Tm in the $\pm$y-direction[7]. The specific shape of the magnetic field is shown in Figure 3.10 (b). The choice of the field shape was influenced by two conflicting requirements - the maximum attainable magnetic field gradient between the VELO and other tracking stations and

[5]The name *Tracker Turicensis* was reverse-engineered from its abbreviation - TT. These stations were to be included into the LHCb trigger and called the *Trigger Tracker*. After these plans were changed, a new name - *Tracker Turicensis* was established, where *Turicensis* stands for a Latin name for the city of Zürich, where this detector was built.

[6]A superconducting magnet design was rejected due to unacceptably high costs and long construction time.

[7]The polarity of the magnetic field can be reversed. This is exceptionally useful for various asymmetry analyses at LHCb.

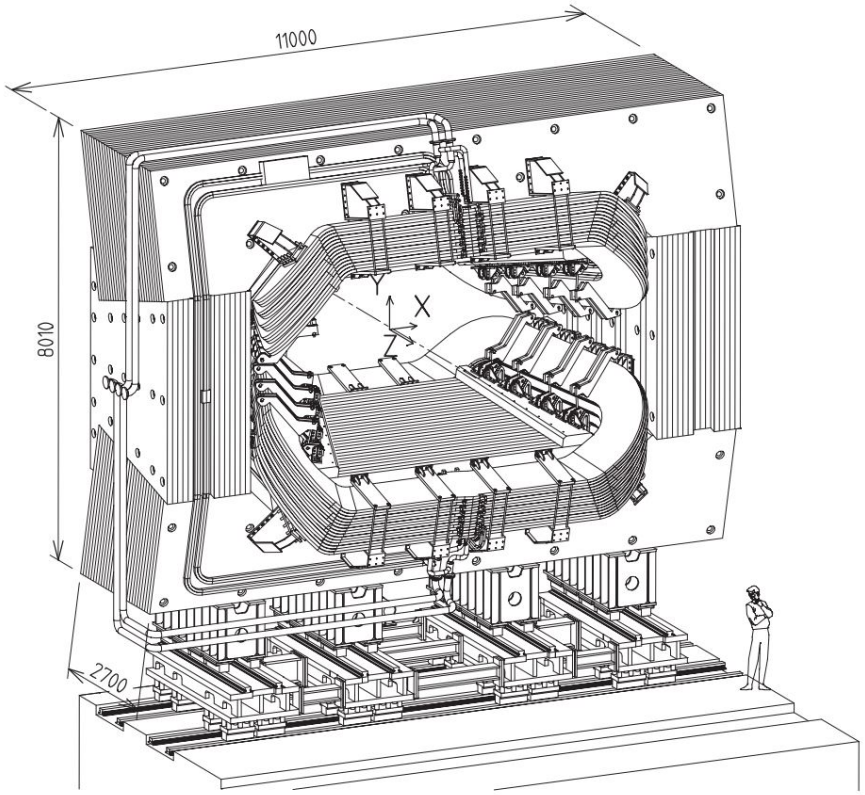

Figure 3.9: The LHCb dipole magnet (units in millimetres) [71].

the minimal field strength inside the VELO and the RICH detectors [71].

The magnetic field measurements were performed using Hall probes mounted on a sensor grid. The spatial precision of the measurements was of the order of a few millimetres and the magnetic field was measured at the relative precision of 4×10^{-4}. Figure 3.10 (a) shows the distribution of the relative field strength uncertainty for the Hall probe measurements, while Figure 3.10 (b) shows the measured magnetic field compared to calculations obtained with TOSCA[8]. In all regions of the detector the agreement between the two is better than 1% [71].

Silicon trackers

TT comprises 4 planar detection layers located upstream of the dipole magnet at a distance of approximately 2.33 m to 2.64 m from the nominal interaction point. These are arranged in an *x*-*u*-*v*-*x* geometry, where the different designations refer to the strip orientation in the different stations, i.e. the four stations have a strip orientation of 0°, +5°, -5° and 0° with respect to the vertical axis. TT is 160 cm by 130 cm in size and covers the entire LHCb acceptance. It is housed in a thermally and electrically insulated

[8] An analysis package for calculation of vector fields.

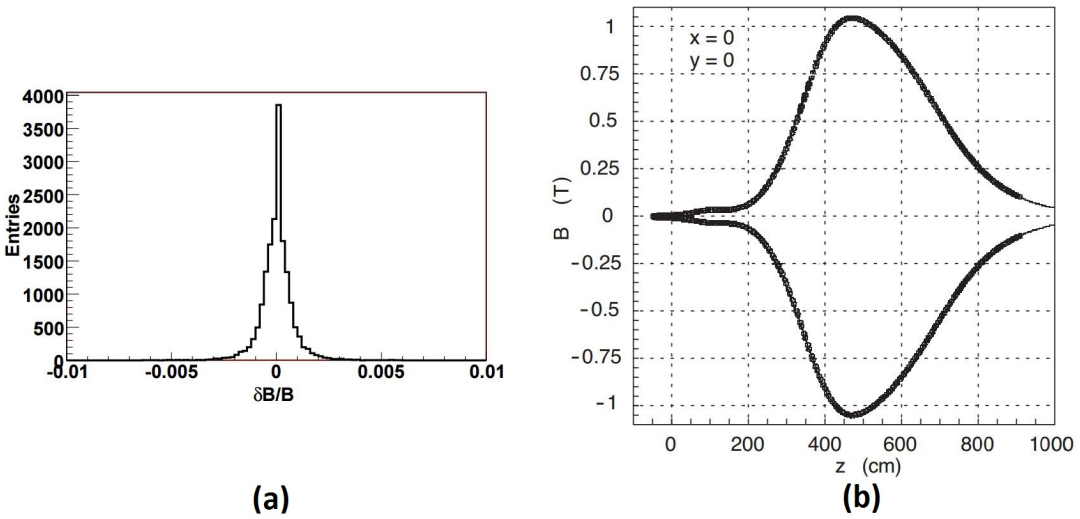

Figure 3.10: The relative uncertainty on the magnetic field measurement using different Hall probes (a). The measured (points) and calculated (line) absolute magnetic field throughout the LHCb detector for the two magnet polarities (b) for the main magnetic field component B_y [71].

and light-tight box at an ambient temperature below 5°C [74].

A TT layer is made from a series of half modules as shown in Figure 3.11. Each module contains 7 or 8 silicon sensors. The modules are staggered by about 1 cm in the z direction and overlap by a few millimetres to ensure full coverage. The sensors used in TT are 500 μm thick, single-sided silicon strips. Each sensor is 9.64 cm wide and 9.44 cm long and contains 512 strips with a strip pitch of 183 μm [71].

The IT forms the innermost region of the three tracking stations downstream of the dipole magnet, T1 - T3, known as the T stations. Similarly to TT, the IT modules are surrounded by a light-tight and thermally and electrically insulated box providing an ambient temperature of 5°C. The three IT detection layers are cross-shaped and cover an area of approximately 126 cm by 41 cm. A single layer of the IT is shown in Figure 3.12.

The sensors of the IT are 7.6 cm wide and 11 cm long and are made out of 384 silicon strips with a strip pitch of 198 μm. Depending on position within a layer, the sensor modules are between 320 μm and 410 μm thick [71].

Outer tracker

The OT is a drift time tracking detector surrounding the IT in the T stations. A drift time detector utilizes the ionization of a gas by a charged particle. The released electrons are attracted by an anode wire and the drift time of these electrons can be

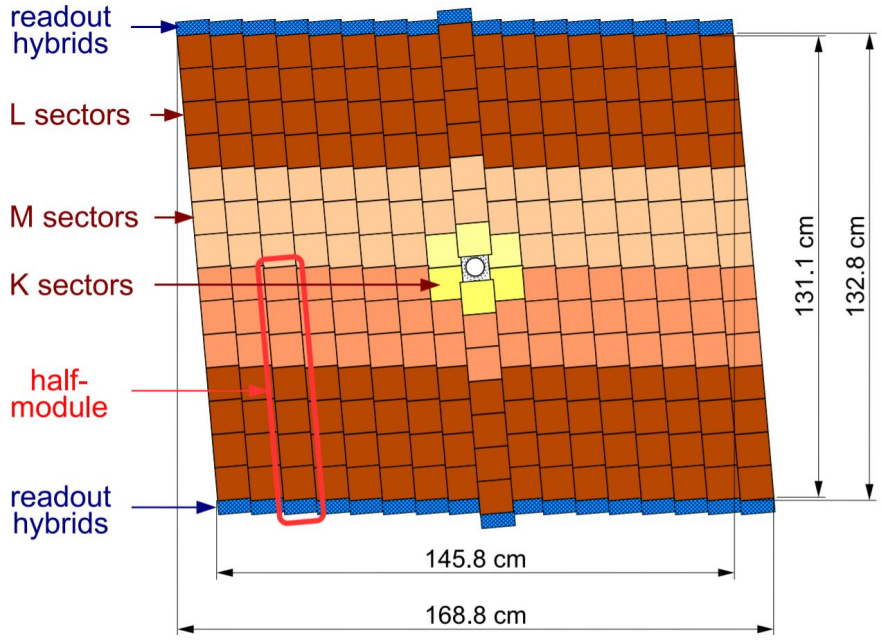

Figure 3.11: Schematic of TT (shown on the near side in purple in Figure 3.8). The third detection layer (v-layer) shown [75].

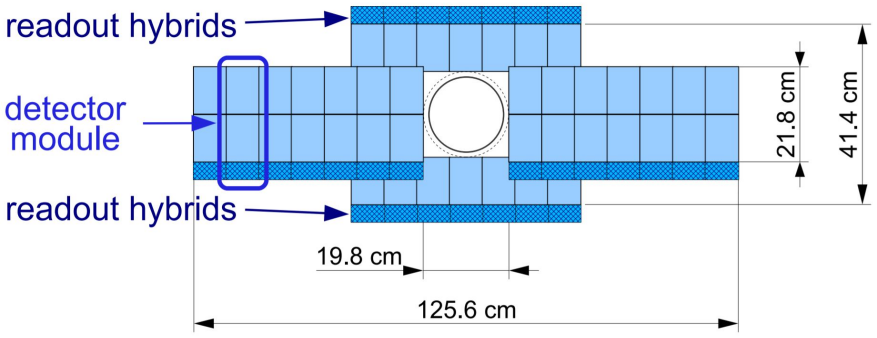

Figure 3.12: Schematic of the IT (shown on the far side in purple in Figure 3.8). The x detection layer shown [76].

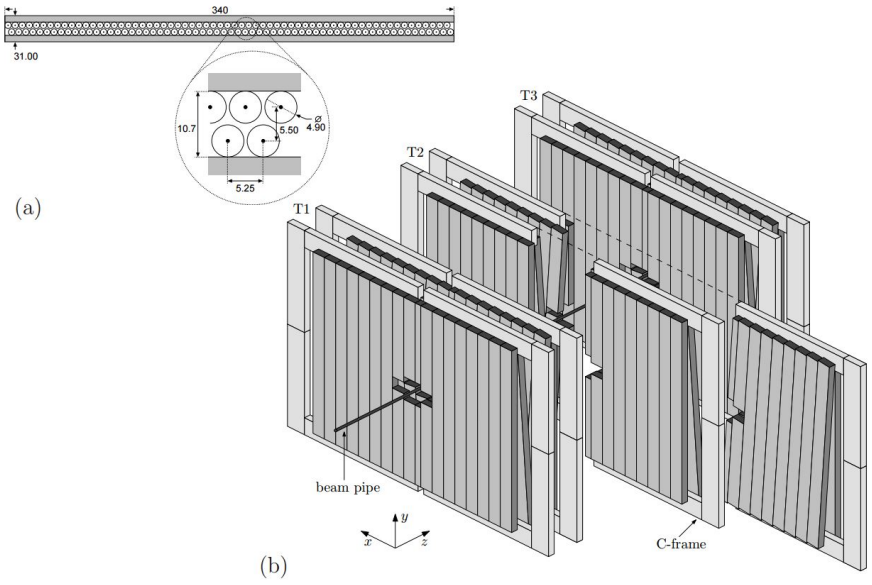

Figure 3.13: Schematic of the OT modules (shown in blue in Figure 3.8). Model cross-section (a) and the arrangement of straw-tube modules in the OT stations (b) shown [77].

used to extract positional information of the track hit. An OT module consists of two staggered layers of drift-tubes with the inner diameter of 4.9 mm. The gas mixture of Argon (70%) and CO_2 (30%) provides a drift time resolution below 50 ns and the drift-coordinate resolution of 200 μm. The arrangement of the straw tubes in an OT module is shown in Figure 3.13 (a).

There are four layers of modules per station. Like in the case of TT, these are arranged in an *x*-*u*-*v*-*x* geometry, where the outer two modules are vertical and the inner two are tilted $\pm 5^\circ$ with respect to the vertical. The active area of these tracking stations is approximately 591 cm by 485 cm and the combination of the IT and OT covers the full LHCb acceptance. The OT stations are shown in Figure 3.13 (b) [71].

3.2.4 Ring imaging Čerenkov detectors

The RICH (Ring imaging Čerenkov[9]) detectors are used as particle identification (PID) tools in high-energy physics experiments. These detectors utilise the Čerenkov effect - a charged particle travelling through a medium faster than light in the same medium will emit a cone of light around its direction of travel. The conical angle of the emitted radiation, θ_c, can be calculated using [66]

$$\cos\theta_c = \frac{1}{\beta n}, \tag{3.2}$$

[9]Letters 'CH' in the acronym are used due to a common English spelling 'Ch' for the letter Č.

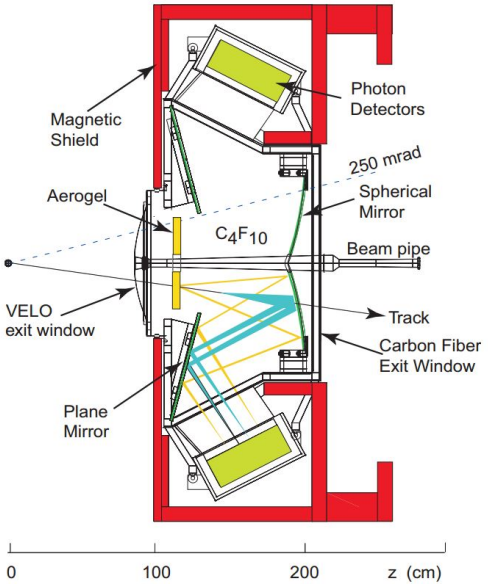

Figure 3.14: Side view of the RICH1 sub-detector [71].

where β is a standard relativistic variable as defined in section 3.1.1, and n is the refractive index of the medium. The velocity obtained for a given particle can be combined with the momentum information of the corresponding track given by the tracking system in order to measure the mass and subsequently identify the species of the particle. PID of charged hadrons, particularly the ability to distinguish between kaons and pions with high efficiency, is of exceptional importance for the main physics program of LHCb. In order to provide PID across a range of momenta in the high rapidity area of interest of LHCb, where the momentum spectrum is relatively hard, the experiment has two of these detectors - RICH1 and RICH2 [71].

RICH1

The RICH1 detector is located between the VELO and TT upstream of the dipole magnet. It uses silicon aerogel and fluorobutane (C_4F_{10}) gas as the radiative media to provide PID in the momentum range of $1 < p < 60$ GeV/c. RICH1 covers the full[10] angular acceptance of the experiment of 25 mrad to 300 (250) mrad in the bending (non-bending) plane. The Čerenkov light is focused onto Hybrid Photon Detectors (HPDs) by an arrangement of spherical and flat mirrors. The setup of the optical system of the RICH detectors was designed to minimise the material budget in the LHCb acceptance. The total radiation length, X_0, of RICH1 is $\sim 8\%$ X_0. The schematic of the RICH1 detector is shown in Figure 3.14 [71].

[10]The low angle acceptance is slightly reduced from 15 mrad to 25 mrad by the beampipe [71].

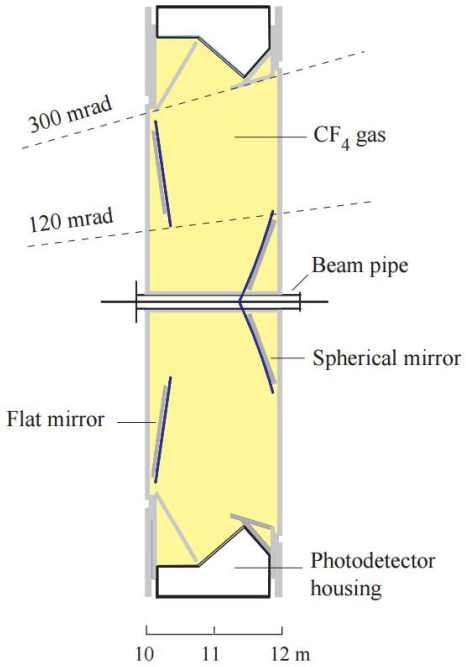

Figure 3.15: Top down view of the RICH2 sub-detector [78].

RICH2

The RICH2 detector is located between the tracking stations and the first muon station downstream of the dipole magnet. The radiative medium used in RICH2 is tetrafluoromethane (CF_4) gas, which provides PID across a range of $15 < p < 100$ GeV/c. RICH2 covers a reduced angular acceptance of 15 mrad to 120 (100) mrad in the bending (non-bending) plane. Similar to RICH1 the Čerenkov light is focused onto HPDs using an arrangement of spherical and flat mirrors minimizing the material budget to about 0.15 X_0. Unlike RICH1, however, the optical systems are arranged in the horizontal rather than vertical direction. The schematic of RICH2 is shown in Figure 3.15 [71].

For the HPDs to operate at full efficiency the magnetic field around them should not exceed 3 mT. Hence both RICH detectors are light-tight and surrounded by a magnetic shielding box [71].

3.2.5 Calorimeters

Calorimeters are detectors used for measuring the energy of particles produced in high-energy collisions as well as being used as particle identification tools. These detectors utilize a property of particle interaction with a bulk material called showering. This occurs as a particle is stopped and absorbed in a material. There are two types of particle

showers - electromagnetic and hadronic. Electromagnetic showers evolve through pair production and through bremsstrahlung of the incident particle and its secondaries. The scale of these showers is determined by the radiation length, X_0, of the material. The hadronic showers are more complex as many factors contribute to the inelastic production of secondary hadrons. The scale of these is determined by the nuclear interaction length, l_a. As the nuclear introduction length in a material is greater than the radiation length, hadronic calorimeters are often deeper than electromagnetic ones. Furthermore, due to various, not fully quantifiable losses during the hadronic showering, such the pion decay to muons and the production of neutrinos, which escape the HCAL, the energy resolution of hadronic calorimeters is usually around 20% worse [66].

There are two types of calorimeters - homogeneous and sampling calorimeters. In the case of the former, the bulk material acts both as the absorber and the detector. A sampling calorimeter consists of layers of material with a high atomic number, usually iron or lead, which acts as the absorber, interlayered by scintillating material that acts as the detector [79]. The LHCb experiment uses sampling calorimeters arranged in the so-called *shashlik*[11] structure. The calorimetry system at the LHCb experiment consists of the electromagnetic calorimeter (ECAL), accompanied by the scintillating-pad detector (SPD) and the preshower (PS) detector, and the hadronic calorimeter (HCAL) [71].

Electromagnetic calorimeter and associated subsystems

The ECAL at LHCb is made from 66 alternating layers of 2 mm thick lead absorber and 4 mm thick scintillator tiles and is situated just upstream of the HCAL and downstream of RICH2 and the first muon station. The lead/scintillator layers form a stack 42 cm deep, which corresponds to a radiation length of 25 X_0. The modules are wrapped in black paper to ensure light-tightness. The scintillator light is transferred to a Photo-Multiplier Tube (PMT) by wavelength-shifting (WLS) fibres. The ECAL is split into three sections with outwardly increasing cell sizes due to the large variance of the incident particle density with radius. The cell structure of the ECAL is shown in Figure 3.16. The ECAL covers the full[12] LHCb acceptance with the outer edge extending to 300 (250) mrad in the bending (non-bending) plane [71]. The energy resolution of the ECAL is given

[11] Named after a skewered meat dish originating from the present day Azerbaijan. In the analogy the wavelength-shifting fibres act as the skewers penetrating the consecutive detector layers through their centre.

[12] Similarly to RICH1, when referring to the entire calorimetry system, full coverage implies the inner edge of the acceptance of 25 mrad due to constraints imposed by the beampipe [71].

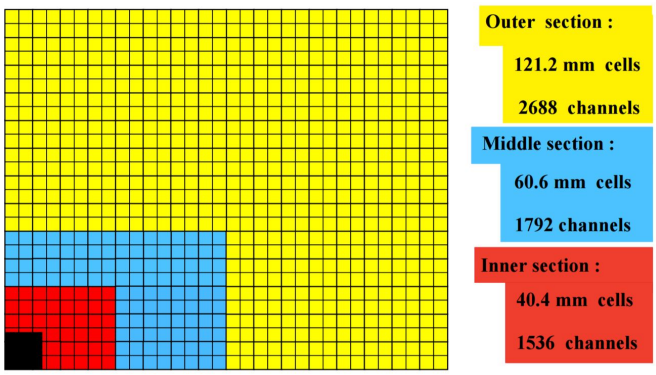

Figure 3.16: Front view of the top right quadrant of SPD/PS and the ECAL. Cell sizes cited are for the ECAL [71].

by [71]

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \tag{3.3}$$

where E is the measured energy in GeV.

As mentioned before, the ECAL is accompanied by two other sub-detectors - PS and SPD - located just upstream of the ECAL. These are two near identical scintillating pads with a 15 mm (2.5 X_0) lead absorber between them covering the same acceptance as the ECAL. Moreover, the readout system is also equivalent. The main reason for these detector planes in terms of the PID is to assist with the separation of photons, pions and electrons. As there is no absorber placed in front of the SPD it is predominantly blind to the incoming photons, which shower upon hitting the lead absorber and thus are visible in PS. Electrons, on the other hand, can shower in the SPD. Furthermore, the SPD/PS system helps to reject more than 99% of incident pions with an above 90% retention of electrons [71].

Hadronic calorimeter

The HCAL at LHCb is the most downstream component of the calorimetry system and is located just before the last four muon stations. It is built as a wall at $z = 13.33$ m with respect to the interaction point and is made of layers of iron and scintillating tiles. In contrast to the ECAL, the scintillating material runs in parallel to the beam line. In the transverse direction the iron absorber forms a 1 cm thick layer between the scintillating tiles whereas in the longitudinal direction the thickness of the absorber corresponds to a hadron interaction length in steel. The HCAL is transversely segregated into two sections with increased cell size in the radial direction. The cell structure of the HCAL is shown in Figure 3.17. The readout system of the PMTs and WLS fibres mimics that

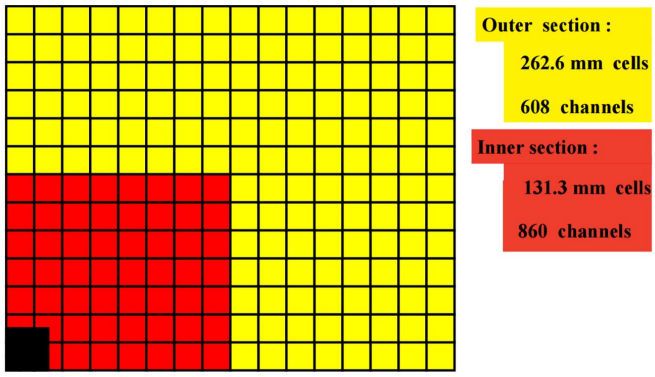

Figure 3.17: Front view of the top right quadrant of the HCAL [71].

of the ECAL [71]. The energy resolution of the HCAL has been determined to be [71]

$$\frac{\sigma_E}{E} = \frac{69 \pm 5\%}{\sqrt{E}} \oplus 9 \pm 2\%, \tag{3.4}$$

where E is the measured energy in GeV.

3.2.6 Muon system

The muon system consists of five stations labelled M1 to M5. M1 is sandwiched between RICH2 and SPD/PS whereas stations M2-M5 form the most downstream[13] particle detection device at LHCb. The muon system is made entirely of multi-wire proportional chambers (MWPC) with the exception of the innermost region of M1 where triple-GEM[14] technology is used. Both of these technologies utilize the ionisation of gas by traversing charged particles, where the ionised electrons are collected on an anode. The muon stations cover an angular acceptance of 20 (16) mrad to 306 (258) mrad in the bending (non-bending) plane. The layout of the five muon stations and the muon filters separating them is shown in Figure 3.18. The muon filters are 80 cm thick iron absorbers. These filter out low momentum particles, retaining only penetrating muons. The minimum momentum required for a muon to penetrate all five muon stations is approximately 6 GeV/c [71].

The stations are separated into logical pads with their size increasing in the radial direction. The breakdown of the M1 station is shown in Figure 3.19. Stations M1-M3 have a high spatial resolution in the bending plane as these are used to determine the track direction and to calculate the transverse momentum of the candidate muon

[13] With the exception of the HeRSCheL detector stations which are located outside the LHCb cavern on both sides of the LHCb experiment and will be briefly discussed in section 3.2.7.

[14] GEM stands for Gas Electron Multiplier.

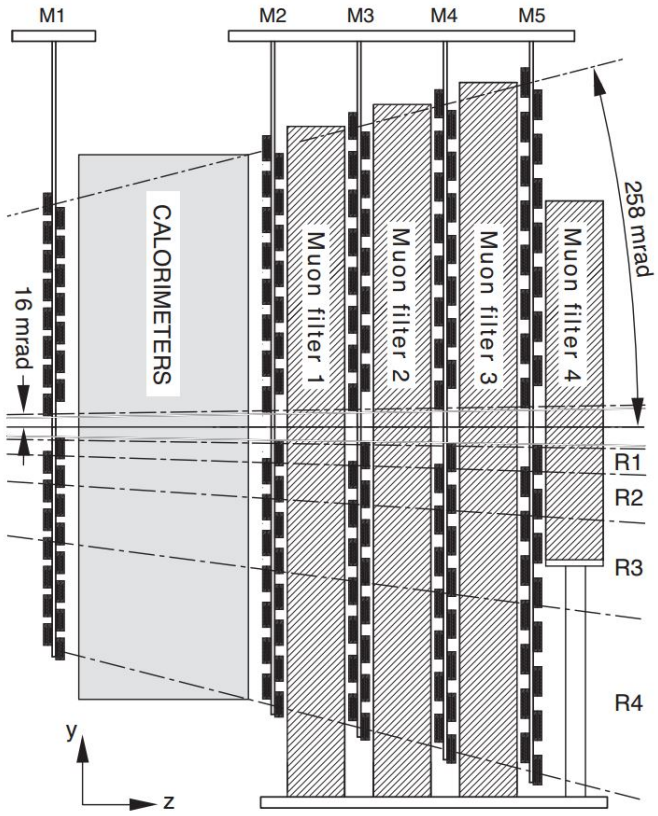

Figure 3.18: Side view of all five muon stations [71].

with a resolution of 20%, while M4 and M5 are used primarily for the identification of penetrating particles and thus have a lower spatial resolution [71].

It is important to note here that the muon systems, along with the calorimeters and VELO, are of great importance for the hardware trigger of the experiment, which will be discussed in detail in chapter 4.

3.2.7 HeRSCheL

HeRSCheL (High Rapidity Shower Counters for LHCb) is a system of scintillating-pad detectors situated at large distances in both the forward and backward region outside of the LHCb cavern [80]. These detectors are hit counters that extend the LHCb charged particle detection coverage in the backwards region from $\eta >$ -3.5 to $\eta >$ -8 and in the forwards region from $\eta <$ 5 to $\eta <$ 8. These counters are used for central exclusive production [81, 82] (CEP) measurements. These require the presence of Large Rapidity Gaps [83] (LRGs) in the production of particles. LRGs are also significant for the physics analysis presented in this thesis and will be discussed further in chapter 5. This detector was installed and tested during 2015 - 2016 and is being used for CEP measurements in Run-II. The schematic of the positioning of the HeRSCheL stations and an image

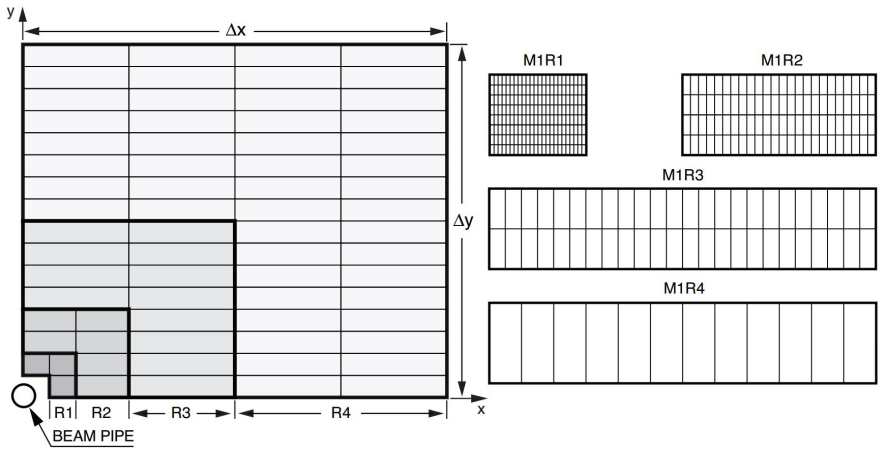

Figure 3.19: Front view of the top right quadrant of the most upstream muon station M1 (left). Division into logical pads with the dependence of their size on the radial position away from the beam axis clearly discernible (right) [71].

showing the two forward counters are shown in Figure 3.20.

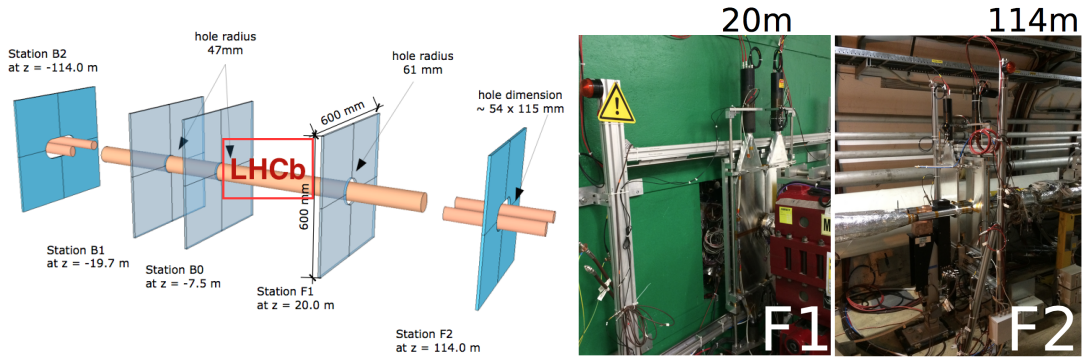

Figure 3.20: Left: a schematic overview of the HeRSCheL stations; right: two forward stations installed in the LHC tunnel [84].

3.2.8 VELO2

As the VELO operates at unprecedented proximity to the beam line at a hadron collider, it was unknown, during its construction and commissioning, if it could withstand the immense radiation doses through the entirety of Runs I and II. As a precaution, a twin detector - VELO2 - was built in parallel to the VELO and transported to CERN. Since its arrival it has been kept in a dual perspex casing where the internal casing surrounding the modules is constantly flushed with dry air in order to minimize damage that might occur due to the fluctuations in temperature, pressure and humidity. The two VELO2 halves are shown in Figure 3.21.

In 2014 a monitoring device - VELO2 Monitor - was designed, built and installed at IP8. The device is based on an Arduino Yún [85] micro-controller board with an

Figure 3.21: VELO2 being transported from the underground cavern to the exhibition area at UX85. The outer perspex casing has been removed during transport.

in-built ATmega32U4 [86] micro-controller with LinioOS [87] based on Unix [88]. The device accommodates a 3D accelerometer and temperature, humidity and pressure sensors. It is connected to the local network and thus the monitoring can be performed remotely, however it also hosts an LCD which cycles through the readings from all but the accelerometer in real time. A pair of these devices were created and installed, one for each half of VELO2. Figure 3.22 shows the device in operation.

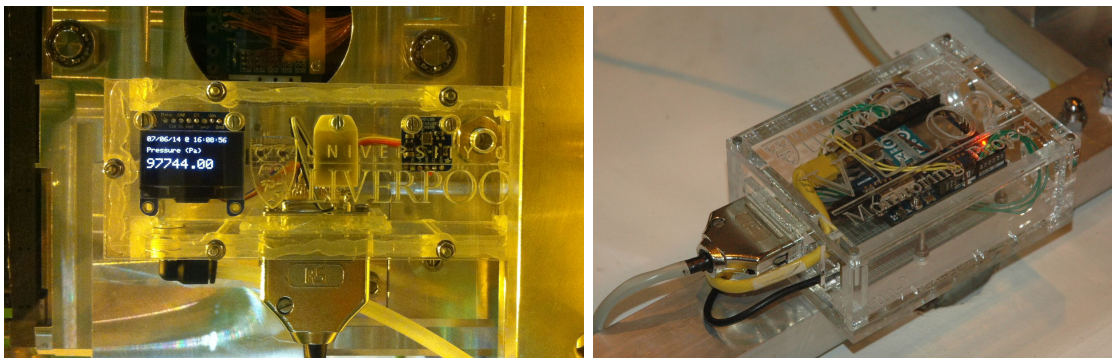

Figure 3.22: Left: upper half of the monitoring device containing the atmospheric sensors and the LCD currently displaying the pressure reading; right: lower half of the monitoring device containing the Arduino micro-controller board, the accelerometer and the power and network connections. The two are connected via a custom-made cable.

3.3 Current performance

3.3.1 Performance of the LHC

The performance of a particle collider can be characterised by the luminosity it has delivered. Luminosity, $\mathcal{L}$, at the LHC is related to the number of pp collisions the accelerator

can deliver to its experiments and is often expressed as instantaneous luminosity, the number of *pp* collisions deliverable to the experiments per unit time, or time-integrated luminosity, the total number of *pp* collisions delivered to the experiments. The former metric is the measure of the optimum peak performance of the collider while the latter suits well to characterise its efficiency. The instantaneous luminosity can be expressed using the parameters of the beam [89],

$$\mathcal{L} = f\frac{N_1 N_2}{4\pi\sigma_x\sigma_y}, \tag{3.5}$$

where N_1 and N_2 are the number of particles in the colliding beams, f is the beam crossing frequency and σ_x and σ_y describe the physical dimensions of the overlapping beam profile. The peak instantaneous $\mathcal{L}$ delivered by the LHC in Run-I was around $\mathcal{L} = 1.0\times10^{34}$ cm^{-2}s^{-1}, while the value for Run-II is around $\mathcal{L} = 1.5\times10^{34}$ cm^{-2}s^{-1} [90].

Run-I took place through 2010 - 2012. The original start in 2008 was abruptly cut short by a major malfunction. Subsequently, upon restart, it was decided to operate the accelerator at a considerably reduced beam energy compared to its design nominal value of 7 TeV per beam. In 2010/11 (2012) the LHC operated at the beam energy of 3.5 (4) TeV. During the first part of Run-II - 2015 - 2016 - the LHC has operated at the beam energy of 6.5 TeV, with the same beam energy planned for the remainder of the run (2017 - 2018). At the moment of writing, May 2017, the LHC and its experiments have just started the data taking for 2017. The integrated luminosity recorded by the four major LHC experiments in the four years of data taking through both Run-I and Run-II is shown in Figure 3.23.

3.3.2 Performance of LHCb

LHCb is a precision experiment where precise measurement of the position of the PVs is essential. As such, maintaining low pile-up is more important than increased statistics. Therefore LHCb employs luminosity levelling [71, 92], meaning that $\mathcal{L}$ is kept at an approximately constant value of around $\mu = 2.2$ throughout the fill. Thus, the total integrated luminosity at LHCb is lower than that of the GPDs, as shown in Figure 3.23.

The constant $\mathcal{L}$ is achieved by initially offsetting the LHC beams transversely and moving them progressively more in-line as the beams degrade. The measurement of μ is performed constantly at LHCb by the trigger pile-up system.

A further figure of merit of the performance of a detector is the efficiency. This can be simply expressed as the time during which the experiment is collecting data with

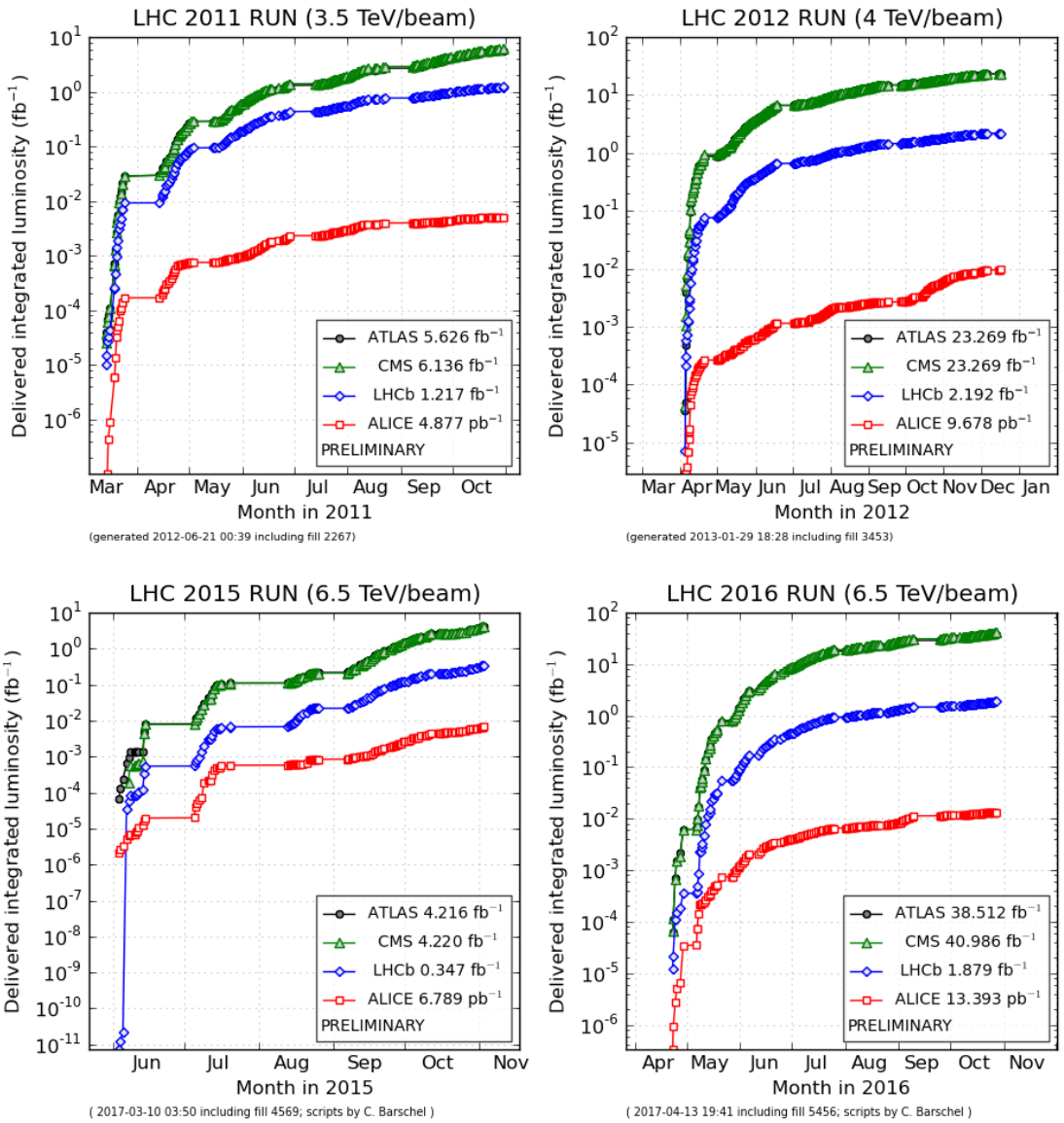

Figure 3.23: Integrated luminosity recorded by the four major experiments at the LHC in four years of operation. Top: 2011 (left) and 2012 (right); bottom: 2015 (left) and 2016 (right) [91]. 2010 is not shown as the majority of operation during it was dedicated to understanding the machine and the amount of data collected in this first year of stable operation was low.

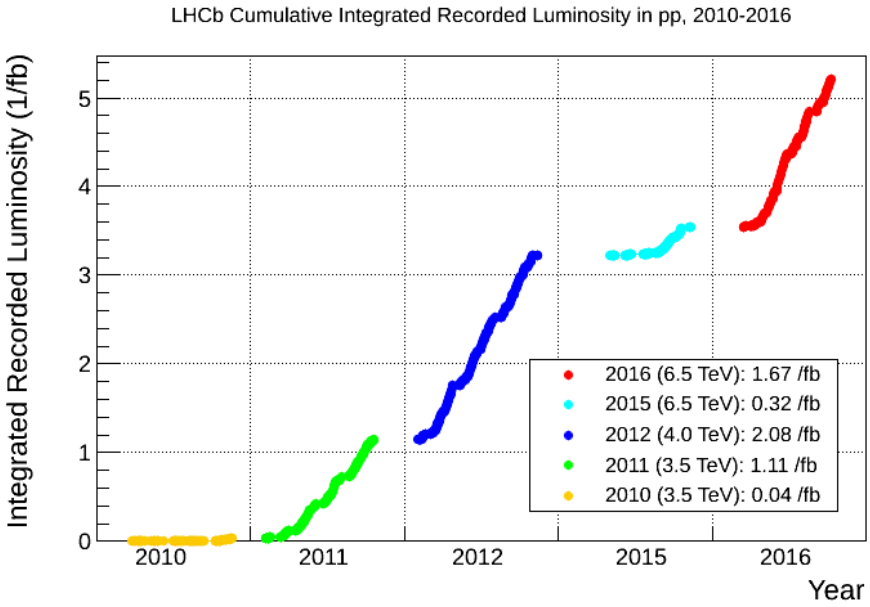

Figure 3.24: Total integrated luminosity recorded at the LHCb experiment between 2011 and 2016. [93].

respect to the total time the collider is providing the experiment with stable beams.

The total integrated luminosity collected by the LHCb experiment is shown in Figure 3.24. An example of the LHCb efficiency breakdown is shown in Figure 3.25.

3.3.3 Experimental output of LHC

The four LHC experiments have only been operating for about half a decade, yet have already produced superb physics results. The highlights, among a plethora of other fascinating physics results, have been the discovery of the Higgs boson [17, 18] by the ATLAS and CMS collaborations, evidence of quark-gluon plasma in pp collisions [94, 95] by the ALICE collaboration and the discovery of tetraquark [96] and pentaquark [97] states, angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay [98] and the observation of tantalising evidence of BSM physics via investigating lepton universality [99, 100] by the LHCb collaboration.

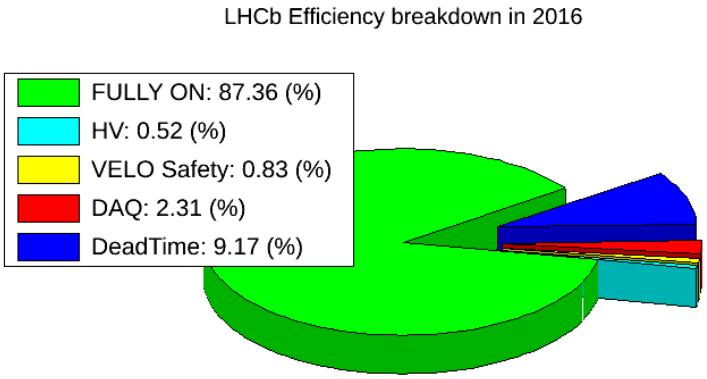

Figure 3.25: LHCb data collection efficiency in 2016 [93]. HV, VELO safety, DAQ and DeadTime show inefficiencies due to problems with the high-voltage supply, VELO sub-detector and data acquisition system and cumulative inefficiencies due to a mixture of other reasons respectively.

CHAPTER
FOUR

Event reconstruction, readout and software

"Non est ad astra mollis e terris via"
("There is no easy way from the earth to the stars")

- Seneca

This chapter is dedicated to the discussion of the operation of the LHCb experiment and the handling of the data it outputs. First, the LHCb trigger is described in detail in section 4.1. Next, the event reconstruction and detector alignment and calibration methods are discussed in section 4.2. This chapter concludes with section 4.3, giving a brief outline of the online systems, data storage and the software used by the LHCb collaboration.

4.1 The LHCb Trigger

The LHC machine delivers collisions at a rate of 40 MHz [71]. In an idealised situation, the experiments around the ring would read out and write to disk the information from every single collision. In practice this is not possible, as the data streams are too vast to be read out and the amount of information they contain is too bulky to be written to disk in full. It could be possible to simply write out as much as possible, at a realistic maximum rate of $\sim 10^3$ - 10^4 events per second, without any regard for the content. However, this is not a valid approach. Most collisions do not contain events that are strictly interesting for physics analyses and most of the processes of interest happen at a rate orders of magnitude lower than the total number of collisions. Therefore a way of selecting interesting events and discarding the rest is needed. It is the job of the trigger systems at the experiments to perform these selections in real time. The LHCb trigger

is split into two parts, the Level-0 hardware trigger and the high level software trigger.

4.1.1 Level-0

The Level-0 [71] (L0) trigger is a hardware trigger and the first stage of the event selection at LHCb. L0, synchronous with the collider, is used to reduce the rate from the output rate of the LHC machine to approximately 1 MHz. The hardware trigger is split into two categories - the calorimeter-based triggers and the muon triggers. The information from the two sections is fed to the L0 decision unit (L0DU), which makes the decision of whether to send a given event to the next trigger stage or to ignore it.

Calorimeter triggers

The calorimeter triggers at L0 use the information from the SPD/PS, ECAL and HCAL to identify the presence of a hadron, an electron or a photon with high transverse energy, E_T. First, the ECAL and HCAL locally select high-E_T deposits, usually with energies over thresholds of a few GeV, as shown in Table 4.1. The information from the selected clusters is merged with the information from the SPD/PS to determine if the clusters were deposited by a photon, an electron or a hadron. The combined information is passed to the Selection Crate, where only the candidate with the highest E_T in each category is selected. The Selection Crate is a modular system comprising eight custom-built selection boards controlled by credit card PCs and is located in a radiation free environment. From the Selection Crate the information is passed to L0DU. In addition, the total E_T deposit, the SPD multiplicity, n_{SPD}, is stored at this stage. The n_{SPD} variable is used by L0DU to veto events with the highest track multiplicities; these would take up a disproportionally large amount of computing resources as high event multiplicity leads to large processing times. This multiplicity cut is referred to as the Global Event Cut (GEC) [71].

Muon triggers

The second constituent of L0 is the muon trigger system. It is controlled by four processing units - one for every transverse quadrant of the muon stations. Each processing unit selects two muon candidates in an event with the highest and second highest transverse momentum, p_T. The high-p_T muon candidates are further required to have left a straight track segment through the muon stations, with the track segment pointing towards the interaction region. The seeds for the extrapolation algorithm are the hits

Trigger	Thresholds
`L0Muon`	$\mathtt{n}_{SPD} < 450$, $\mathtt{p}_T > 2.8$ GeV/c
`L0MuonEW`	$\mathtt{n}_{SPD} < 10000$, $\mathtt{p}_T > 6.0$ GeV/c
`L0Electron`	$\mathtt{n}_{SPD} < 450$, $\mathtt{E}_T > 2.8$ GeV
`L0Hadron`	$\mathtt{n}_{SPD} < 450$, $\mathtt{E}_T > 4.175$ GeV
`L0Photon`	$\mathtt{n}_{SPD} < 450$, $\mathtt{E}_T > 2.8$ GeV
`L0DiMuon`	$\mathtt{n}_{SPD} < 900$, $\mathtt{p}_T^1 \times \mathtt{p}_T^2 > 33.8$ (GeV/c)^2

Table 4.1: Examples of L0 trigger thresholds in Run-II with the `0x00A8 L0TCK` used in August 2015. $\mathtt{n}_{SPD}$ refers to the multiplicity at the SPD sub-detector and $\mathtt{p}_T$ and $\mathtt{E}_T$ are transverse momentum and energy, respectively. `L0MuonEW` is a trigger line designed specifically for the use of the electroweak physics working group at LHCb in order to study the impact of the Global Event Cut at L0 on the physics analyses using the `L0Muon` channel.

in M3. For each hit a straight line extrapolation to M2, M4 and M5 is performed by requiring the hits in these stations to be within defined fields of interest (FOIs). If a satisfactory set of hits for a candidate is found, the hit in M2 is used as a new seed for extrapolation to M1 and M3. If again a satisfactory set is found within the FOI, the muon candidate is selected. The pad size of the muon stations is scaled such that a single pad would provide the same angular coverage from station to station. This allows for the track extrapolation to be based on simple logical decisions and for the L0 muon trigger to make the decisions in the time allocated to it. The information of the selected candidates is then transmitted to L0DU [71].

L0 decision

The sub-detector trigger information arrives at L0DU at different times, so the time alignment is first performed. Then the information is combined and evaluated. This combined knowledge is compared with various requirements pre-installed via an L0 trigger configuration key (L0TCK). The L0TCK contains previously decided-upon selection thresholds for various L0 trigger lines. These trigger lines and examples of their respective thresholds are given in Table 4.1. The computation is performed by L0DU which then sends the information to the Readout Supervisor called ODIN, which makes the final decision of whether to accept an event at L0. The various trigger decisions are logically OR-ed by ODIN prior to the final decision being made. The latency, the time between an interaction taking place and the arrival of the information at L0DU, which includes the time-of-flight of particles, the interaction with the detectors and the

transmission of signal via cables, is fixed at 4 μs. This leaves L0DU 2 μs to make the decision [71].

Bandwidth division

The process during which the 1MHz of the L0 bandwidth allocation is performed is known as the bandwidth division. This process is based on the available computing resources and the planned physics program of the experiment. As the focus of the LHCb experiment is high precision flavour physics, the hadron stream, the event stream passing the `L0Hadron` trigger, is allocated the highest bandwidth. However many flavour physics decay channels are semi-leptonic and therefore a large bandwidth is allocated to the muon/di-muon and electron/photon streams as well. The availability of these channels allows LHCb to be competitive in a varied physics program.

The total L0 bandwidth is divided between three streams as shown in Figure 4.1. The charged hadron stream's allocation is 45% of the bandwidth (450 kHz), the muon and di-muon stream is given 40% (400 kHz) and the electron and photon stream is given 15% (150 kHz) of the L0 bandwidth [101].

4.1.2 High level trigger

The high level trigger (HLT) at LHCb is a two-part fully software trigger written in `C++` consisting of two stages - HLT1 and HLT2 [71]. The entirely software nature of HLT has made it immensely flexible and it has evolved drastically over the years of LHC operation. Part of this evolution is shown in Figure 4.1 comparing the whole trigger chain in Run-I and Run-II.

During 2011 the two HLT applications ran in sequence synchronously to the LHC and L0. The output rate achieved at the end of the applications was 3.5 kHz or 0.2 GB/s. In 2012 the output rate was increased to 5 kHz or 0.3 GB/s and approximately 20% of the events were deferred to disk between the two levels of HLT, with HLT2 now running partially asynchronously. For Run-II major improvements were introduced in the form of extra available computing resources, allowing for an output rate of 12.5 kHz or 0.6 GB/s and for HLT1 and HLT2 to be run fully orthogonally with 100% of the events buffered to disk [101, 103].

Both stages of the HLT run on the Event Filter Farm (EFF) located at LHCb. It consists of 1,700 computing nodes with ∼27,000 physical and ∼52,000 logical CPU cores and hosts 5 (10) PB of mirrored (un-mirrored) disk storage space [101, 103].

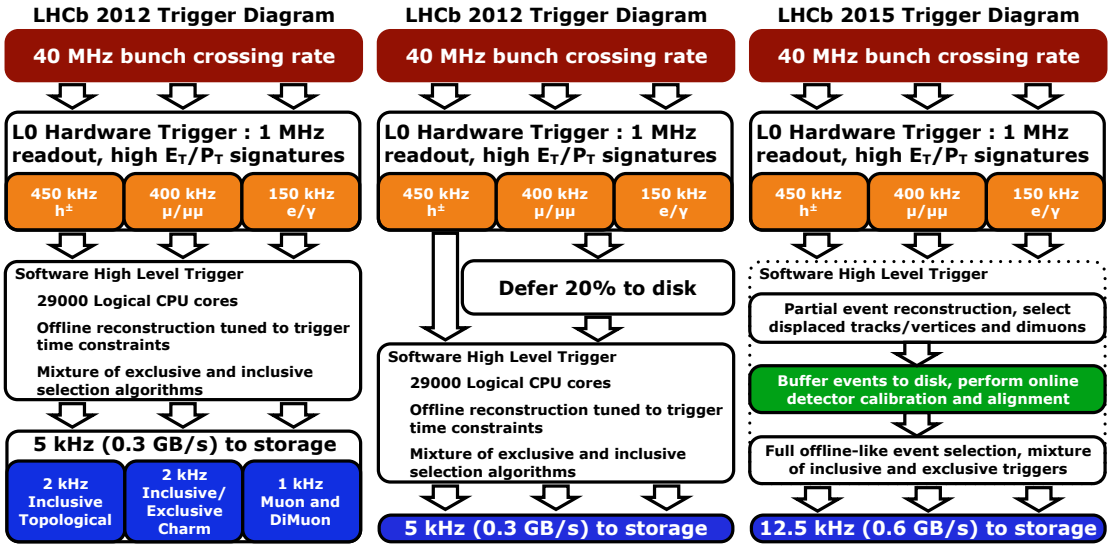

Figure 4.1: Left to right: Run-I trigger with no data deferral, Run-I trigger with 20% of data deferred to disk and the trigger scheme to be used through Run-II (2015-2018) [102].

HLT1

HLT1 is the first software stage of the trigger. It accepts the events from L0 and performs a simplified track fitting procedure. First, T seeds, the hits in the T stations, are reconstructed with the assumption that these originate from the interaction region. Next, the simplified 2D VELO tracks are reconstructed in the $r-z$ plane. If these satisfy a track χ^2 criterion and originate from the interaction point, the ϕ - sensor information is added and a 3D VELO track is reconstructed. Furthermore, the 2D tracks are used to reconstruct the primary vertices. Then the information from the VELO and the T stations is combined [71].

The primary objective of HLT1 is to validate or discard decisions made by L0 and to reduce the rate from L0 output to approximately 150 kHz to be passed on to HLT2. The selections, similarly to L0, are performed by a set of trigger lines which are pre-installed via a trigger configuration key (TCK) [71]. Examples of typical HLT1 lines are given in Table 4.2. Moreover, HLT1 also provides dedicated samples of data for the real-time alignment and calibration, further discussed in section 4.1.2 [103].

HLT2

The average event size output from the HLT1 is ∼55 kB, meaning that the available disk buffer allows for 150 hours of data taking before the HLT2 application must be run [101, 103]. This does not immediately present problems, as due to the degradation of the LHC beam no fill lasts for such a period of time. Care, however, needs to be taken

HLT1 line	Selection requirements
`HLT1SingleMuonHighPT`	`L0Req` = {`L0Muon or L0MuonEW`} `p` > 6.0 GeV $\mathtt{p}_T$ > 4.34 GeV `TrChi2` < 3 `MaxOTHits` = 15000 `MaxITHits` = 3000 `MaxVeloHits` = 6000
`HLT1SingleElectronNoIP`	`L0Req` = `L0Electron` `p` > 20.0 GeV $\mathtt{p}_T$ > 10.0 GeV `TrChi2` < 3 `MaxOTHits` = 15000 `MaxITHits` = 3000 `MaxVeloHits` = 6000
`HLT1DiMuonHighMass`	`L0Req` = {`L0Muon or L0DiMuon`} `p` > 3.0 GeV `VxDoca` < 0.2 `VxChi2` < 25 `IPChi2` < 4 `TrChi2` < 3 `MaxOTHits` = 15000 `MaxITHits` = 3000 `MaxVeloHits` = 6000

Table 4.2: Selected Run-II HLT1 line examples with 0x00FB0051 TCK used in August 2015. The `L0Req` slot is used to select only the events passing the given L0 line. `p` and $\mathtt{p}_T$ slots are cuts on the candidate particle's minimum momentum and minimum transverse momentum. `TrChi2`, `VxChi2` and `IPChi2` slots are the quality cuts on the track and vertex fits and the determination of the impact parameter, respectively. `VxDoca` slot is the cut on the maximum distance of closest approach of the vertex to the nominal beamline. Finally, the `MaxOTHits`, `MaxITHits` and `MaxVeloHits` slots are Global Event Cuts cutting on the maximum multiplicity of the event in the respective detectors.

as the interval between the fills is often not long enough to run the HLT2 application over the whole buffer and the computing resource priority has to be handed back to running HLT1 once data taking resumes, which can lead to cumulative data build-up on the disk.

The job of the HLT2 application in Run-II is to fully reconstruct the events to offline-quality[1], perform the event selection and cut the output data rate to 12.5 kHz. The event selection, similarly to the previous two stages, is done via a set of trigger lines which are pre-installed via a TCK. Unlike at the previous two stages, the HLT2 triggers can be considerably more sophisticated with a maximum allowable event processing time of 650 ms/event compared to 35 ms/event at HLT1 [103]. Examples of typical HLT2 trigger lines are given in Table 4.3.

The triggers at HLT2 and the previous stages can also be prescaled. A prescale means that only a designated fraction of the events passing a certain trigger line will be accepted. The use of a prescale introduces an extra selection efficiency uncertainty and tuning of the trigger line cuts for rate control, whilst maintaining a high signal efficiency, is preferred. Nevertheless, a use of a prescale can be beneficial when an interesting type of event either has a higher than expected data retention rate or very high multiplicity, resulting in a higher than desired event size.

Real-time alignment and calibration

The switch to a fully deferred HLT2 for Run-II facilitated an introduction of a novel feature at LHCb. Dedicated data samples are collected by HLT1 to be used for detector alignment and calibration in real time. The calibration takes a few minutes and is performed before HLT2 is run, thus the new calibration constants can be used in the second stage of the software trigger, improving the data quality. During this process the alignment of more than 1,700 detector components are examined and nearly 2,000 calibration constants are calculated [104].

The alignment and calibration process can be performed at the start of every LHC fill or more often if deemed necessary. Every update of the calibration constants triggers a run change[2] at LHCb and thus a range of values around a given central figure of merit

[1] In Run-I there was a difference of the data quality at the online stage (trigger) and the offline stage (end-user) due to the detector misalignments and the unavailability of information from some sub-detectors, such as RICHs, in the online regime; with the improvements made over the LS1 and the introduction of a fully deferred HLT2 these differences have been all but eliminated.

[2] Run is a assignment of a unique number to a certain amount of continuously taken data. If no other reason for triggering of a run change occurs, a run is changed every hour as a minimum. This process limits the amount of data that can be corrupted for any one reason upon data taking.

HLT2 line	Selection requirements
HLT2EWSingleMuonLowPt	Input = BiKalmanFittedMuons p > 4.8 GeV prescale = 0.002
HLT2EWSingleElectronVHighPt	Input = BiKalmanFittedElectrons L0Req = L0Electron Hlt1Req = ElectronDecision Pt > 1.5 GeV PrsMin = 50 MeV EcalMin = 0.1 HCALMax = 0.05 TkChi2 < 20
HLT2EWDiMuonDY1	Inputs = TrackFittedDiMuon MinMass = 2.5 GeV MinExcMass = 3.0 GeV MaxExcMass = 3.2 GeV MuPt > 0.8 GeV/c TkChi2 < 10
HLT2EWDiElectronHighMass	Inputs = TrackFittedDiElectron L0Req = L0Electron Hlt1Req = ElectronDecision MinMass = 20 GeV VtxChi2 = 25 TkChi2 = 10 ElecPt = 10 GeV PrsMin = 50 MeV EcalMin = 0.1 HCALMax = 0.05

Table 4.3: Selected Run-II HLT2 line examples with 0x00FB0051 TCK used in August 2015. The Inputs slot is used to define the particle container from which the trigger line should select the required particle candidates. The L0Req and Hlt1Req slots are used to select only the events passing the given L0 and HLT1 lines. The p and Pt slots are cuts on the candidate particle's minimum momentum and minimum transverse momentum. The prescale slot is used to define the fraction of the events that pass this trigger line that are written to disk. The PrsMin, EcalMin and HCALMax slots define the minimum or maximum required deposit in the PS, ECAL and HCAL respectively. The TkChi2 and VtxChi2 slots set the cut on the minimum fit quality of the track and vertex respectively. The MinMass and MinExcMass and MaxExcMass slots define the minimum mass cut and the mass exclusion window respectively. Finally, the MuPt and ElecPt slots define the minimum transverse momentum requirements of the daughter particles to the particle on which the trigger line is designed to fire.

of alignment is allowed to avoid over-fragmenting the data stream. The alignment and calibration process of various detectors will be further discussed in section 4.2.4.

Turbo stream

With the use of a deferred trigger, increased computing time allowance in HLT2 and the introduction of real-time alignment and calibration, another novel improvement to the physics analyses at LHCb has been made available. This is the so-called Turbo stream [105].

For certain analyses, such as the analysis of charm meson decays, it is not necessary to store the whole 55 kB event to disk. Instead, these analyses only need information about their respective signal candidates. Such analyses are often statistically limited and would benefit greatly from an increased trigger throughput. With the introduction of the real-time alignment and calibration, the data quality at the trigger stage already matches the quality obtained by re-running over data in an offline regime. This means that, for certain channels, there is no longer any need for the entire event to be written to disk, as the raw event information is no longer necessary for offline reprocessing.

These events, having been fully reconstructed in the online regime and passed through the Turbo stream, have an event-size of around 5 kB. Thus a tenfold increase of statistics for these analyses can be achieved with no increase to the overall bandwidth used [101, 103, 105, 106].

No-bias data

No-bias data is a set of data that is collected with no explicit selection bias. A preset dedicated fraction of the L0 bandwidth of 1 MHz is used up for selecting events using random ODIN triggers. This means that in this regime the trigger works as a throughput by randomly selecting as many events as it is capable of writing to disk. Such datasets are dominated by the events of soft QCD processes. The physics analysis presented in this thesis uses a no-bias data sample, the specifics of which will be introduced in section 5.2.2.

4.2 Event reconstruction

The events triggered upon by L0 and HLT1 undergo a full offline-quality reconstruction in HLT2 using all the available information from the LHCb sub-detectors.

4.2.1 Track reconstruction

When traversing the LHCb detector, a charged particle leaves a series of charge deposits, hits in the tracking sub-detectors described in sections 3.2.2 and 3.2.3. These hits are described by coordinates in the four-dimensional space, ie. both their three-dimensional coordinates and the time of occurrence is recorded by the detector. A track is simply a trajectory connecting such hits that satisfies certain quality requirements. There are five track type designations at the LHCb experiment:

- **Long tracks** traverse the entire tracking system; these have the most precise momentum estimate and thus are most useful for physics analyses;
- **Upstream tracks** pass only through the VELO and TT; these are typically low momentum tracks bent out of the LHCb acceptance by the dipole magnet;
- **Downstream tracks** leave hits only in the TT and T stations; these typically come from long-lived particles that decay outside the VELO acceptance, such as K_s^0 and Λ baryons;
- **VELO tracks** pass only through the VELO; these include all backwards tracks[3] and those with large angle with respect to the beam line;
- **T tracks** have hits only in the T stations; these are mostly tracks produced in secondary interactions.

A visual representation of the track types at LHCb is shown in Figure 4.2.

The charged track reconstruction begins with track seeding. Straight line segments in the VELO are constructed from matching hits in three[4] VELO stations. These straight line segments are denoted as the VELO seeds. A similar procedure of local hit matching is done at the T stations and the resultant line segments are denoted as T seeds [71,108].

After seeding is performed, the track matching process begins. First, the forward tracking is done, where the VELO seeds are extrapolated forwards and matched with hits in TT and the T stations. This process approximates the deflection of a charged particle trajectory while traversing the magnet to a single point-like change of direction at the centre of the magnet. Next, unused VELO and T seeds are extrapolated towards

[3]The VELO sub-detector provides a backwards coverage of -3.5 $< \eta <$ -1.5. The backwards tracks are defined as tracks traversing this region.

[4]As discussed in section 3.2.2, the spacing of the VELO stations is chosen such that any track in the LHCb acceptance would traverse no less than three of these stations.

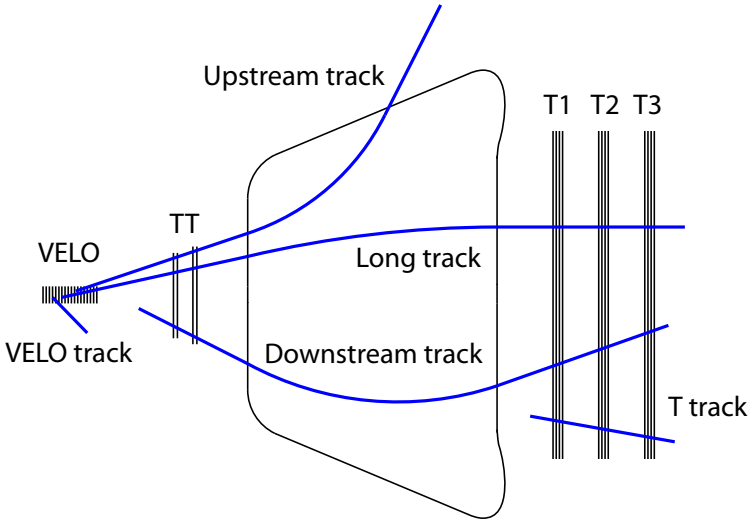

Figure 4.2: Pictorial description of the different track types at LHCb overlaid over a top-down view of the experiment [107].

the centre of the magnet in search for a match. The trajectories matched here and during forward tracking become long tracks. Next, the algorithm attempts to match VELO and T seeds to hits in TT, resulting in the upstream and downstream tracks, respectively. After all the matching processes are completed, the remaining unmatched VELO and T seeds are denoted as VELO and T tracks [71, 108].

Following the matching process track fitting algorithms are used to fit the obtained tracks. The trajectories described by these fits are used to determine the momentum of the particles responsible for the tracks. At LHCb a Kalman filter [109] is used for the track fitting procedure. The Kalman filter is an iterative procedure that can remove hits within a trajectory and re-fits the trajectories until they match designated quality requirements [71, 109].

A key measure of the performance of the tracking system is the track reconstruction efficiency. In the estimation of the track reconstruction efficiency only long tracks are used, as the other track types are rarely used in physics analyses. The reconstruction efficiency is determined using the tag-and-probe method with the muons from the $J/\psi \to \mu^+\mu^-$ decays. The long track reconstruction efficiency is consistently found to be greater than 95% for long tracks with $p_T > 500$ MeV throughout the operation of the LHCb experiment [110].

An example of track reconstruction efficiency is shown in Figure 4.3 for the data gathered in 2012. Furthermore, the reconstruction efficiency in the Monte-Carlo used in 2012 is also shown for comparison.

Another important metric of measuring the performance of the tracking system is

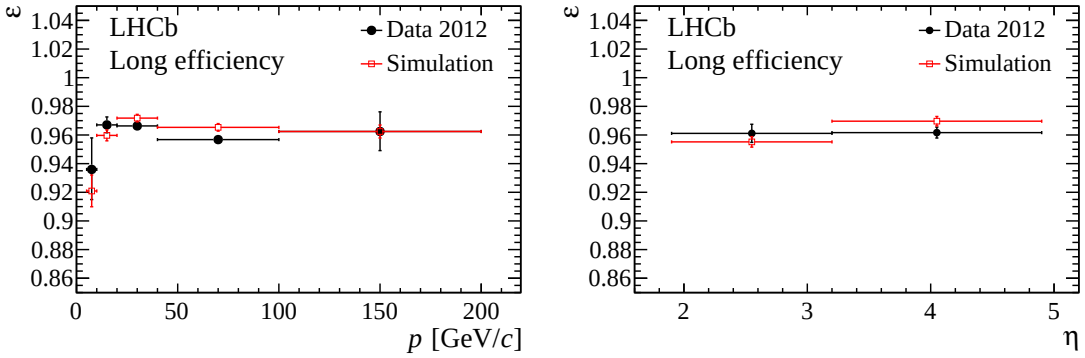

Figure 4.3: Long track reconstruction efficiency in the 2012 data (black) and 2012 Monte-Carlo (red) as a function of track momentum (left) and pseudorapidity (right) [110]. Equivalent figures for the 2015 data are not yet publicly available.

the momentum resolution. For this also the $J/\psi \to \mu^+\mu^-$ decays are used. The mass resolution of the J/ψ is given by the combination of the momentum resolution of the two muons and their masses can be neglected. Considering decays where the momentum of the two muons is similar, the momentum resolution $\delta p/p$, can be approximated as

$$\left(\frac{\delta p}{p}\right)^2 = 2\left(\frac{\sigma_m}{m}\right)^2 - 2\left(\frac{p\sigma_\theta}{mc\theta}\right)^2, \tag{4.1}$$

where m is the invariant mass of J/ψ candidate, σ_m, is the uncertainty on this mass obtained by fitting a Gaussian to the resonance and the second term is a correction for the muon opening angle, θ, with σ_θ denoting the percentage error on it. Again, only the long tracks are considered. The momentum resolution of charged tracks at LHCb has been determined to be $\delta p/p = 0.5\%$ for $p < 20$ GeV/c rising to $\delta p/p = 0.8\%$ at 100 GeV/c [111]. The momentum resolution as a function of charged particle momentum is shown in Figure 4.4.

Furthermore, the quality and efficiency of both the track reconstruction discussed here and the vertex reconstruction discussed in the following section depend on the intrinsic hit efficiency of the tracking detectors. Hit efficiency is measured using reconstructed tracks and extrapolating these through all tracking detectors to determine whether or not an expected hit is recorded. The overall hit efficiency is determined to be above 99.7% for all silicon sensors [111] and above 99.2% for the OT [77].

4.2.2 Vertex reconstruction

A point from which particles in a collision at a particle collider originate is referred to as a vertex. The points of the primary interaction of the particle beams are known as the

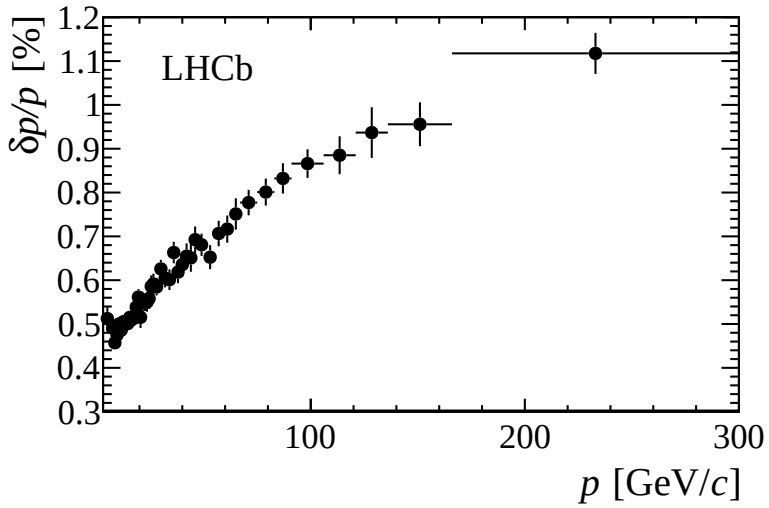

Figure 4.4: Momentum resolution of charged particle tracks as a function of their momentum [111].

primary vertices (PVs) while the vertices formed by the decays of the particles arising from PVs are called secondary vertices (SVs). The primary vertex reconstruction is based on reconstructed tracks [71, 112].

The PV reconstruction begins with a seeding process which looks for clusters of tracks along the beam line defined by their z coordinate, z^{clu}, and the uncertainty associated with it, σ_Z^{clu}. Here, z coordinate is the intercept of the Kalman fitted VELO track with the nominal beam line. The reconstructed clusters are sorted in a descending order of multiplicity which results in reconstruction of higher multiplicity PVs first. All resulting clusters are considered in pairs in turn and the pair with the minimum distance, D^{pair}, is selected, where this distance is defined as

$$D^{pair} = \frac{|z^{clu_A} - z^{clu_B}|}{\sqrt{\left(\sigma_z^{clu_A}\right)^2 + \left(\sigma_z^{clu_B}\right)^2}}. \tag{4.2}$$

If D^{pair} of a given pair is less than 5, it is merged into a new cluster, with z_{pair}^{clu} and $\sigma_{Z\ pair}^{clu}$ determined using the weighted mean method. The pairing process continues until no more pairs pass the D^{pair} requirement. The merged clusters are used as seeds in the vertex fitting procedure [112].

The PV fitting employs an adaptive weighted least squares method where the Tukey biweight method [113] is used to assign weights to the tracks forming the vertex seed based on their χ^2_{IP}. Here χ^2_{IP} is defined as the increase of the χ^2 of the PV fit with the inclusion of a given track in the fitting procedure; it is required to be $\chi^2_{IP} < 9$. The

track weight, W_T, is assigned using

$$\begin{aligned} W_T &= \left(1 - \frac{\chi^2_{IP}}{C^2_T}\right)^2 \quad \text{for } \chi_{IP} > C_T \\ W_T &= 0 \qquad\qquad\qquad\quad \text{for } \chi_{IP} \leq C_T, \end{aligned} \tag{4.3}$$

where C_T is the Tukey constant. All track types with hits in the VELO - long, upstream and VELO - are used for vertex reconstruction [112].

The position of a PV is obtained by minimizing χ^2_{PV}, where

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \chi^2_{IP,i} \cdot W_{T,i}. \tag{4.4}$$

The position finding described above is an iterative process. After each iteration the tracks of the vertex seed are extrapolated towards the new z position of the PV. The process stops once the following conditions are satisfied:

- $|\Delta z| < 0.5$ μm, where Δz is the shift of the PVs z coordinate between the iterations;

- a PV has at least 5 good quality ($\chi^2_{IP} < 9$) tracks assigned to it.

To avoid convergence at local minima, a high value of C_T is used initially and decreased during the first five fitting iterations to the nominal value of 3 [112].

Similarly to the track reconstruction, a measure of performance of the vertex reconstruction procedure is the resolution with which the PV position is obtained. For PVs, this resolution varies with the number of tracks originating from it as shown in Figure 4.5. It ranges from around 10 (50) μm to 35 (250) μm in the transverse (longitudinal) direction as a function of number of tracks present in a given vertex [114].

Additionally, the impact parameter (IP), the minimum distance between a track and the PV is measured with a resolution, σ_{IP}, given by

$$\sigma_{IP} = 15 \pm \frac{29}{p_T} \quad \mu m, \tag{4.5}$$

where p_T is the track's transverse momentum in GeV/c [115].

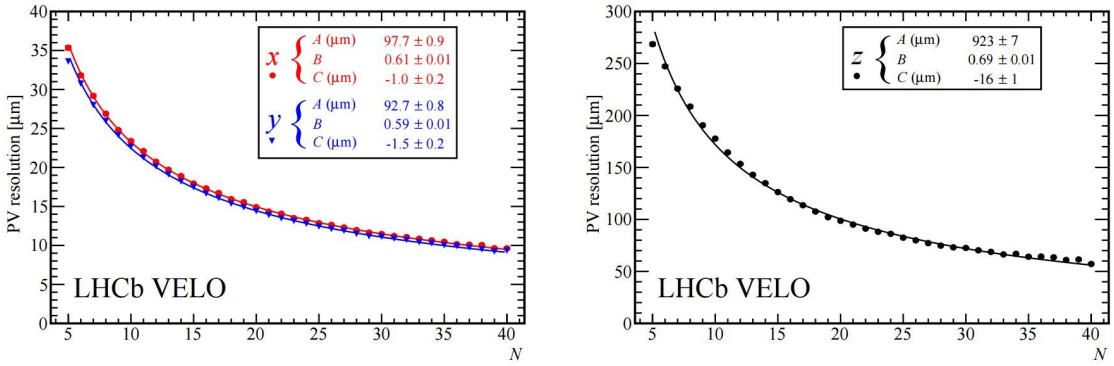

Figure 4.5: PV resolution in single PV events in 2011 ($\sqrt{s} = 7$ TeV) data as a function of the track multiplicity for x (red) and y (blue) coordinates (left) and z coordinate (right). It varies from 10 to 35 μm in the transverse direction and between 50 and 250 μm in the longitudinal direction. The fit (solid lines) is parametrised in terms of the number of tracks, N, using $\sigma_{PV} = \frac{A}{N^B} + C$. The fit parameters are shown in the boxes [114].

4.2.3 Particle identification

PID is an important tool for many physics analyses that look for final states containing specific species of particles. Furthermore, it is necessary for a precise estimation of particle momentum. PID at the LHCb experiment is performed by the two RICH detectors, the calorimeters and the muon system, all of which are described in chapter 3.

Photon identification

The photons at LHCb are identified using the ECAL by looking for energy deposit clusters with no associated track. To veto charged particles, all charged tracks are extrapolated to the face of the ECAL and a cluster-to-track estimator, χ^2_γ, is calculated. Clusters with sufficient separation from tracks, $\chi^2_\gamma > 4$, are identified as photons. Further improvement on $\gamma - e$ separation is achieved using SPD/PS. Unlike electrons, which leave charge deposits in both the SPD and PS, the photon-induced showers are more likely to begin after the photon has traversed the lead layer between the two [71, 111].

Neutral pions, π^0, are identified by matching two neighbouring photon clusters to the invariant mass of the π^0. If such a match is made, the two photons are merged into one particle denoted as `ResolvedPi0` with a mass resolution of 8 MeV/c^2 . If a $p_T^{\pi^0} > 2$ GeV/c the clusters cannot be resolved due to the finite granularity of the ECAL. Such overlapping clusters are denoted as `MergedPi0`. The mass resolution for such π^0 candidates drops to around 30 MeV/c^2 [111].

A photon identification efficiency of 95% can be achieved with a `MergedPi0` background rejection of 45% [111].

Electron identification

The electron candidates are identified using the ECAL and SPD/PS. A cluster-to-track estimator is built similarly to that for the photon to match the calorimeter cluster with tracks while looking for corresponding hits in the SPD and PS. Electron energy losses through bremsstrahlung are also recovered at this stage by attempting to match all photon clusters to an electron track before it traverses the magnet. Any photon emission by an electron after traversing the magnetic field leads to both particles depositing their energy in the same cluster reconstructed as an electron with their combined energy [71, 111].

A cut of $p_T > 0.5$ GeV/c is applied to the electron candidates resulting in electron reconstruction efficiency of around 95% with a charged pion mis-identification rate of 0.7% [71].

Muon identification

For a track to reach M2 and M3 it has to have a minimum momentum of $p > 3$ GeV/c, hence only such tracks are examined when identifying muons. Hits are sought in the FOIs in the same way as the L0 muon trigger candidates are found. Straight tracks with hits in a sufficient number of muon stations are labelled as muons, with tracks with $p > 10$ GeV/c being required to leave hits in all four farthest muon stations M2-M5. In order to be identified as a muon, a track must also have $p_T > 0.8$ GeV/c [71, 111].

The muon identification efficiency is measured to be above 94% with a 3% ($\pi \rightarrow \mu$) misidentification rate for all muons with $p_T > 1.7$ GeV/c [116]. Combining these with PID performed by the RICH detectors, the mis-identification rate drops to 1% [71].

Hadron identification

The PID of charged hadrons is performed using the RICH sub-detectors. The separation of the different charged hadron species using the C_4F_{10}, as found in RICH1, is shown in Figure 4.6. Algorithms based on the log-likelihood approach [117] are employed for distinguishing between different PID hypotheses. Averaging over the momentum range covered by the RICH detectors of $2 < p < 100$ GeV/c, this results in a kaon identification efficiency of 95% with a ($\pi \rightarrow K$) mis-identification rate of 10% when requiring that the kaon mass hypothesis is simply greater than the pion mass hypothesis - $\Delta log\mathcal{L}(K - \pi) > 0$. Tightening this requirement to $\Delta log\mathcal{L}(K - \pi) > 5$ reduces the mis-identification rate to 3% while only slightly reducing the efficiency [71, 111].

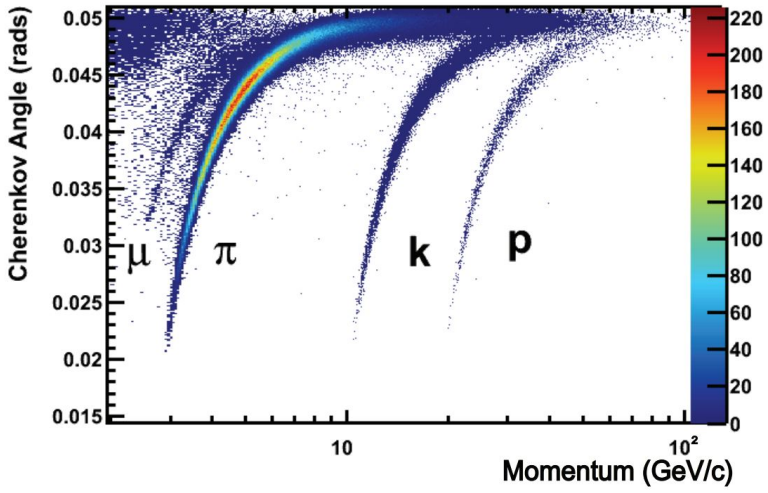

Figure 4.6: Reconstructed Čerenkov angle for different particle species in the C_4F_{10} radiator found in RICH1 [111].

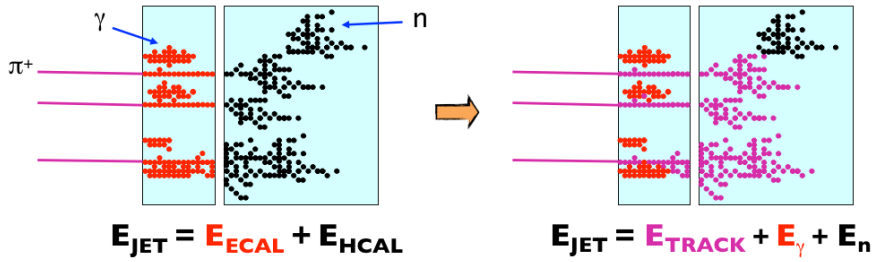

Figure 4.7: Left: regular calorimetry method; right: use of the particle flow approach [118].

Particle flow algorithm

The use of the particle flow (PF) algorithm is an approach of using calorimetry information at high-energy physics experiments. A successful use of the PF approach at an experiment requires a high-precision tracking system combined with a high-granularity ECAL.

The algorithm traces the paths of charged particles through the calorimeters using the information from the tracking system and combines the individual track four-momenta with the deposits in the calorimeters as shown in Figure 4.7. The transverse and longitudinal granularity of the calorimeters has to be sufficient to avoid 'confusion' between different clusters in the calorimeters as shown in Figure 4.8.

The particle flow algorithm at LHCb utilizes its high-precision tracking to determine the charged particle momentum with high accuracy. The charged long tracks are matched with the calorimeter clusters, which are clustered based on their separation in $(\eta \times \phi)$ space and, combining the momentum and PID information, a correspond-

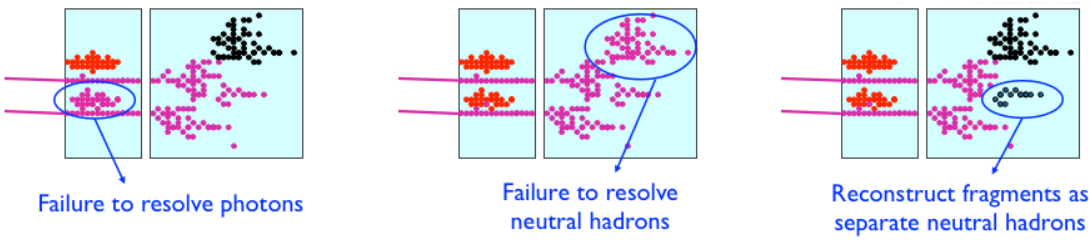

Figure 4.8: Various sources of 'confusion' using the PF algorithm, which need to be minimised [118].

ing amount of energy is subtracted from the energy deposits in the calorimeters. The muons are identified and no calorimeter clusters are associated with this species of particle. This is repeated by the algorithm until all charged tracks are removed. The remaining energy deposits are assigned to the neutral particles by taking the neutral clusters in the ECAL to be photons and associating the neutral clusters in the HCAL with being neutral hadrons. The use of the entire detector by a PF algorithm allows for improved energy reconstruction and further improves particle identification [118, 119]. The analysis outlined in this thesis uses the particle flow approach.

4.2.4 Detector alignment and calibration

In order to successfully reconstruct the events at LHCb alignment and calibration of the LHCb sub-detectors has to be performed. As discussed in section 4.1.2, the calculation of the alignment and calibration constants is performed in the online regime before the HLT2 application is run. The recalculation of these constants offline is also possible if required.

The alignment of the various tracking components is performed by the minimisation of residuals - the distance between the fitted trajectory of a track and the hits said track left in the tracking stations. This process is shown in Figure 4.9.

VELO alignment

In order to deliver high quality vertex and track reconstruction, the VELO detector has to be aligned to a precision of a few μm in the x and y directions and a few tens of μm in the z direction. The alignment of such precision is especially challenging as the VELO is a moving detector and is closed and retracted with every fill of the LHC.

The alignment of the R and ϕ sensors at the module level is performed upon construction via a precision survey and is known to an accuracy of 3 μm in translation and 20 mrad in rotation. Furthermore, the movement of the VELO detector is performed

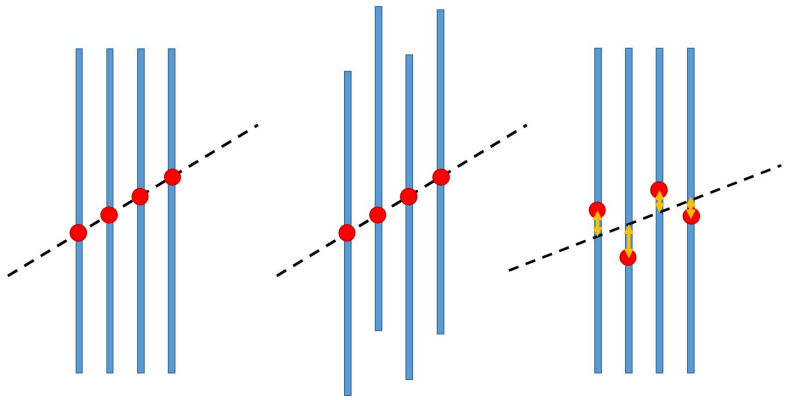

Figure 4.9: An exaggerated view of the alignment process of a tracking detector using the minimization of residuals. Left: a perfectly aligned tracking stations in blue, traversing track in black and the hits of the track in the stations in red; middle: real situation of a slightly misaligned detector; right: reconstructed track hits and trajectory if the misalignment is not taken into account. The yellow arrows show the residuals that are to be minimised to obtain the best alignment.

by a stepping motor with motion control precision of 10 μm [71, 111].

The alignment of the two detector halves relies primarily on the tracks passing through the overlap of the two. Minimization of a χ^2 function of the residuals of such tracks, as shown in Figure 4.9, is used to align the two VELO halves to a precision of a few μm in x, y translation and to 0.2 mrad in rotation around the z axis [71,101,104,111].

As discussed in section 4.1.2, the alignment procedure is done online prior to the HLT2 application being run, to guarantee the best reconstruction quality. Figure 4.10 shows the variation in μm in the VELO alignment during a time period in 2016. It can be seen that the realignment was deemed necessary approximately 50% of the time during this period.

Alignment of the tracking system

The required precision of the alignment of the silicon trackers and the outer tracker of tens of μm and around 100 μm respectively cannot be performed with a mechanical measurement of the physical location of the detectors as the ramping of the magnetic field in the LHCb dipole magnet can shift the tracking stations by as much as 5-15 mm. In order to determine the alignment of these detectors long tracks are used to align them with respect to the VELO sub-detector. This alignment is performed using the minimization of the residuals from the Kalman fitted track trajectories. Furthermore, the drift time of charge, t_0, in the OT is aligned to the global LHCb clock during every alignment process to a value of $\Delta t_0 < 0.1$ ns [101, 104, 111].

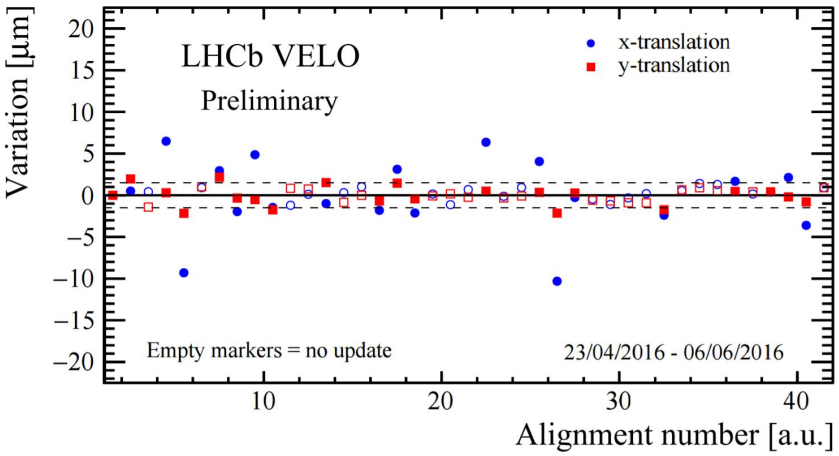

Figure 4.10: The variation in the alignment of the VELO sub-detector in μm as a function of the number of times the alignment task was run. The empty markers show times when the realignment was deemed unnecessary while the full markers show the times the realignment was performed. [103].

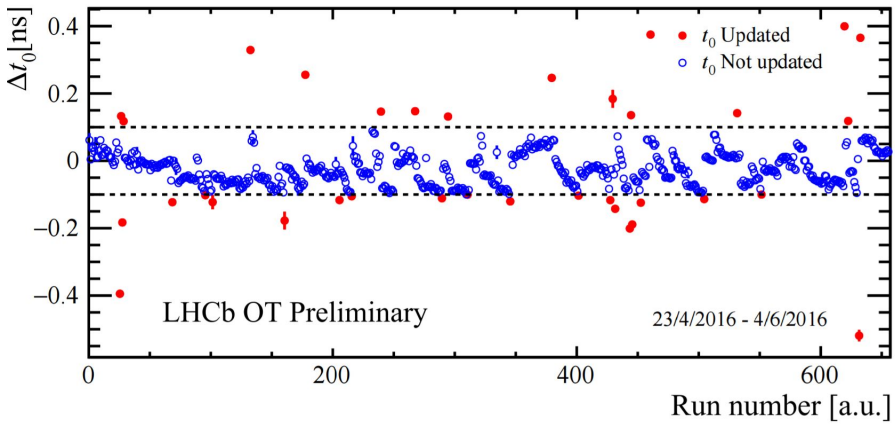

Figure 4.11: The variation in the time alignment of the OT sub-detector in ns as a function of the number of times the alignment task was run. The empty blue markers show times when the realignment was deemed unnecessary while the full red markers show the times the realignment was performed. [103].

The variation in the OT drift time alignment in ns is shown in Figure 4.11 for the same time period as for the VELO. Unlike the VELO, the OT is not moved for every LHC fill and the realignment procedure is deemed necessary only about 10% of the time.

Calibration of the RICH detectors

Both RICH detectors use two sets of mirrors to focus the Čerenkov light. The first set of spherical mirrors focuses the light onto a set of flat mirrors which reflect it onto the HPDs Therefore any slight misalignment of the mirrors results in unwanted distortions in the reconstructed Čerenkov angles.

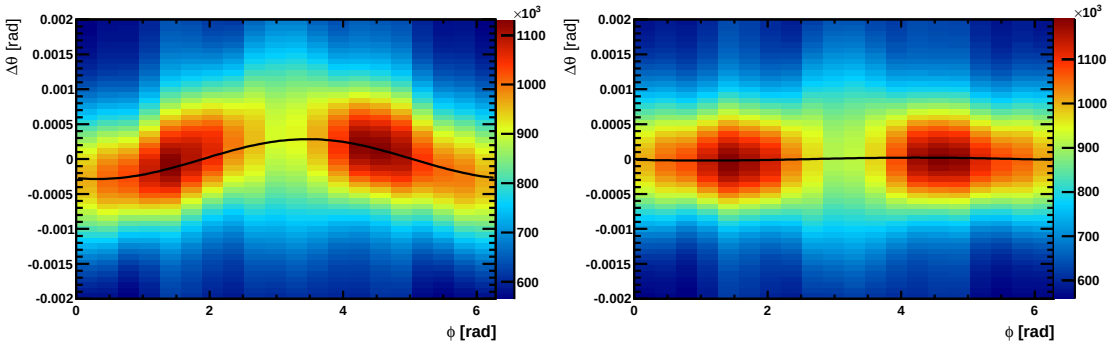

Figure 4.12: The variance of the Čerenkov angle as a function of ϕ in RICH2 fitted with $\theta_x cos(\phi) + \theta_y sin(\phi)$. Left: prior to the realignment; right: after the realignment is performed [120].

The mirror alignment process is performed similarly to that of the tracking detectors. The variation of the Čerenkov angles prior to and after the alignment process for RICH2 are shown in Figure 4.12.

Calibration of the calorimeters

The calibration of the calorimeters is important to maintain their design energy resolution discussed in section 3.2.5. The calibration of the ECAL is performed by fitting the $\pi^0 \rightarrow \gamma\gamma$ mass peak. The HCAL is calibrated using a ^{137}Cs source transported through the scintillator cells by a hydraulic system. As the process takes a considerable amount of time, the HCAL calibration takes place during the technical stops which occur approximately bi-monthly [71, 111].

A good detector alignment is invaluable for maximization of the momentum resolution and PID performance and for optimization of the quality of information used by the reconstruction software and the particle flow algorithm.

4.3 Online system, data storage and software

The volumes of data recorded by the LHCb detector would not be of any use without appropriate systems managing the detector, the data flow it outputs and permanent storage of the data.

4.3.1 Online system

At LHCb the readout of the data from the detector front-end electronics and detector management is performed by the online system consisting of the Data Acquisition system

(DAQ), Timing and Fast Control (TFC) and Experiment Control System (ECS) [71].

The design of the DAQ is based on cost-effectiveness and, wherever possible, commercially available components are used. The readout of the data from all front-end electronics of all sub-detectors, with the exception of the RICH sub-detectors, is done by an LHCb-wide standardized readout board - TELL1 [71]. The point-to-point transfer of data is carried out using optical and Gigabit-Ethernet links [71].

The TFC system is responsible for synchronisation of the LHCb detector operation and readout via a beam-synchronous clock and runs the Readout Supervisor [71].

The control and monitoring of the entire detector is performed by the ECS. It is responsible for the supply of the low and high voltages to the detector components as well as controlling the gas flows, temperatures, pressures and monitoring of the DAQ, TFC and the trigger systems [71].

4.3.2 Data storage

The data collected by the LHCb experiment, as well as data from other experiments around CERN, are sent to the CERN Computing Centre (CCC), where they are stored on tape[5]. This monolithic block of data is too large to be efficiently used for individual physics analyses and therefore it is reprocessed into streams based on specific types of events required by different analysis working groups at the experiment. This is done via a process called stripping [71].

The data storage and reprocessing along with the production of Monte-Carlo simulation samples is performed by a system of computing clusters located at research centres and universities around the globe. These sites are divided into tiers: Tier-0, Tier-1 and Tier-2. CERN is the only Tier-0 site and a copy of all data gathered by the experiments is stored on tape at the CCC. A second copy of data is distributed to all Tier-1 sites. These are where the reprocessed data are also stored and are available, upon a request by a physics analyst, to be copied to disk for more efficient access by a system called Castor. CERN also acts as a Tier-1 site. Tier-2 sites are predominantly universities, including the University of Liverpool, and are mostly tasked with the production of Monte-Carlo samples [71].

[5]Storage of data on tape might seem archaic. However, it is far more useful for prolonged storage than disk due to inevitable disk failure in long term.

4.3.3 Software

The copious number of tasks required to successfully run a high-energy physics experiment requires complex dedicated software. A custom built software framework - Gaudi [121] - is used as a base for software projects running various aspects of the LHCb experiment.

A project called Gauss [122] is used for configuring the stand-alone Monte-Carlo generators used by LHCb and simulating the physical detector and its response using Geant4 [123]. The process is continued by the Boole [124] application which simulates the readout of the LHCb detector and digitises the simulated data in a manner indistinguishable from the real data. From there the Brunel [125] application is employed for the reconstruction of the digital data output either by the real detector or Boole. The trigger software project is called Moore [126] with the Turbo stream implemented with an application named Tesla [105].

The distribution of the vast amount of data between the Tier-1 and Tier-2 sites and the end-user is performed using a global computing network created for the specific purpose - the Grid [127, 128]. The submission of required analysis code to be run over this data using the Grid is handled by the Dirac [129] and Ganga [130] applications.

Finally, the framework predominantly used by individual users to perform physics analyses at LHCb is called DaVinci.

CHAPTER

FIVE

Measurement of the forward energy flow

"All have their worth and each contributes to the worth of the others"

- J.R.R. Tolkien

In the following chapter the measurement of particle multiplicity and energy flow in pp collisions at a centre of mass energy of $\sqrt{s} = 13$ TeV with the LHCb detector is presented. Section 5.1 outlines the scientific motivation behind this measurement, while sections 5.2 and 5.3 introduce the datasets, the machine configuration during the data taking and the various theoretical predictions used in the analysis. The event selection is discussed in section 5.4 and section 5.5 describes the analysis process in detail. Finally, the results of the analysis are presented in section 5.6.

5.1 Motivation

High-energy collisions of hadrons, such as the collisions of protons at the LHC, are extremely complex. Unlike the collision process in lepton colliders, where two point-like objects scatter in a single interaction, the composite nature of hadrons leads to a considerably less clean environment. In such collisions the main pp scatter is accompanied by the so-called underlying event (UE). The UE consists of the outcome of the showering and hadronisation process of the remnants of the hadrons involved in the main scattering event as well as multi-parton interactions (MPIs).

MPIs, as introduced in section 2.2.6, are additional interactions between the constituent partons of the two colliding hadrons and give rise to enhancements in the observable final states through both hard and soft regime contributions. While the hard

scattering component of the interaction is computable using perturbative QCD calculations, the softer contributions cannot be described using the same tools.

The contribution from MPIs tends to increase with increasing c.o.m. energy of the hadron collider. This can be experimentally observed by comparing the increase in the total pp cross-section, with the increase in the collision energy, to the increase in the cross-sections of various observable processes as shown in Figure 5.1. It can be seen that the rise in the total pp cross-section is much flatter with increasing c.o.m. energy than the rise of the production cross-sections of various final states. An increased MPI contribution with increasing c.o.m. energy allows for enhancements in production cross-section of, for example, multi-jet final states to be accounted for in Monte-Carlo models in a consistent manner [47].

The reason for the increase in MPIs with the c.o.m. energy is due to the increase in the energy of individual partons, both valence and from the quark-gluon sea. As the energy of the hadron grows, partons with progressively lower Björken-x, and therefore increasing fluxes, are sufficiently energetic to participate in a scattering event, enhancing the MPI contribution [47].

5.1.1 Multi-parton interactions at LHCb

The study of MPIs at progressively higher energies is important in order to fully understand their contribution to the observables in hadron collisions and, ultimately, aid in the development of a coherent theoretical description of such collisions in both the hard and the soft regime. Thus it is important to perform measurements sensitive to MPIs at LHC energies. Such studies were carried out with data taken at $\sqrt{s} = 7$ TeV by the ATLAS [133], CMS [134] and LHCb [135] collaborations. With the aforementioned MPI dependence on the c.o.m. energy, these studies should be repeated in Run-II, at $\sqrt{s} = 13$ TeV.

As the c.o.m. energy increases, partons with a lower fraction of the proton's momentum can be probed. Furthermore, the flux of partons increases with a decrease in the fraction of momentum they carry. Therefore an increased flux of partons becomes relevant to an interaction process as the Björken-x is reduced. The above implies that MPIs become increasingly relevant at larger values of pseudorapidity, as for a collision to produce a lot of deposits in the high-η region at the LHC the participating parton from one of the proton beams has to have a lower Björken-x than the parton from the other beam. The forward instrumented coverage of the LHCb detector is unique among the

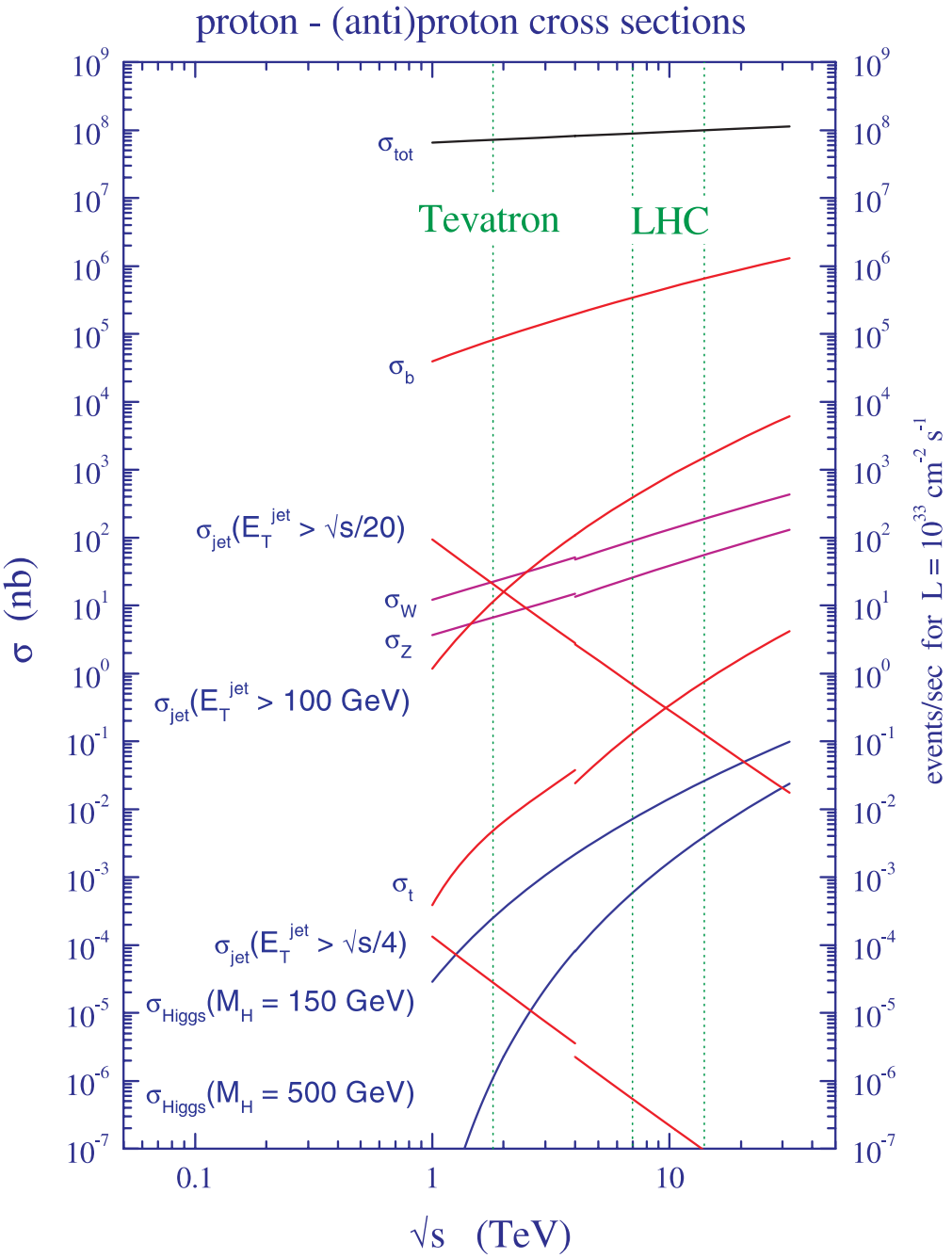

Figure 5.1: Standard Model cross-section at the Tevatron [131] and LHC energies. The cross-section of production of various final states rises more steeply than the total pp cross-section [132].

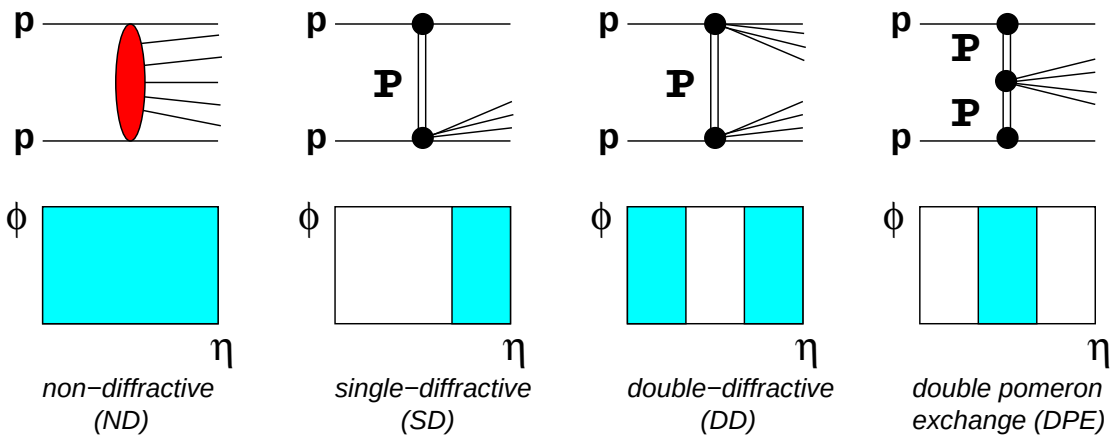

Figure 5.2: Non-diffractive and diffractive collision processes. The top row shows the topology of the interaction; the bottom row shows the $(\eta \times \phi)$ phase-space coverage of the products of the corresponding interaction. From left to right: Non-diffractive (ND) interaction with a colour octet (gluon) being exchanged; single-diffractive (SD) process, a Pomeron exchange, followed by the disassociation of a single proton; double-diffractive (DD) process, a Pomeron exchange followed by the disassociation of both protons; double Pomeron exchange (DPE), central-exclusive $\mathbb{PP}$ interaction [47].

experiments around the LHC and, with the collider currently providing the highest-ever energy hadron collisions, LHCb is the best-suited experiment to study MPIs.

5.1.2 Diffractive processes

One way to examine the UE at a hadron collider is through the study of diffractive events. Diffractive scattering is a soft regime interaction which dominates the no-bias data collected by the high-energy experiments. In the diffractive regime the interaction of the two protons is assumed to be mediated by a two-gluon colour singlet state - a Pomeron - coupling directly to the incident protons [47]. The difference between non-diffractive and diffractive events is depicted in Figure 5.2.

In a diffractive interaction no colour is exchanged and thus the protons either scatter elastically or one or both are excited to a higher energy state, which later decays into a multi-particle final state via dissociation. In the case of one or both of the protons dissociating, the interaction is referred to as single- or double-diffractive. Finally, an exchange of two Pomerons can take place, where the two $\mathbb{P}$ interaction results in the observable final state and the two incident protons remain intact. The lattermost is known as central-exclusive production [47].

As shown in Figure 5.2, diffractive events can be identified by observing particle production in the $(\eta \times \phi)$ phase-space. The topology of the diffraction products result in either one (single- and double-diffractive events) or two (CEP events) empty regions

of particle rapidity, also known as Large Rapidity Gaps (LRGs). The contribution of the MPIs is used by the Monte-Carlo generators to account for the observed survival probability [136] of LRGs in hard diffraction processes [47].

5.1.3 Study of UE and MPIs at LHCb

As discussed in the previous sections, the UE and MPIs become more prominent with increasing c.o.m. energy, as well as at higher values of pseudorapidity, and dominate the no-bias data samples collected by the high-energy physics experiments. With its forward instrumentation and the ability to trigger on low energy objects, LHCb is an excellent experiment at which to study these soft QCD phenomena. Furthermore, a correlation between the MPI activity and the particle multiplicity observed at an experiment is expected. This implies that an analysis studying the particle multiplicity in the forward region at LHCb provides a method of investigating the UE and MPIs. In addition to the multiplicity, the energy flow in the region can be determined. This provides a fuller understanding of the parton showering and hadronisation processes as it encompasses the information on the amount of energy and types of particle produced in a given region of the phase-space. Moreover, the separation of the analysis sample into event classes in order to distinguish between non-diffractive and diffractive processes can yield further information on particle production processes at various scales of momentum transfer.

The analysis presented in this thesis is the measurement of the forward particle multiplicity and energy flow in pp collisions at $\sqrt{s} = 13$ TeV recorded with the LHCb detector using a no-bias data sample. As previously stated, a similar analysis was performed using the Run-I data [135]. Although some aspects of the analysis described here, such as the event selection requirements and the definitions of the event classes, mimic the Run-I analysis, most of the analysis strategy is fundamentally different.

Firstly, in this analysis the multiplicity and average energy are measured over a two-dimensional energy versus pseudorapidity space. Then, the multiplicity distributions are unfolded to correct for the detector effects. Finally, the multiplicity and average energy distributions are combined to obtain the energy flow distributions. Furthermore, the neutral multiplicity and energy component is measured separately. In addition, the use of the particle flow algorithm, unique VELO ID requirement and the handling of the various systematics, manifest major differences in the approach with respect to the previous analysis performed using the Run-I data. The following sections give a detailed account of the analysis procedure.

5.2 Machine configuration and datasets

5.2.1 Configuration of the LHC and LHCb

The restart of the LHC for Run-II began in late December of 2014 after nearly two years of dormancy during Long Shutdown 1 (LS1). During the first half of 2015 the cooling down of the collider to its operating temperature took place followed by the introduction of particle beams. Then the beam energy was gradually built up by the LHC machine until the first stable beams were reached on the 3rd of June 2015 at the Run-II nominal centre-of-mass energy of $\sqrt{s} = 13$ TeV. Over the next two months the machine went through the phase of intensity ramp-up. This is a process during which the charge in the beam, and subsequently the instantaneous luminosity delivered to the experiments by the LHC, is gradually increased. During this stage the LHC beams were run with a bunch spacing of 50 ns and a maximum of 400 bunches per beam.

The LHCb detector was fully calibrated, tested and commissioned for Run-II during the first six months of 2015.

5.2.2 Data samples

The data samples used for the analysis presented in this thesis were collected in July and August 2015 during the so-called 'Early Measurements' (EM) period. It is during this period that the increased bunch spacing and reduced intensity beams were used. The datasets are no-bias data, which, as discussed earlier in section 4.1, means that the events are selected using ODIN's random sampling trigger. The same trigger configuration key was used for data collected with both magnet polarities. Further to the random trigger, the `HLT2MBNoBiasLeadingCrossing` trigger line was used to filter the recorded events. This trigger line operates as a pass-through for an HLT1 line with an equivalent name, which takes in the L0 decisions made by ODIN. The purpose of this HLT2 trigger is to only select events from the leading bunch crossings. A leading bunch crossing is defined, in this case, as any beam-beam crossing for which the previous two beam crossings have not been beam-beam. The designation 'beam-beam' refers to a crossing where both LHC beams traverse the interaction region simultaneously. This requirement allows for any possible backgrounds due to spillover to be removed. Spillover refers to backgrounds that occur when the charge deposits in any of the sub-detectors from an event of a given beam crossing are assigned to an event from the following beam crossing as a consequence of the limited readout time resolution of the detector.

Property	Magnet Up	Magnet Down
Centre-of-mass-energy	13 TeV	
Bunch spacing	50 ns	
TCK	`0x00F9014E`	
Number of events (data)	289,821,375	514,879,308
ν (MC)	1.6	
Number of events (MC)	10,031,879	10,006,120

Table 5.1: Summary of the properties of the datasets used.

5.2.3 Analysis Monte-Carlo samples

The Monte-Carlo samples used in the analysis process were produced with an LHCb tune [137] of the Pythia 8 [53, 54, 138] generator. The LHCb tune's default PDF set, CT09MCS, was used, with a value of $\alpha_S(M_Z)$ of 0.118 for the calculation of the elements of the hard scattering matrix. A brief outline of the Pythia 8 MC generator is given in section 5.3.1.

The detector simulation was performed using the LHCb simulation version, Sim09b, further tailored to specifically match the conditions present during the 'Early Measurements' data taking period - Sim09b EM. The main difference between the Sim09b and Sim09b EM samples is the bunch spacing of the beams being set to 50 ns instead of the nominal 25 ns to match the running conditions of the collider at the time.

Two minimum-bias samples were created, one for each magnet polarity, containing 10,031,879 and 10,006,120 events for magnet up (MU) and magnet down (MD) configurations respectively. The average number of visible interactions in the simulated samples is $\nu = 1.6$. The properties of the data and MC samples used for this analysis are summarized in Table 5.1.

5.3 Monte-Carlo predictions

In addition to the MC sample used in the analysis process, four more sets of theoretical predictions were obtained using the three MC generators introduced in section 2.2.8. These are two further sets of predictions created using two tunes of the Pythia 8 MC generator - LHCb [137] and Monash 2013 [139] -, as well as two sets of predictions produced using two versions of the cosmic-ray MC generators - Sibyll 2.1 [58] and EPOS LHC [43].

5.3.1 Pythia 8.212

Pythia 8.212 is the latest stable version of the Pythia MC generator and is a standard tool for generating events for high-energy physics experiments.

Pythia 8.212 provides a reliable description of pp interactions for collisions in the $10 < \sqrt{s} < 10^5$ GeV c.o.m. energy range. The ISR and FSR contributions are evolved using p_T-ordered algorithms. The values of $\alpha_S(M_Z)$ are set separately for the hard process, ISR, FSR and the MPIs and single, double and central diffraction [138]. Processes in Pythia are ordinarily calculated to LO, though NLO predictions can be produced.

MPIs in Pythia are modelled in order of descending p_T [140] and both soft and additional hard scatterings are described in a single unified framework. In addition to the partonic QCD 2→2 scattering, other activity, such as charmonium and bottomium production and γ+jet process, among others, have been added to the mix of MPI processes [138].

The momentum transfer cutoff for MPIs, p_{T0}, at a given energy scale, λ, is governed by

$$p_{T0}(\lambda) = p_{T0}^{ref} \left(\frac{\lambda}{\lambda^{ref}} \right)^{\beta}, \tag{5.1}$$

where p_{T0}^{ref} is the p_T reference parameter, λ^{ref} is the energy scale reference parameter and β is a parameter governing the energy rescaling pace [141].

Pythia handles all three diffractive topologies - single diffraction, double diffraction and central diffraction (DPE) - according to the Ingelman-Schlein model [142]. In this model the Pomeron is treated as a glueball-like hadronic state [138].

LHCb Tune

The LHCb tune [137] of Pythia is a version of the standard Pythia 8.212 MC generator tailored for use at the LHCb experiment and is the tune used for generating the analysis MC described in section 5.2.3. The parameters regarding the particle production have been tuned using the data from the CLEO [143], BABAR [144] and the LHCb experiments. As with the version of the tune used for producing the analysis MC, this set uses the CT09MCS PDF set with the $\alpha_S(M_Z)$ value of 0.118. The MPI cutoff governing parameters for the LHCb tune are set as follows: $p_{T0}^{ref} = 2.88$ GeV/c, $\lambda^{ref} = 7000$ GeV, $\beta = 0.238$.

Monash 2013 Tune

The Monash 2013 tune of PYTHIA 8.212 has been tuned using the minimum-bias, Drell-Yan and UE data collected by the LHC in combination with the data from the SPS [145] and Tevatron. The former data are used to constrain the ISR, FSR and MPI parameters, while the latter are used to constrain the energy scaling. The Monash 2013 tune uses the NNPDF2.3 QCD+QED LO PDF set and the default value of $\alpha_s(M_Z) = 0.130$ [139]. The MPI cutoff governing parameters for the Monash 2013 tune are set as follows: $p_{T0}^{ref} = 2.28$ GeV/c, $\lambda^{ref} = 7000$ GeV, $\beta = 0.215$ [139].

5.3.2 EPOS LHC

EPOS LHC is an updated version of the EPOS 1.99 cosmic-ray event generator for minimum-bias hadronic interactions and has been tuned with the minimum-bias, *p-Pb* and *Pb-Pb* data from the LHC experiments. The non-diffractive process in EPOS is modelled as QCD 2→2 scattering regulated by a p_T cutoff parameter. MPIs are modelled, with their probability calculated from the geometrical overlap function of the proton matter distribution. The overlap function and the value of p_T are tuned using experimental minimum-bias and UE data [43]. The diffractive events are modelled using Regge-based Pomeron model [146, 147].

Uniquely among the MC generators used here, EPOS accounts for collective effects, such as the collective flow and the core-corona treatment of the string fragmentation introduced in section 2.2.8. The high density core in the model is produced by the overlap of the string segments due to MPIs [43, 55].

The default parameters of the tune were used in producing the theoretical predictions.

5.3.3 Sibyll 2.1

SIBYLL 2.1 is another cosmic-ray MC generator for minimum-bias hadronic interactions. This generator was designed to describe the experimental data up to $\sqrt{s} = 2$ TeV and tuned using data from experiments at HERA [148], but has since been re-tuned using the $\sqrt{s} = 7$ TeV data from the LHC experiments.

As with the other generators discussed here, SIBYLL 2.1 incorporates Regge theory for the description of the soft scattering component. The soft interactions are simulated as a pair of gluons. Multiple soft interactions are simulated, with at least one of the multiple interactions always involving a valence quark [58, 149].

Unlike with the other models, diffraction is not split into different topologies and is treated with a strict kinematic cutoff. This results in the maximum energy loss of a proton via quasielastic diffraction processes of 20% [58]. The default parameters of the generator were used in producing the theoretical predictions.

5.4 Event selection

The first step in the process of producing the measurements described in this thesis is the definition of the four event classes in which these measurements are divided. The four classes are

- inclusive minimum-bias,
- hard scattering,
- diffractive enriched,
- non-diffractive enriched,

where inclusive minimum-bias covers all selected events and the three following classes are sub-classes of the inclusive minimum-bias event class.

Before the selection requirements of the four event classes are discussed any further, the definitions of three terms - 'detector level' and 'generator level' and 'fiducial selection' - need to be introduced.

Detector level, or reconstruction level, refers to the events as they are recorded by the detector, meaning that all real data collected by the LHCb experiment is recorded at the detector level. Similarly, all events produced by a Monte-Carlo generator that have had the detector simulation applied to them and have been reconstructed using the BRUNEL application are referred to as detector level.

Generator level, or truth level, refers to the events as they were produced by the *pp* scattering process or, in simulation, as they were created by the MC generator before the detector simulation is applied. Generator level, therefore, is fully accessible in MC. The extrapolation of the data recorded by the LHCb experiment to what was produced in the *pp* collision is the main goal of the analysis outlined in the following sections.

Fiducial selection is a set of cuts restricting the event selection to a specific portion of a phase-space. In the case of this analysis the fiducial acceptance covers $2.0 < \eta < 5.0$ and $2.0 < p < 1000.0$ GeV/c in $(\eta \times p)$ space. The η cuts are chosen to match the

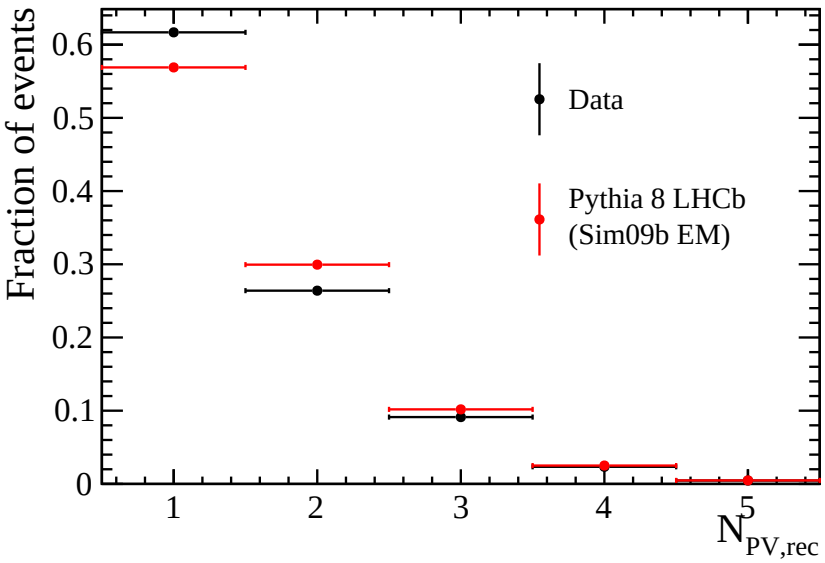

Figure 5.3: The number of reconstructed primary vertices in data (black) and Monte-Carlo (red). The distributions are normalized to unity.

detector acceptance. Likewise, the lower p bound arises from the difficulty to reconstruct particles with momentum lower than 2 GeV/c. The upper momentum bound of 1000.0 GeV/c was chosen as tracks with momentum higher than this have a large reconstruction uncertainty due to their large bending radius when traversing the magnetic field. Moreover, such tracks are extremely rare, making up less than $10^{-3}\%$ of all tracks in the data samples used for the measurement described in this thesis. The fiducial cuts are applied to all events both at the detector and generator level.

5.4.1 Inclusive minimum-bias

The selection of the inclusive minimum-bias events at the reconstruction level for both data and Monte-Carlo begins by requiring that there is no more than one primary vertex (PV) in a selected event, $\mathrm{N}_{PV,rec} < 2$. This requirement is imposed in order to remove pile-up, the occurrence of more than one pp collision in a single beam interaction. The presence of pile-up constitutes a background, which, although suppressed by the above requirement, is still present and will be discussed in section 5.5.6. Ideally $\mathrm{N}_{PV,rec} = 1$ would be used, however not all primary vertices are reconstructed when an insufficient number of reconstructed tracks is associated with them, as discussed in section 4.2.2. Thus events with $\mathrm{N}_{PV,rec} = 0$ are not discarded at this point so as not to introduce a bias. $\mathrm{N}_{PV,rec}$ in data and MC is shown in Figure 5.3.

Events are required to contain at least one well reconstructed track with $p > 2$ GeV/c in the fiducial acceptance. A candidate track must be an LHCb long track with χ^2/ndof

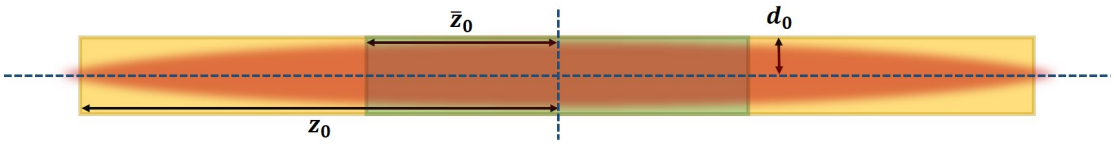

Figure 5.4: A side view of the interaction region at LHCb (red) with the nominal centre line in z and y shown in blue. The gold and green overlays show the geometrical cut acceptance for individual tracks ($d_0 \times z_0$) and the event averaged tracks ($d_0 \times \bar{z}_0$) respectively.

less than 3. It is also required to pass a set of geometrical cuts. The geometrical cuts applied are the distance of the track's first state[1] being no farther from the interaction point than 1 cm in the transverse ($x \times y$) and 25 cm in the longitudinal (z) direction, defined as $d_0 < 10$ mm and $|z_0| < 250$ mm respectively. A visual representation of these geometrical cuts is given in Figure 5.4. These criteria are used to ensure that the track used in the selection of the event originated from the interaction region. A further cut on the average z position of all tracks in the event, $\bar{z}_0 < 100$ mm is applied to ensure that the pp interaction being selected took place in the nominal interaction region inside LHCb.

The selection of events at the MC generator level starts with requiring the presence of exactly one inelastic pp interaction per beam crossing, $\mathrm{N}_{pp,inel}$ per $BX = 1$, essentially selecting only events with exactly one Monte-Carlo primary vertex (MCPV). This requirement fully removes pile-up at the generator level as cutting on the true number of PVs is unambiguous. Furthermore, a presence of at least one charged track with $p > 2$ GeV/c and $\chi^2/\mathrm{ndof} < 3$ within the acceptance is required.

5.4.2 Hard scattering

A selected inclusive minimum-bias event is denoted as a hard scattering event if, in addition to the selection requirements outlined in section 5.4.1, it also contains at least one long track with transverse momentum $p_T > 3$ GeV/c. This requirement ensures that the selected event contains a hard pp scattering component. This, in principle, biases this event class towards higher-than-average particle multiplicity.

Equivalently to the selection at the detector level, to select a hard scattering event at the truth level, the presence of at least one charged track with $p_T > 3$ GeV/c, in $2.0 < \eta < 5.0$ is imposed.

[1]Here the track's first state is defined as its distance of closest approach (DOCA) to the interaction point.

5.4.3 Diffractive and non-diffractive enriched

The selection of these two event classes employs the backwards coverage of the VELO sub-detector of -3.5 $< \eta <$ -1.5. The diffractive and non-diffractive events are defined as inclusive minimum-bias events which contain at least one, $N_{back} > 0$, or exactly zero, $N_{back} = 0$, tracks within -3.5 $< \eta <$ -1.5. Therefore these two sub-classes are fully mutually exclusive, although they are free to overlap with the hard scattering event class. The backwards veto of the diffractive enriched event class is used to look for events with LRGs, as introduced in section 5.1, which are a key property of diffractive events.

The selection of the diffractive and non-diffractive events at the generator level is performed in a similar way to that at the detector level. The presence of at least one or exactly zero charged tracks within -3.5 $< \eta <$ -1.5 is required for the event to be classed as non-diffractive or diffractive, respectively.

To study the fraction of true diffractive events in the diffractive enriched event class, two stand-alone Pythia 8 - LHCb samples are created. For the first sample, the generation of the three diffractive event topologies, shown in Figure 5.2, is enabled, while for the second sample these are disabled. 2,120 and 260 events pass the selection requirements of the diffractive enriched event class from the two samples, respectively, with total cross-sections of 103.4 and 59.4 mb. Each number of events is normalized to the respective simulated cross-section and the two are compared. It can be concluded from this study that, at the generator level, approximately 79% of the events in the diffractive enriched event class are true diffractive events.

The event selection both at the detector and generator levels for all four event classes is summarized in Table 5.2 with the trigger pre-selection requirements included. Additional information on the studies of the event selection cuts can be found in appendix A.

5.4.4 Selection efficiencies

The event selection efficiencies at the detector level are calculated for both the data sample and the Monte-Carlo used in the analysis. This is carried out separately for the samples of the two magnet polarities and the combined sample. These efficiencies are shown in Table 5.3.

First, the percentage of the pile-up suppressed events that are selected as inclusive minimum-bias, N_{incl}/N_{noPU}, is given[2]. The values for N_{incl}/N_{noPU} for the total samples

[2]In various expressions the names of the event classes - inclusive minimum-bias, hard scattering, diffractive enriched and non-diffractive enriched - are contracted for brevity as *incl*, *hard*, *diff* and *ndif*, respectively.

Pre-selection (data only)	Trigger: `Hlt2MBNoBiasLeadingCrossing` BX type: beam-beam	
Event class	Detector level	Generator level
Inclusive minimum-bias	$N_{PV} < 2$ and $N_{long} > 0$ in $2.0 < \eta < 5.0$ having $\chi^2/\text{ndof} < 3$, $p > 2$ GeV/c, $d_0 < 10$ mm, $\lvert z_0 \rvert < 250$ mm, and $\bar{z_0} < 100$ mm	$N_{pp,inel}$ per BX = 1 and $N_{char} > 0$ in $2.0 < \eta < 5.0$ with $p > 2$ GeV/c
Hard scattering	inclusive minimum-bias events with $N_{long} > 0$ having $p_T > 3$ GeV/c in $2.0 < \eta < 5.0$	inclusive minimum-bias events with $N_{char} > 0$ in $2.0 < \eta < 5.0$ with $p_T > 3$ GeV/c
Diffractive enriched	inclusive minimum-bias events with $N_{tracks} = 0$ in -3.5 $< \eta <$ -1.5	inclusive minimum-bias events with $N_{char} = 0$ in -3.5 $< \eta <$ -1.5
Non-diffractive enriched	inclusive minimum-bias events with $N_{tracks} > 0$ in -3.5 $< \eta <$ -1.5	inclusive minimum-bias events with $N_{char} > 0$ in -3.5 $< \eta <$ -1.5

Table 5.2: Summary of the selection requirements for the four event classes. N_{tracks}, N_{long} and N_{char} refer to the number of tracks, number of long tracks and number of charged particles, respectively; other variables are as defined in text. The event classes and their selection requirements are similar to those used in Ref. [135].

Sample	$N_{trigger}$	N_{noPU}	$\frac{N_{incl}}{N_{noPU}}$, %	$\frac{N_{hard}}{N_{incl}}$, %	$\frac{N_{diff}}{N_{incl}}$, %	$\frac{N_{ndif}}{N_{incl}}$, %
MC_{MU}	-	6,470,941	73.71	7.00	8.36	91.64
MC_{MD}	-	6,439,762	73.65	7.00	7.10	92.90
MC_{total}	-	12,910,703	73.68	7.00	7.73	92.27
$Data_{MU}$	237,836,575	194,359,511	45.30	6.79	11.44	88.56
$Data_{MD}$	400,254,187	295,767,653	51.89	6.84	8.97	91.03
$Data_{total}$	638,090,762	490,127,164	49.27	6.82	9.87	90.13

Table 5.3: Detector level event selection efficiencies for Monte-Carlo (top three rows) and data (bottom three rows). $N_{trigger}$ is the number of events passing the imposed `Hlt2MBNoBiasLeadingCrossing` trigger and N_{noPU} is the number of events that pass the $N_{PV} < 2$ requirement. The fourth column shows the selection efficiency of inclusive minimum-bias events from N_{noPU} events and columns five to seven show the selection efficiency of the event sub-classes with respect to the inclusive minimum-bias. The statistical uncertainties are negligible.

are 74% in MC and 49% for real data respectively. Next, the percentage of inclusive minimum-bias events passing the selection of the three sub-classes is shown. N_{hard}/N_{incl} is 7% for both MC and data, while N_{diff}/N_{incl} and N_{ndif}/N_{incl} is 8% and 92% and 10% and 90% for MC and data respectively.

5.5 Measurement

The measurement of the forward particle multiplicity and energy flow is performed over a 2D energy versus pseudorapidity, $(e \times \eta)$, space[3].

In this 2D space the total multiplicity, M_{total}, normalised to the number of pp interactions, N_{pp}, in a given energy and pseudorapidity interval, $(\Delta e, \Delta\eta)$, can be defined as

$$\frac{1}{N_{pp}}\frac{dM_{total}}{d\eta de} = \frac{1}{\Delta\eta\Delta e}\left(\frac{1}{N_{pp}}N_{\eta,e}\right), \tag{5.2}$$

where $N_{\eta,e}$ is the number of particles within the given $(\Delta e, \Delta\eta)$ interval. Similarly, the total energy flow, E_{total}, in the same interval can be defined as

$$\frac{1}{N_{pp}}\frac{dE_{total}}{d\eta de} = \frac{1}{\Delta\eta\Delta e}\left(\frac{1}{N_{pp}}N_{\eta,e}\langle E_{\eta,e}\rangle\right), \tag{5.3}$$

where $\langle E_{\eta,e}\rangle$ is the average energy deposited in the $(\Delta e, \Delta\eta)$ interval and the rest of the variables are unchanged from Equation 5.2.

[3]To determine the energy of each particle the best guess PID, assigned by the particle flow algorithm, is used to determine its mass.

5.5.1 Binning

In an idealised experiment the detector resolution would be perfect and the available statistics would be infinite and thus the optimum choice of binning would tend towards a continuum, i.e. $N_{bins} \to \infty$ bins with a bin width $\Delta bin \to 0$. In a real experiment, however, this is not possible and an appropriate binning scheme, balancing the available statistics and coarseness of the binning, has to be chosen.

High statistics in the available samples allows for a relatively fine binning. In the η (x) direction, 20 uniform bins with a bin width of $\Delta\eta = 0.15$ covering the fiducial acceptance of $2.0 < \eta < 5.0$ are chosen. In the e (y) direction a non-uniform binning scheme is chosen. The need for non-uniformity of binning in energy arises from the orders of magnitude difference between the number of particles with $e \sim 3$ GeV and the ones with $e \sim 100$ GeV and higher. The specific binning scheme for energy is chosen such as to keep the multiplicity distribution relatively uniform. To cover the energy range of $2.0 < e < 1000.0$ GeV, it is therefore split into 10 bins with bin edges (in GeV) of 2.0, 3.0, 4.0, 6.0, 8.0, 11.0, 15.0, 20.0, 30.0, 50.0 and 1000.0. The energy range is chosen to match the fiducial p acceptance.

5.5.2 Methodology

The approach of this analysis is to, first, separate the measurement into two components - charged and neutral. Then, the multiplicity distributions of both components are measured. After that the reconstructed multiplicity distributions are unfolded to account for the detector effects as described in section 5.5.7. Following that the unfolded multiplicity and the average energy are combined as defined by Equation 5.3. Finally, the results in section 5.6 are given as one-dimensional multiplicity and energy flow distributions as a function of η.

5.5.3 Reconstructed multiplicity

In addition to the event selection requirements discussed in section 5.4 all individual particles have further requirements imposed on them at the detector level.

All charged particles in the selected events are required to be of long track type, have ghost probability, GhostProb < 0.3, satisfy a unique VELO ID requirement and be within the fiducial acceptance. Only long type tracks are selected as these are the highest quality tracks at LHCb, with the most hits in the various tracking stations and with the best momentum estimate. The ghost probability cut refers to the probability that the

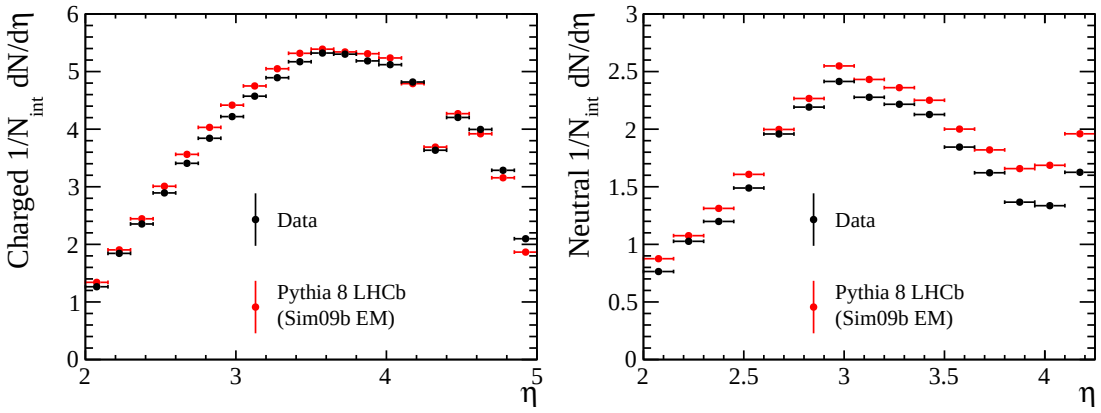

Figure 5.5: Event averaged reconstructed multiplicity for the inclusive minimum-bias event class for the charged (left) and neutral (right) components. The black and red data points denote the reconstructed multiplicity from data and Monte-Carlo respectively.

reconstructed track was made up from spurious hits in the detector and was not left by a real particle traversing the tracking detectors. The calculation of the GhostProb variable is done by a multi-variate analysis based on a neural network [150] and is performed centrally by the LHCb reconstruction software as described in Ref. [151]. The unique VELO ID requirement is imposed by tallying the ID values for the VELO segments the selected long track has traversed. If more than one long track is found with identical hits in VELO only the one with the lowest χ^2/ndof is kept.

For the neutral component, all particles in the selected events are required to be of the `photon` particle flow type and be within the fiducial p acceptance and a reduced fiducial η acceptance of $2.0 < \eta < 4.25$. The reduction of the η range is needed due to the limited acceptance of the calorimetry system at LHCb. Only the `photon` type is selected as other types have either negligible contribution to the neutral multiplicity or have an unacceptably high rate of fake particles.

The event averaged reconstructed multiplicity as a function of η in data and MC is shown in Figure 5.5 for the inclusive minimum-bias event class for both the charged and neutral components. It can be seen that the event-averaged reconstructed charged multiplicity is well modelled across the entire η range, with the exception of the final rapidity bin, where a slightly larger discrepancy is visible. The event-averaged reconstructed neutral multiplicity is more discrepant between MC and data, especially in the bins covering the η range of $3.80 < \eta < 4.25$. The MC and data reconstruction efficiencies are discussed further in section 5.5.5.

In order to estimate the rate of fake particles, as well as to perform the background subtraction and detector unfolding, which will be discussed in further sections, the

Track type	Abundance, % Data	Abundance, % MC	Fake rate, % MC	Selected
Charged track types				
VELO	28.49	-	-	No
Long	47.00	66.59	4.04	Yes
Upstream	13.19	17.81	5.95	No
Downstream	11.32	15.60	28.05	No
Neutral particle types				
Photon	39.67	39.12	$<10^{-2}$	Yes
Merged π^0	0.08	0.08	0	No
Resolved π^0	3.36	3.21	40.70	No
Neutral Hadron	8.80	9.05	N/A	No
Neutral Recovery	7.65	7.52	N/A	No
Composite V0	4.58	4.20	11.22	No
Bad Photon	0.04	0.06	0	No
Isolated Photon	35.83	36.74	N/A	No

Table 5.4: The abundance of reconstructed charged track and neutral particle types in data and Monte-Carlo. The fake rate of the corresponding track and particle types in MC is also shown. The statistical uncertainties are negligible.

Monte-Carlo samples are used. This requires linking between the generator and detector level particles, which is performed using standard LHCb software tools. The rate of fake particles is estimated using MC by examining the fraction of reconstruction level particles with no link to a generator level particle. For some particle types there are no matching algorithms presently available and thus the current methods of estimating the fake rate are not applicable; these particle types are not used in the analysis. The abundance of charged track and neutral particle types in the data and MC samples is given in Table 5.4. The full 2D reconstructed multiplicity distributions for data and MC are provided in appendix C.1.

5.5.4 Generated multiplicity

Contrary to the detector level multiplicity, there are no further cuts applied to the individual particles at the generator level, with the exception of the fiducial acceptance cuts. The abundance of species of particles at the generator level is shown in Table 5.5. The charged component consists mainly of charged kaons and pions, together accounting for 89.3% of generated charged particles, whereas the neutral component is made up

Particle	Abundance, %	Particle	Abundance, %
Charged particle species			
e^-	0.62	e^+	0.62
μ^-	0.02	μ^+	0.02
π^-	38.57	π^+	39.44
K^-	5.55	K^+	5.73
p^-	3.49	p^+	4.73
Σ^-	0.58	Σ^+	0.62
Neutral particle species			
γ	82.81	D^0	$< 10^{-6}$
K_L^0	4.36	K_S^0	4.37
n	3.46	$\bar{n}$	2.82
Λ^0	1.18	$\bar{\Lambda}^0$	1.00

Table 5.5: The abundance of generated particle species in the total PYTHIA 8 Sim09b EM sample. The statistical uncertainties are negligible.

predominantly of photons, accounting for 82.8% of the generated neutral multiplicity.

5.5.5 Reconstruction efficiency

The reconstruction efficiency, ε_{rec}, is the fraction of truth level particles that are reconstructed in the LHCb detector. ε_{rec} can be studied using the Sim09b EM sample by examining the fraction of generator level tracks with a link to a detector level counterpart. The reconstruction efficiency as a function of η is shown in Figure 5.6 for both the charged and the neutral components.

In principle, the process of detector unfolding, discussed in section 5.5.7, accounts for the ε_{rec}. This efficiency, however, differs slightly between the data and the MC sample. As the latter is used to build the response matrix, which, in turn, is used to unfold the former, a correction accounting for this difference has to be applied for both the charged and the neutral components.

To account for the difference in ε_{rec} in the charged component a standard LHCb procedure is used. A table of Data/MC ε_{rec} ratio, R_ε, shown in Figure 5.7, is produced for Sim09b EM and EM data centrally by the collaboration, where ε_{rec} in data is calculated with the tag-and-probe method using muons from the $J/\psi \to \mu^+\mu^-$ decay as described in Ref. [110].

The R_ε values are used to either remove some of the reconstructed MC tracks, when

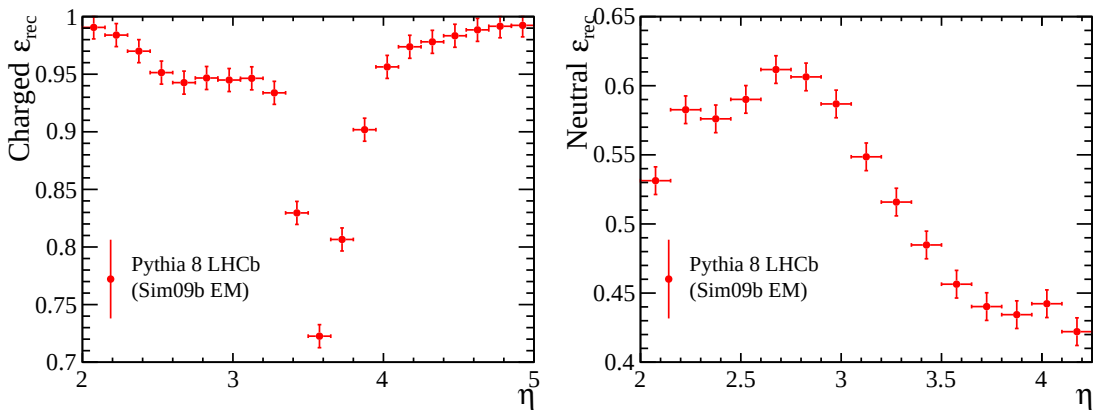

Figure 5.6: Particle reconstruction efficiency for the charged (left) and neutral (right) components. Only the `long` track type for charged component and the `photon` particle flow type for the neutral component are considered.

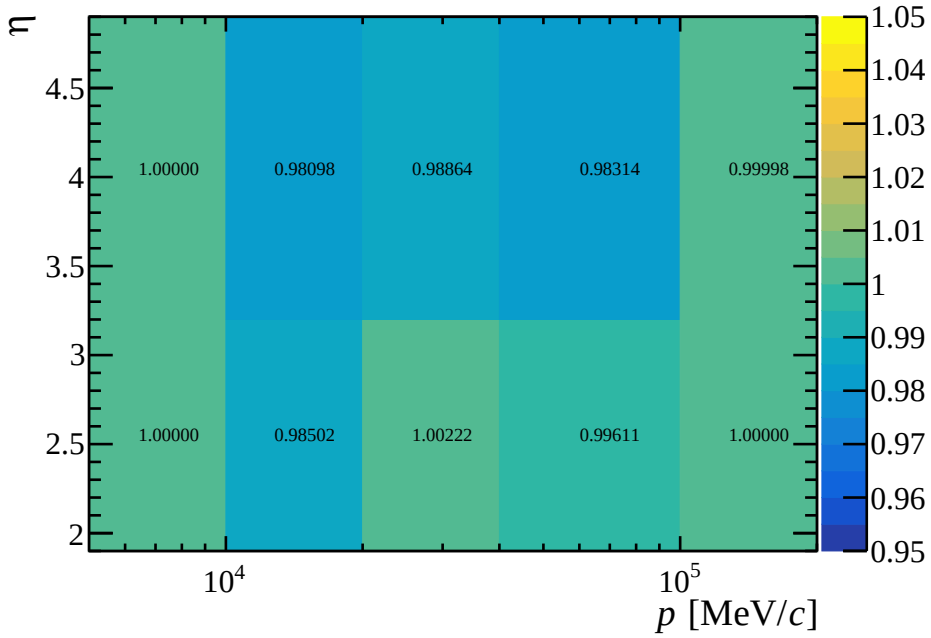

Figure 5.7: Centrally produced Data/MC ε_{rec} table for Sim09b EM MC and EM data samples [152].

$R_\varepsilon < 1$, or to fully reconstruct a certain number of generator level tracks that were not reconstructed, when $R_\varepsilon > 1$. This process shifts the ε_{rec} in MC to match that in the data and is performed during the building of the response matrix as described in section 5.5.7.

Moreover, the statistical and systematic uncertainties on R_ε are used to determine the uncertainty associated with the track reconstruction. This R_ε table, however, does not cover the $2.0 < p < 5.0$ and $200.0 < p < 1000.0$ GeV/c ranges included in the fiducial selection. Furthermore, due to insufficient statistics, R_ε is set to 1.0 for the $5.0 < p < 10.0$ GeV/c bins and the bin covering (2.0 - 3.2 $\times$ 100.0 - 200.0 GeV/c) in the $(\eta \times p)$ space. For all the above bins, therefore, the value of $\mathrm{R}_\varepsilon = 1$ is used with no associated statistical uncertainty [153]. The method of obtaining the tracking reconstruction uncertainty on the final charged multiplicity distribution is discussed in section 5.5.9.

The discrepancy of ε_{rec} between data and MC for the neutral component is less well understood and no standard procedure to account for this in the framework of the analysis presented here currently exists.

A method of examining this discrepancy is the study of the photon and neutral pion reconstruction efficiency using $B^+ \to J/\psi K^{(*)+}$ decays, adopting an approach similar to that in Ref. [154]. This method uses the signal yields of $B^+ \to J/\psi(K^{*+} \to K^+\pi^0)$ and $B^+ \to J/\psi K^+$ decays and the well known ratio of the branching fractions [13] of the two decay channels to determine the neutral ε_{rec} in data and MC.

The studies of the neutral ε_{rec} in data and MC are ongoing and no ε_{rec} correction is applied for the neutral component at this stage. Instead, a systematic uncertainty is assigned to account for the effects on the imperfect knowledge of ε_{rec} as discussed in section 5.5.9.

5.5.6 Purity

The multiplicity and energy flow are measured per pp scattering, which means that any pile-up contribution to the particle production constitutes a background. This background is mainly removed by the pre-selection requirement on the number of reconstructed primary vertices, $\mathrm{N}_{PV,rec} < 2$, introduced in section 5.4.1. However, some impurity persists through residual pile-up.

This pile-up contribution occurs when one or more primary vertices in an event go unreconstructed, which leads to the event passing the $\mathrm{N}_{PV,rec} < 2$ requirement, yet hav-

ing more than one MCPV. The tracks from the second primary vertex thus constitute a background. The pile-up conditions are well modelled in the MC, as indicated earlier by the $N_{PV,rec}$ distributions in data and MC shown in Figure 5.3. Therefore this background can be evaluated and subtracted using the MC sample.

After the event selection the reconstructed PV in the selected events is linked to a MCPV. The vertex linking is performed by counting the number of matched tracks coming from the PV and MCPV. Standard LHCb software tools are used to link the reconstruction level tracks coming from a PV to generator level tracks coming from a MCPV. Where such a link exists, the two tracks are considered matched. The PV and MCPV sharing the most matched tracks are then considered linked. Once a linked pair of vertices is found, the PV and MCPV are further required to be no more than 5 mm apart in the z direction.

For the events containing a single MCPV matched to a single PV no further action is taken. If an event contains more than one MCPV and one of those is matched to the PV, tracks from all unmatched MCPVs are removed as fake tracks. Finally, for events with no reconstructed PVs, one MCPV is selected at random with the tracks from the rest of the MCPVs removed as fake tracks.

The background subtraction is performed during the unfolding process described in section 5.5.7. This background is found to be around 1-2%. Furthermore, the background from the pile-up is also taken as a systematic uncertainty in full as described in section 5.5.9.

5.5.7 Unfolding

In a high-energy physics experiment the reconstructed observable differs from the corresponding true observable due to various detector effects such as the efficiency, purity and acceptance, and needs to be corrected. One way of performing such corrections is through the process of deconvolution or detector unfolding. In the analysis presented here, the reconstructed observables, the measured multiplicity distributions, have to be unfolded to their corresponding true distributions.

Unfolding method

Consider a reconstructed multiplicity distribution, $\mathbf{M}_r$, and a true distribution, $\mathbf{M}_t$. The two distributions can be related via the response matrix, $\mathbf{R}$, through

$$\mathbf{M}_r = \mathbf{R}\mathbf{M}_t. \tag{5.4}$$

The two multiplicity distributions can be obtained using the Monte-Carlo simulation as discussed in the previous sections and therefore $\mathbf{R}$ can be extracted in a straightforward manner. This is not immediately useful, as for the real data only the reconstructed distribution is available. One can, however, relate $\mathbf{M}_r$ to an unfolded distribution, $\mathbf{M}_u$, using the unfolding matrix, $\mathbf{U}$, via

$$\mathbf{M}_u = \mathbf{U}\mathbf{M}_r. \tag{5.5}$$

Therefore, given that the aim of the unfolding procedure is to obtain an unfolded distribution that matches the true distribution, ie. $\mathbf{M}_u \simeq \mathbf{M}_t$, Equations 5.4 and 5.5 can be combined and $\mathbf{U}$ and $\mathbf{R}$ related as

$$\mathbf{U} = \mathbf{R}^{-1}, \tag{5.6}$$

which leads to Equation 5.5 becoming expressible as

$$\mathbf{M}_u = \mathbf{R}^{-1}\mathbf{M}_r. \tag{5.7}$$

Thus the data distribution can be unfolded by using the reconstructed data distribution and an inverse of the response matrix obtained from MC.

There are two immediate issues with this approach. First, this assumes that $\mathbf{R}$ is invertible, which is not a definite *a priori*. Secondly, bins of lower statistics in $\mathbf{R}$, having larger statistical uncertainties and being related to higher bins in $\mathbf{R}^{-1}$, would disproportionally distort the unfolded distribution. It is therefore evident that approaching the unfolding through the inversion of the response matrix is not always feasible.

An alternative approach is to measure $\mathbf{U}$ directly from Monte-Carlo, side-stepping the inversion of $\mathbf{R}$. In a directly measured $\mathbf{U}$, an element of the matrix shows the probability of an entry to be present in an (e, η) bin of the true multiplicity distribution given a corresponding entry in the reconstructed multiplicity distribution.

The method of measuring **U** directly from MC implies an assumption that the shape of the true distribution in MC is known to match the shape of the true distribution in data. This assumption is the underlying principle in the use of the generalized bin-by-bin unfolding approach

$$m_{(i,j),u} = \frac{m_{(i,j),r}}{\varepsilon_{(i,j)}}, \tag{5.8}$$

where $m_{(i,j),u}$ is the $(i,\, j)$th element of $\mathbf{M}_u$, $m_{(i,j),r}$ is the $(i,\, j)$th element of $\mathbf{M}_r$ from data and the correction factor, or efficiency, $\varepsilon_{(i,j)}$, can be expressed as

$$\varepsilon_{(i,j)} = \frac{m_{(i,j),r}}{m_{(i,j),t}}, \tag{5.9}$$

with $m_{(i,j),t}$ and $m_{(i,j),r}$ being the $(i,\, j)$th elements of $\mathbf{M}_t$ and $\mathbf{M}_r$ from MC. The above assumption, however, is not necessarily wholly true. The approach can be improved upon by using further statistical evaluation. In this analysis, the D'Agostini method [155] based on Bayes' theorem is used.

Using Bayesian statistics one can relate the probability of having a certain value of $m_{(i,j),t}$ given a known value of $m_{(i,j),r}$ expressed as $p(m_{(i,j),t}|m_{(i,j),r})$, to the probability of having a certain value of $m_{(i,j),r}$ given a known value of $m_{(i,j),t}$, expressed as $p(m_{(i,j),r}|m_{(i,j),t})$, via [156]

$$p(m_{(i,j),t}|m_{(i,j),r}) = p(m_{(i,j),r}|m_{(i,j),t})\frac{m_{(i,j),t}}{m_{(i,j),r}}. \tag{5.10}$$

The left hand side of the Equation 5.10 is equivalent to an element of **U**, $U_{(i,j)}$, while $p(m_{(i,j),r}|m_{(i,j),t})$ is equivalent to an element of **R**, $R_{(i,j)}$, giving

$$U_{(i,j)} = R_{(i,j)}\frac{m_{(i,j),t}}{m_{(i,j),r}}, \tag{5.11}$$

where $R_{(i,j)}$, $m_{(i,j),t}$ and $m_{(i,j),r}$ can be obtained from MC.

After **U** is built using the analysis MC sample, it is used to deconvolve the $\mathbf{M}_r$ from the desired sample of real data.

It must be noted here that, in principle, the designations of $\mathbf{M}_u$ and $\mathbf{M}_t$ both refer to the truth level multiplicity distributions throughout this section and, if the $\mathbf{M}_r$ upon which **U** is applied comes from the same MC sample is used in building of **U**, the two are indeed identical. The separate designations are used to illustrate that the latter is known directly, while the former is obtained via the unfolding process.

The `RooUnfold` package [157] is used for the unfolding of the multiplicity distri-

bution. The unfolding method which uses the iterative Bayesian approach [158] - `RooUnfoldBayes` - is employed. This unfolding procedure is governed by the regularization parameter **k**, which, for the given method, corresponds to the number of iterations. The regularization parameter is chosen such as to balance the systematic and statistical uncertainties of the unfolding process; the recommended value of $\mathbf{k} = 4$ is used in this analysis. A detailed description of the iterative procedure and the treatment of uncertainties during the unfolding process, and the choice of the value of **k** is given in Refs. [155, 158].

Building of the response matrix

The response matrix is built using the reconstructed and true multiplicity distributions from Monte-Carlo. There are, however, a few differences for these distributions in the response matrix compared to those discussed in sections 5.5.3 and 5.5.4:

- only the events which pass the reconstruction level event cuts are used;
- the truth distribution is background subtracted;
- the Data/MC discrepancy in ε_{rec} for the charged[4] component is taken into account.

The first point is necessary as the unfolding technique used requires the linking of the detector and generator level objects, which can only be done consistently if the same event selection is employed at both levels. The second and third points are necessary to account for the background due to the pile-up and correcting for the observed differences between data and MC in the reconstructed multiplicity for the charged component as discussed in sections 5.5.6 and 5.5.5, respectively.

First, the truth and reconstruction level distributions are simultaneously filled with all linked particles within the fiducial acceptance at the detector level, requiring that these particles originate from a linked PV and MCPV pair. Following that, all remaining truth level particles in the selected events are further added into the generator level distribution as missed particles with the exception of the particles that did not originate from the MCPV linked to the PV, as these constitute a background due to the pile-up and thus are discarded. The reconstruction level particles that originate from a PV which was not linked to a MCPV are also treated as background due to the pile-up and are input to the response matrix as fake particles. Furthermore, all reconstruction

[4]The method of accounting for the discrepancy in ε_{rec} for the neutral component is still under development, as discussed in section 5.5.5 and thus is not applied here at this time.

level particles that did not link to a truth level counterpart or fell outside the fiducial acceptance are input to the response matrix as fake particles.

During the above process the Data/MC ε_{rec} discrepancy is corrected for as discussed in section 5.5.5. A random number generator is used to select a number of reconstructed tracks to be input as missed and a number of unreconstructed generator level tracks to be set as fully reconstructed[5] for $R_\varepsilon < 1$ and $R_\varepsilon > 1$, respectively, effectively equating the ε_{rec} in data and MC.

Two unfolding matrices are built for each event class, one for the charged and one for the neutral component.

Testing the unfolding matrices

In order to test the validity of the unfolding method, an unfolding matrix is built for each event class using only even-numbered events in the total MC sample. This is then used to unfold the reconstructed multiplicity distributions obtained using only odd-numbered events in the total MC sample. Then, the reconstructed multiplicity distributions are unfolded using the corresponding unfolding matrices. The unfolded distributions of the odd-numbered events are then compared to the truth level distributions from the odd-numbered MC events. No significant deviation is observed.

For the use in these "closure tests", the unfolding matrices are built without applying the correction for the reconstruction efficiency described above.

5.5.8 Correcting for the event selection

The application of the detector level event selection for the generator level events when building the response matrix leads to a discrepancy between the unfolded and truth level multiplicity distributions, as the detector level geometric cuts, d_0, $|z_0|$ and $\bar{z_0}$, lead to some truth level events, which would have passed the truth level selection, being discarded. The discrepancy is around 1% for all event classes with the exception of the diffractive enriched event class where it is around 10%.

To correct the above discrepancies, a table of bin-by-bin correction factors is obtained by taking the ratio of the regular generator level multiplicity distribution and the generator level distribution in the response matrix. In order not to introduce any bias, the latter generator level distribution is taken from a response matrix where the ε_{rec} discrepancy correction for the charged component is not performed. These correction

[5] Full reconstruction here implies that a generator level track is perfectly reconstructed at the reconstruction level.

factors are applied to the final unfolded distributions for both the charged and neutral components.

5.5.9 Uncertainties

Reconstruction efficiency

To compute the uncertainty due to ε_{rec} on the charged multiplicity distribution, all reconstructed tracks in data are first weighted by the corresponding R_ε, discussed in section 5.5.5. Then, each R_ε value is shifted by its statistical uncertainty in turn and the resultant table is again used to weight all reconstructed tracks. In addition, the R_ε table also provides a 0.4% overall systematic uncertainty which is used to shift the entire R_ε table and weight all reconstructed tracks again. For tracks where $R_\varepsilon = 1$ and no statistical uncertainty is available, an uncertainty of 5%, coming directly from Monte-Carlo, is used [153]. Furthermore, when the 0.4% systematic shift is made, it is not applied to these bins, as it is already covered by the Monte-Carlo systematic.

After all reconstructed tracks have been weighted by the central and the uncertainty-shifted values of R_ε, the new central reconstructed multiplicity distribution is compared to each uncertainty-shifted distribution in turn. The differences are summed in quadrature and expressed as a percentage error of the reconstructed multiplicity in each bin. To obtain the final uncertainty on the unfolded event-averaged charged multiplicity due to the ε_{rec}, the reconstructed charged multiplicity distribution from data is shifted both up and down by this percentage error. Then, after unfolding the central and the two uncertainty shifted reconstructed multiplicity distributions, the maximum deviation between the central unfolded and either up- or down-shifted unfolded distribution is taken as the final uncertainty in each bin.

This systematic uncertainty cannot yet be rigorously obtained for the neutral component as the tools for comparing the ε_{rec} in data and MC are currently still under development, as discussed in section 5.5.5. For now, to cover the possible effects arising from the imperfect knowledge of the neutral reconstruction efficiency a set overall value of the systematic uncertainty is assigned.

The overall systematic uncertainty due to ε_{rec} for the charged component is found to be 3.3 - 3.4% for all four event classes. For the neutral component a set conservative value of 20% is used as a placeholder to cover for the unknown neutral reconstruction differences between data and MC.

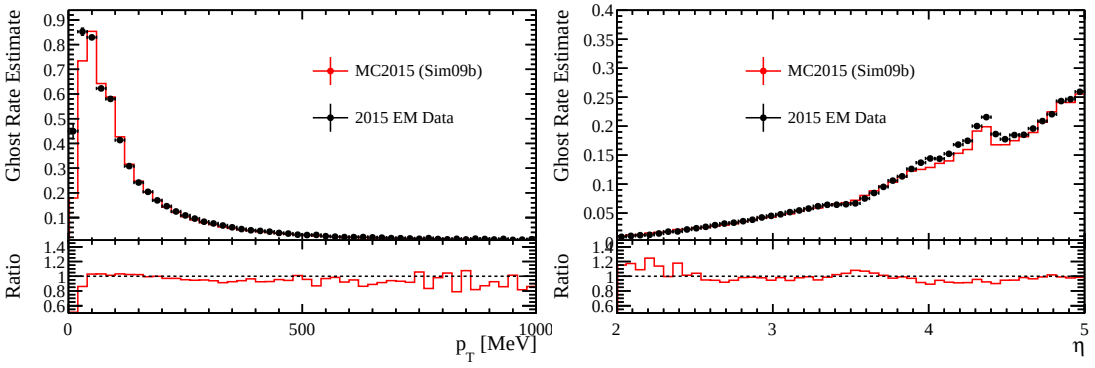

Figure 5.8: The fake rate estimates for the EM data, Sim09b as a function of p_T (left) and η (right) for the charged component [159]. The ratio plots show an approximate overall discrepancy of 6%, this value is thus used in this analysis.

Fake rate

The fake rate is the fraction of mistakenly reconstructed particles and can be estimated using the MC samples by linking the detector and generator level particles. In cases where such a link does not exist the particle is determined to be a fake. The systematic uncertainty arises from an imperfect match of the fake rate in MC and data samples.

For the charged component, the uncertainty of the fake rate is estimated by assuming that the fakes arise from real VELO tracks that are mismatched with track segments in the downstream detectors. The VELO clusters of a single long track from a different event are added into a selected event. The rate at which this track is reconstructed as a good long track is then used as an estimate of the fake rate in both data and MC. The fake rate in data is then compared to the fake rate in the corresponding MC sample. Such studies were carried out as discussed in Ref. [159] and Data/MC comparison plots are shown in Figure 5.8. A percentage uncertainty on the fake rate is estimated using the above studies; a global 6% uncertainty [153] on the fake rate is used. Thus all particles, which are determined to be fakes, as discussed above, are weighed by 6% while filling a reconstructed multiplicity distribution. Following that, the regular reconstructed multiplicity distribution is compared to the one where fake tracks have been weighted and the percentage difference is extracted for each multiplicity bin. Then, like with the reconstruction efficiency uncertainty, the up and down uncertainty shifted reconstructed multiplicity distributions from real data are unfolded and compared to the central unfolded multiplicity distribution.

For the estimation of the fake rate uncertainty for the neutral component, an approach of examining events of empty-empty beam crossing type was explored. Empty-

empty beam crossings are ones where neither beam has a proton bunch traversing an interaction point. In such events any activity in the detector would not be due to an actual particle traversing it and thus, if reconstructed as such, would be labelled as a fake. The fake rate uncertainty could then be determined by taking the average reconstructed multiplicity in the empty-empty events as a fraction of the average number of unlinked neutral particles in the selected events. However, no reconstructed particle in empty-empty events passes the lower bound of the fiducial p acceptance of 2 GeV/c, hence this approach cannot be used.

As described in section 5.5.3, however, only the `photon` particle flow component is selected in this analysis. For this component the fake rate is extremely small, but non-zero. Thus, a conservative approach of using a 100% fake rate uncertainty can be used.

Furthermore, the rate of fake photons due to the bremsstrahlung of the electrons is negligible. The energy deposits from the bremsstrahlung photons are assigned to the respective electrons by the LHCb reconstruction software [71]. A photon emitted after the magnet will deposit its energy in the same calorimeter cluster as the electron, while the photons emitted before the magnet are accounted for by comparing the energy deposit of the electron to its assigned momentum. Moreover, the possible rate of unidentified bremsstrahlung photons is further suppressed by the low electron content in the samples, shown in Table 5.4.

The following steps of obtaining the final uncertainty for the neutral component is the same as for the charged component.

The total fake rate uncertainty for the charged component is determined to be around 0.2% for all four event classes. For the neutral component this uncertainty is negligible.

Particle mis-identification

The `ParticleFlow` algorithm assigns the best-guess PID to all reconstructed particles using the information from the RICH detectors, calorimeters and the muon chambers and, for the charged tracks, combining this information with the momentum estimates from the tracking system. To account for possible mis-identification, different PIDs can be assigned to all selected particles. This is done by checking which PID has been assigned to the particle and replacing the mass of that particle with a different one. For the charged component the most common mis-identification mode is the mis-identification of charged pions and kaons. Therefore all charged particle PIDs, and

subsequently the masses, are changed to a charged pion and a charged kaon in turn.

For the neutral component, where only photons are selected, the most common mis-identification mode is the mis-identification of neutral pions. However, as the entire energy of the particle is deposited in the calorimeters regardless of its species, the mis-identification of a photon as a neutral pion does not alter the neutral multiplicity distributions. The PID uncertainty, therefore, is not applied to the neutral component.

For the charged component, the distribution of the central multiplicity values is compared to the distributions which have been shifted by assigning the different PIDs. The maximum deviation from either the $\pi^{\pm}-$ shifted or $K^{\pm}-$ shifted distributions is used for each bin.

The remaining procedure is equivalent to that employed when determining the previous two systematic uncertainties.

The total uncertainty due to the particle mis-identification on the final unfolded event-averaged charged multiplicity distributions is found to be approximately 1% for all four event classes.

Magnet configuration

As discussed in section 3.2.3, the magnetic field polarity of the dipole magnet at LHCb can be switched between two opposite configurations. Due to the charge asymmetric response of the LHCb detector, this change of the magnetic field configuration induces a systematic uncertainty. To estimate this uncertainty the reconstructed multiplicity distributions for the magnet up and magnet down data samples are extracted separately. The uncertainty on the final multiplicity distribution is obtained by taking the maximum deviation between the two distributions with respect to either of the two samples in each bin.

The following procedure is again identical to the method used for the previous systematic uncertainties. The total systematic uncertainty due to the charge asymmetric response of the LHCb detector is found to be sub-% for both the charged and the neutral component for all event classes with the exception of the diffractive enriched event class where this uncertainty is found to be around 19% for both components.

Unfolding

The process of unfolding has both an associated systematic and a statistical uncertainty.

The systematic uncertainty is the model dependence. This is an uncertainty that

arises from the use of a specific Monte-Carlo generator and tune to model the truth level multiplicity distributions. In order to account for any bias arising from this, the four sets of theoretical predictions are used. First, the truth multiplicity distribution for each of the four theoretical predictions is divided by the truth multiplicity distribution from the analysis MC sample to obtain four sets of weights. Afterwards, these weights are used in turn to weight individual particles, both at the truth and reconstruction level, when creating the multiplicity distributions. These reconstruction level distributions are then unfolded using the response matrix and corrected using the correction weights discussed in sections 5.5.7 and 5.5.8 respectively. The model dependence uncertainty is obtained by comparing the weighted unfolded distributions with their corresponding weighted truth distributions for each set of theoretical predictions and taking the maximum deviation from all models in each bin.

The model dependence uncertainty is found to be around 2% and 8% for the charged and neutral components respectively for all event classes except for the diffractive enriched event class, where this uncertainty is around 3% and 11%.

The total statistical uncertainty is accounted for during the unfolding process. The unfolding package used propagates the statistical uncertainty from the reconstruction level to the unfolded multiplicity distribution as described in Refs. [155, 158]. As the statistics of data samples are of considerable size, the statistical uncertainty both pre- and post-unfolding is negligible compared to other uncertainties considered. For completeness, however, it is included when computing the total uncertainty.

Pile-up

As discussed in section 5.5.6, residual pile-up occurs when there are multiple PVs at the generator level, but one or none are reconstructed. Tracks originating from any unreconstructed PVs in an event that is accepted constitute a background. The effect of the background contribution can be expressed as a fraction using

$$\left(\frac{dM(i,j)}{M(i,j)}\right)_{pile-up} = \frac{M(i,j)_{pile-up} - M(i,j)}{M(i,j)} \cdot \frac{N_{pile-up}}{N_{tot}}, \tag{5.12}$$

where M is the multiplicity in events with $N_{PV} < 2$ and exactly one MCPV (events with no pile-up), $M_{pile-up}$ is the multiplicity in events with $N_{PV} < 2$ and more than one MCPV (events with pile-up), $M_{pile-up}$ is the number of events that have pile-up, N_{tot} is the total number of events, and i and j are indices denoting the bin. This background due to pile-up, although subtracted during the unfolding as described in section 5.5.7,

is applied as a systematic in full.

The systematic uncertainty on the final distributions is obtained by shifting the central reconstructed multiplicity distribution by the obtained pile-up contributions and following the same procedure as used for the reconstruction efficiency uncertainty. This systematic uncertainty is found to be between 1% and 3% for both components in all event classes.

Total uncertainty

As the various systematic uncertainties are uncorrelated, the total systematic uncertainty, Δ_{syst}, is obtained by adding all individual systematic uncertainties in quadrature as given by the Equation 5.13,

$$\Delta_{tot} = \sqrt{\Delta_{rec}^2 + \Delta_{gst}^2 + \Delta_{PID}^2 + \Delta_{mag}^2 + \Delta_{sta}^2 + \Delta_{mod}^2 + \Delta_{pup}^2}, \tag{5.13}$$

where Δ_{rec} is the uncertainty due to the reconstruction efficiency, Δ_{gst} is the uncertainty due to the fake particles, Δ_{PID}[6] is the uncertainty due to particle mis-identification, Δ_{mag} is the uncertainty due to the effects of the change in the magnetic field configuration, Δ_{sta} and Δ_{mod} are the statistical and systematic uncertainties from the unfolding process and Δ_{pup} is the systematic due to the pile-up.

The total contribution from all uncertainties is calculated by shifting the final unfolded event-averaged distributions by the total uncertainty and comparing the total integral of the shifted and central distributions. The uncertainties for both components are summarized in Table 5.6.

[6]This is set to zero for the neutral component.

Uncertainty	Event class Incl.	Hard.	Diff.	Ndif.
Charged component				
Reconstruction efficiency	3.40	3.25	3.39	3.40
Ghost rate	0.21	0.27	0.14	0.21
Particle mis-identification	1.09	1.02	1.07	1.09
Magnet configuration	0.91	0.85	19.47	0.86
Unfolding (syst.)	1.85	2.11	2.77	1.85
Unfolding (stat.)	3×10^{-5}	2×10^{-4}	7×10^{-5}	3×10^{-5}
Pile-up	1.99	1.19	1.72	1.62
Total uncertainty	5.06	4.78	20.68	4.91
Neutral component				
Reconstruction efficiency	20.00	20.00	20.00	20.00
Ghost rate	6×10^{-5}	0.00	8×10^{-4}	4×10^{-5}
Magnet configuration	0.63	0.59	18.56	0.90
Unfolding (syst.)	7.77	8.22	11.50	7.91
Unfolding (stat.)	3×10^{-5}	2×10^{-5}	7×10^{-5}	3×10^{-5}
Pile-up	2.29	1.35	2.45	1.92
Total uncertainty	22.46	22.47	31.22	22.52

Table 5.6: Overall uncertainties in percent on the final unfolded event-averaged multiplicity distributions for each source of uncertainty for both the charged and the neutral component. The overall total uncertainty, obtained by adding each uncertainty in quadrature, is also given.

5.6 Results

The results of the analysis discussed in this thesis are presented as 1D multiplicity[7] and energy flow distributions. The 1D multiplicity distributions are obtained by integrating over all energy bins per η bin. The energy flow distributions are obtained by first applying the Equation 5.3 and then integrating over all energy bins per η bin. These are compared to the four sets of theoretical predictions discussed in section 5.3:

- Pythia 8.212 LHCb + CT09MCS;
- Pythia 8.212 Monash 2013 + NNPDF2.3 LO QED+QCD;
- EPOS LHC;
- Sibyll 2.1.

The charged and neutral multiplicity and energy flow distributions are shown separately in Figures 5.9, 5.10, 5.11 and 5.12. The full 2D unfolded multiplicity distributions for data and the truth level multiplicity distributions from MC are provided in appendix C.2. In addition, the 2D multiplicity distributions for the four sets of theoretical predictions are provided in appendix C.3.

For the charged multiplicity, the best agreement with data for the inclusive minimum-bias and non-diffractive enriched event classes is observed for Pythia 8 LHCb and Sibyll 2.1, while Pythia 8 Monash 2013 and EPOS LHC overshoot the data considerably. For the hard scattering event class a good agreement is observed only with the Pythia 8 LHCb set of predictions, with all of the rest overestimating the multiplicity across the η range. The diffractive multiplicity is equally well described by all sets of theoretical predictions, with the exception of EPOS LHC, which underestimates the multiplicity considerably across the η range. Out of the other three, both tunes of Pythia 8 overestimate and Sibyll 2.1 underestimates the multiplicity within the total uncertainty on the data with Pythia 8 Monash 2013 matching the data best.

For the charged energy flow, Pythia 8 LHCb and Sibyll 2.1 are again in best agreement with the data for the inclusive minimum-bias and the non-diffractive enriched event classes, while Pythia 8 Monash 2013 and EPOS LHC both overestimate the energy flow considerably across the η range. Similarly to the charged multiplicity, Pythia 8

[7] In addition, tables of the two-dimensional event-averaged multiplicity and the total uncertainty associated with it in each bin are given in appendix B for all four event classes for both the charged and the neutral component.

LHCb again describes the data best for the hard scattering event class with the other three sets of predictions overestimating the energy flow. The energy flow distribution for the diffractive enriched event class is well described by both sets of theoretical predictions using Pythia 8, but is underestimated by both cosmic-ray MC generators, with EPOS LHC performing the worst.

For the neutral multiplicity component, the data in the inclusive minimum-bias and non-diffractive enriched event classes is equally well described by all four generators in the lower half of the η range, while the agreement in the η range of $3.5 < \eta < 4.25$ is poor. The same mismatch in high-η is observed for the other two event classes. This, however, is not necessarily due to poor modelling of the theoretical predictions as the discrepancy is at least partially if not fully due to the mismatch in ε_{rec} between the data and the analysis MC sample. The studies of improving the understanding of the Data/MC difference in ε_{rec} and subsequently the results for the neutral component are ongoing as discussed in section 5.5.5. For the hard scattering event class the neutral multiplicity is seemingly underestimated by Pythia 8 LHCb, with the other three sets of predictions being closer to the data. The four sets of predictions diverge considerably for the diffractive enriched event class, with the two tunes of Pythia 8 describing the data quite well, while both Sibyll 2.1 and EPOS LHC underestimate the neutral multiplicity to a differing degree.

For the neutral energy flow, all four sets of theoretical predictions describe the data similarly well for all four event classes, with a tendency of slightly underestimating the energy flow compared to data. The exception is the diffractive enriched class, where both cosmic-ray generators underestimate the energy flow considerably, with EPOS LHC, again, performing the worst. It has to be reiterated that the discrepancy is quite large in the η range of $3.5 < \eta < 4.25$. This enlargement of the disagreement between data and theory is at least partially if not fully due to the poor understanding of ε_{rec} in data and MC, studies of which are ongoing.

Overall, the best agreement with the data is shown by the theoretical predictions generated with Pythia 8 LHCb + CT09MCS, while EPOS LHC is generally performing the worst.

The large difference between the two Pythia tunes can be predominantly attributed to the use of the different PDF sets, however the higher value of $\alpha(M_Z)$ of 0.130 and the earlier MPI cutoff, $p_{T0}^{ref} = 2.28$, in the Monash 2013 tune also contribute. The importance of the latter is also backed by the relatively better agreement between the tunes for the

diffractive enriched event class, where a lesser prominence of MPIs is expected.

The disagreement between EPOS LHC and the other Monte-Carlo predictions, as well as with the unfolded data, showcases possible underdevelopments in the collective flow framework incorporated in the model. The split core-corona approach to string hadronisation is the most probable cause of the considerably overestimated multiplicity in inclusive minimum-bias, hard scattering and non-diffractive enriched event classes. This suggests that the core contribution, in its current implementation, is underestimated, as the expectation is that an increased core-like string hadronisation would decrease the observed multiplicity [43, 55]. This observation is further backed by the considerably underestimated multiplicity for the diffractive enriched event class, as the core density is connected to the prevalence of the MPIs.

Furthermore, although Sibyll 2.1 shows good agreement with the inclusive minimum-bias and non-diffractive enriched event classes, it does fail to describe the diffractive enriched and hard scattering event classes. The considerable discrepancy in the diffractive enriched event class can be attributed to the fact that Sibyll has the least developed description of the diffractive disassociation out of the models studied here. The maximum energy loss by a proton in a diffractive process is limited to 20% and diffraction is not split into different topologies. The overestimation of the multiplicity and, especially, energy flow for the hard scattering event class can be attributed to the requirement of the participation of a valence quark in at least one of the MPIs.

Although the above conclusions can be drawn for the charged component with some certainty, it must be re-stated here that the neutral component is still under investigation and the large error bars are mainly constituted by the blanket 20% uncertainty assigned to the neutral reconstruction efficiency. Nevertheless, the trends observed in the neutral distributions strongly mimic those observed in the charged distributions, which gives some degree of confidence that the above conclusions are valid for the neutral component as well.

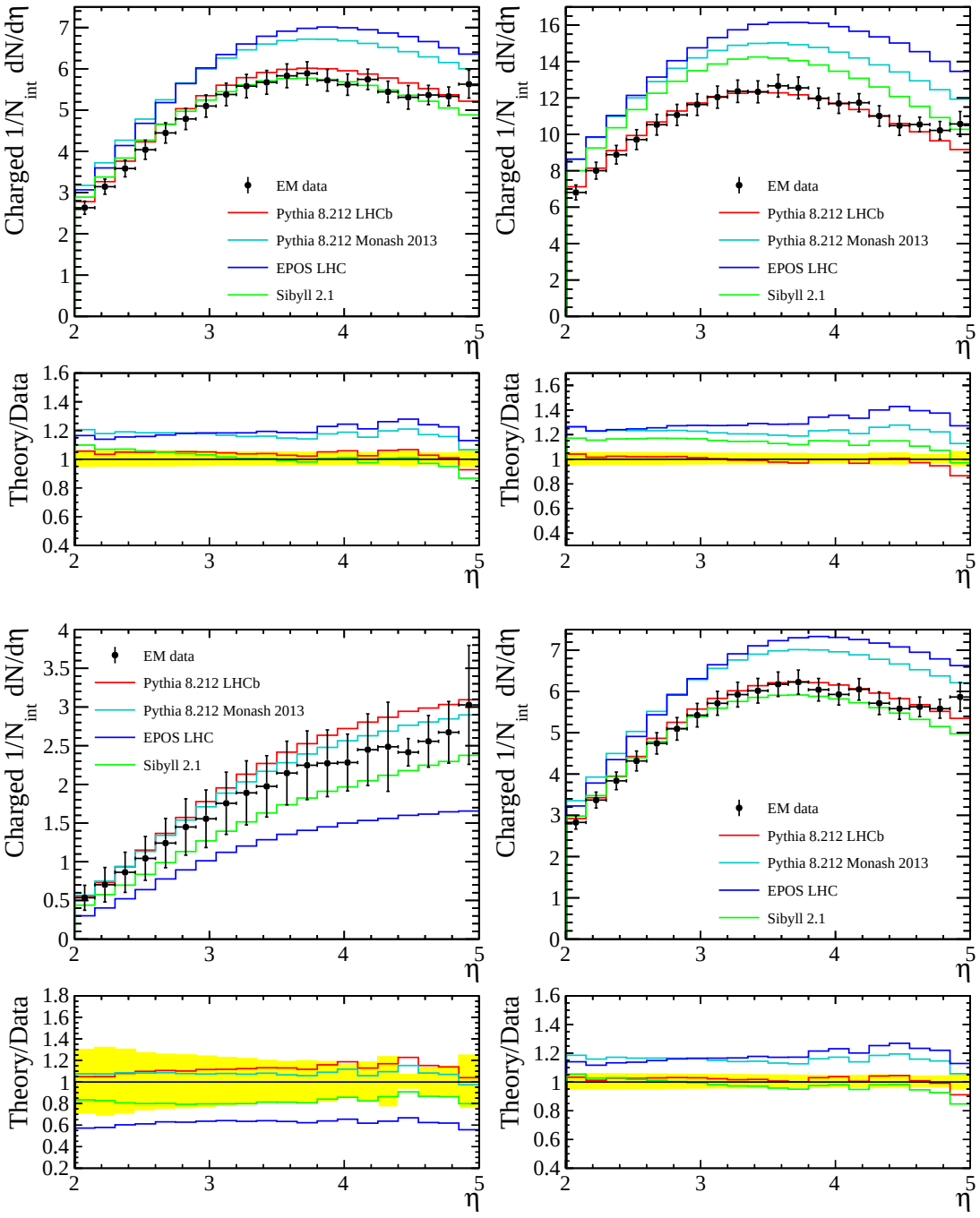

Figure 5.9: Charged event-averaged multiplicity distributions as a function of η for the four event classes. Clockwise from top left: inclusive minimum-bias, hard scattering, non-diffractive enriched and diffractive enriched. Data (black markers), with the error bars showing the total uncertainty, is compared with the theoretical predictions generated using Pythia 8.212 LHCb (red), Pythia 8.212 Monash 2013 (light blue), EPOS LHC (dark blue) and Sibyll 2.1 (green). The theory/data ratio is also shown; the yellow band represents the total uncertainty.

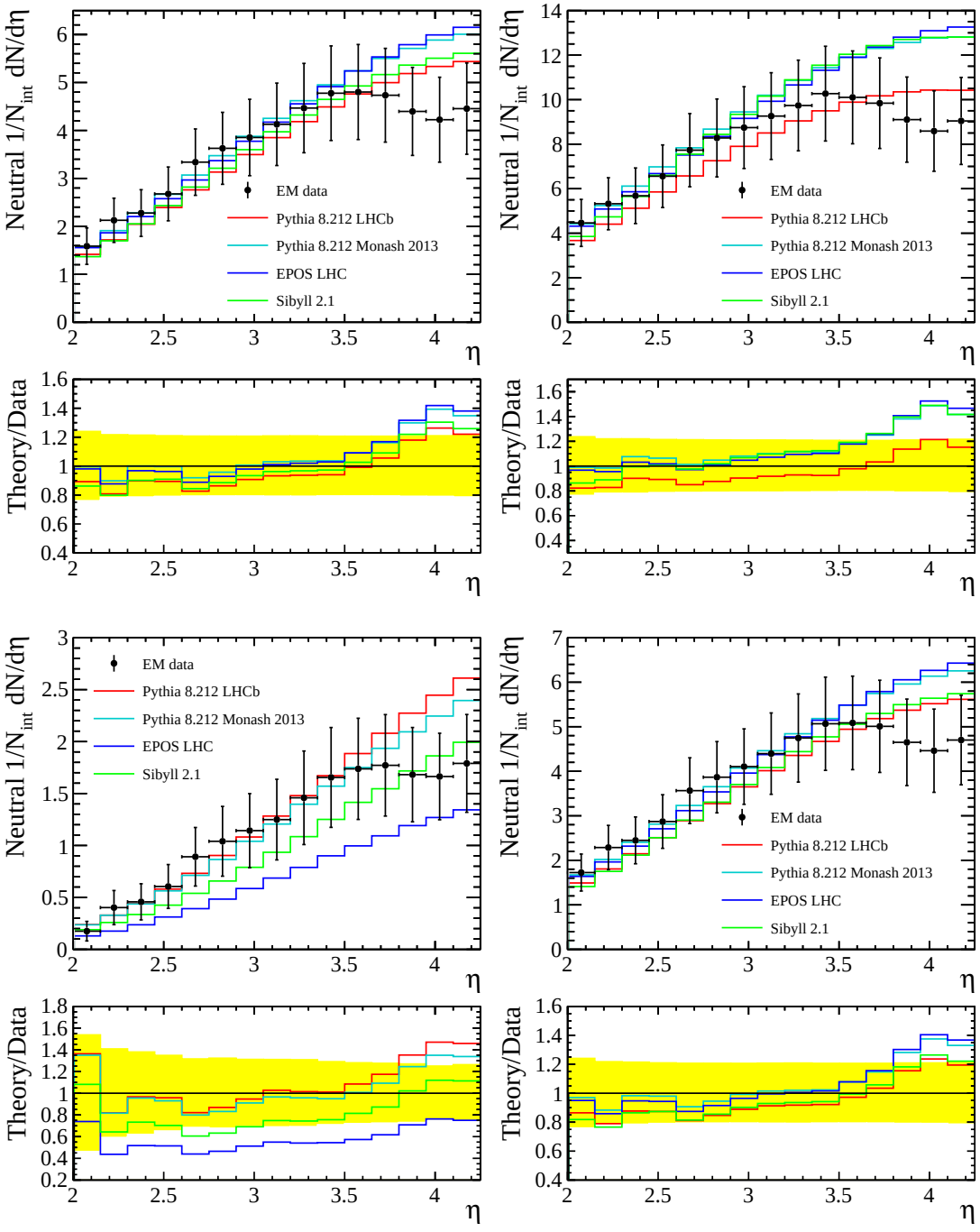

Figure 5.10: Neutral event-averaged multiplicity distributions as a function of η for the four event classes. Clockwise from top left: inclusive minimum-bias, hard scattering, non-diffractive enriched and diffractive enriched. Data (black markers), with the error bars showing the total uncertainty, is compared with the theoretical predictions generated using Pythia 8.212 LHCb (red), Pythia 8.212 Monash 2013 (light blue), EPOS LHC (dark blue) and Sibyll 2.1 (green). The theory/data ratio is also shown; the yellow band represents the total uncertainty.

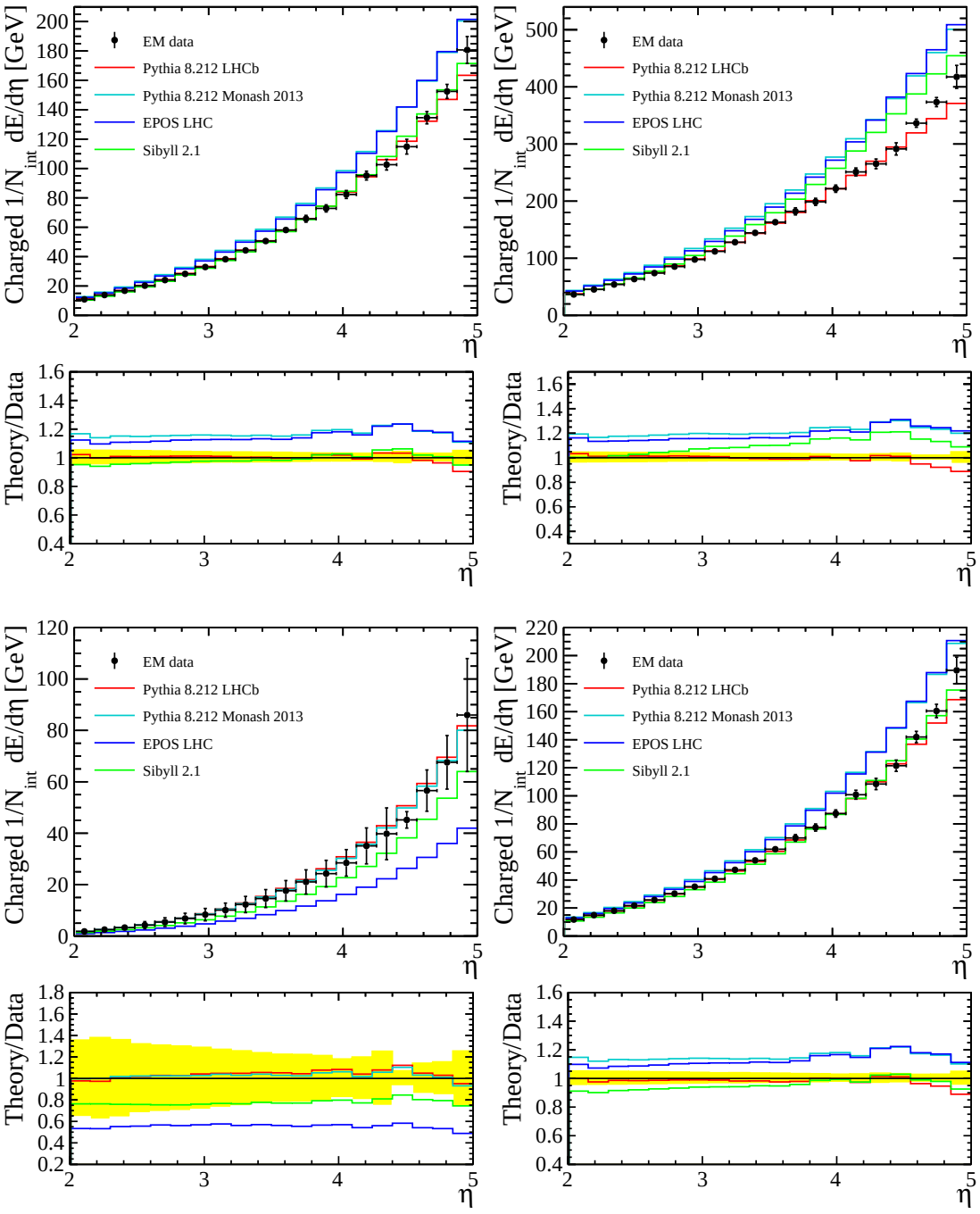

Figure 5.11: Charged event-averaged energy flow distributions as a function of η for the four event classes. Clockwise from top left: inclusive minimum-bias, hard scattering, non-diffractive enriched and diffractive enriched. Data (black markers), with the error bars showing the total uncertainty, is compared with the theoretical predictions generated using Pythia 8.212 LHCb (red), Pythia 8.212 Monash 2013 (light blue), EPOS LHC (dark blue) and Sibyll 2.1 (green). The theory/data ratio is also shown; the yellow band represents the total uncertainty.

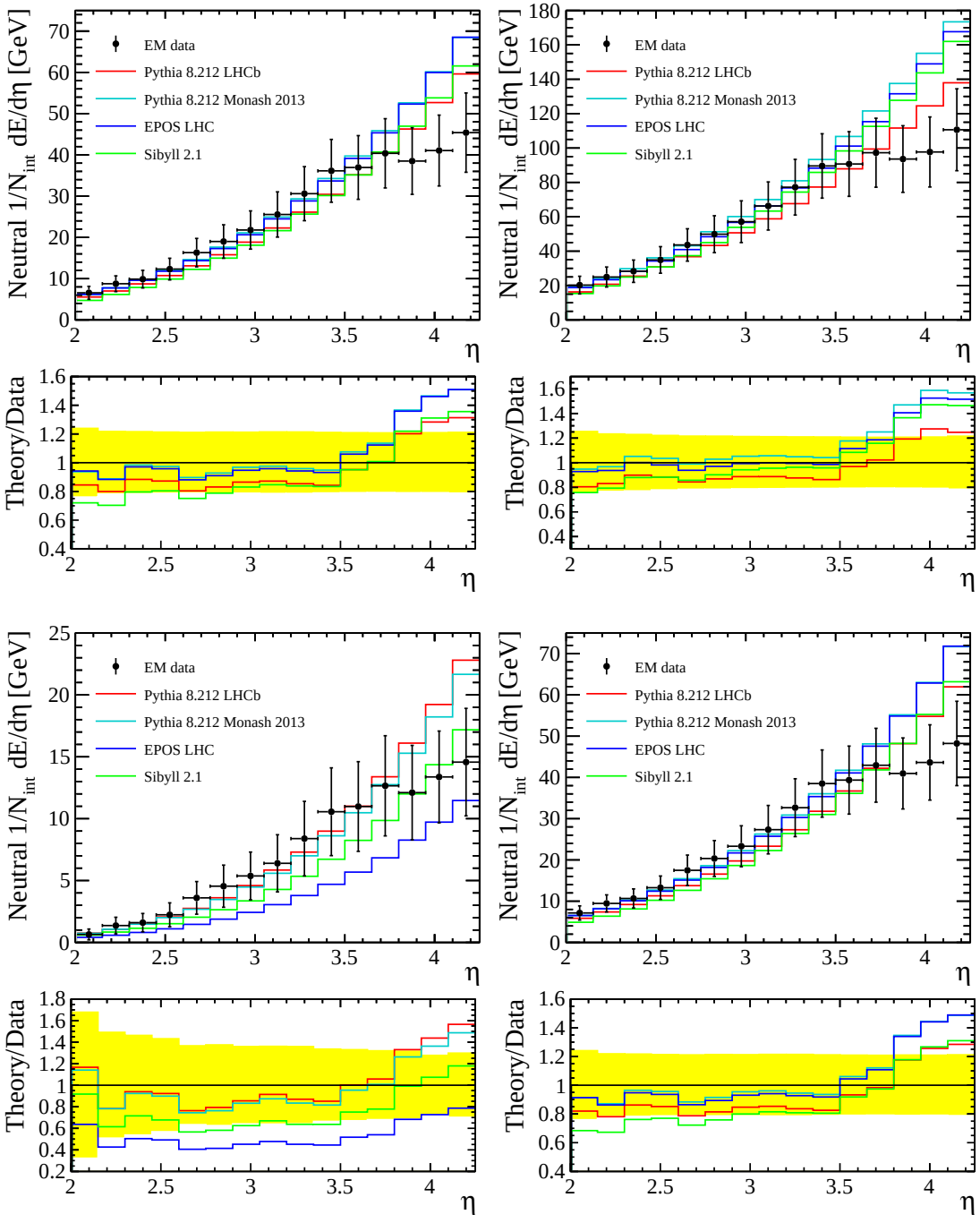

Figure 5.12: Neutral event-averaged energy flow distributions as a function of η for the four event classes. Clockwise from top left: inclusive minimum-bias, hard scattering, non-diffractive enriched and diffractive enriched. Data (black markers), with the error bars showing the total uncertainty, is compared with the theoretical predictions generated using Pythia 8.212 LHCb (red), Pythia 8.212 Monash 2013 (light blue), EPOS LHC (dark blue) and Sibyll 2.1 (green). The theory/data ratio is also shown; the yellow band represents the total uncertainty.

CHAPTER
SIX

Conclusions

"Art is never finished, only abandoned"

- Leonardo DaVinci

This thesis gives an in-depth description of the measurement of particle multiplicity and energy flow in pp collisions at the c.o.m. energy of $\sqrt{s} = 13$ TeV with the LHCb detector. The discussion of the analysis is complemented by an outline of the theoretical basis of particle physics and the experimental environment and tools used to gather and analyse the data used for the measurement. Furthermore, a couple of sections of this thesis give a more detailed insight on a few aspects of the LHCb experiment on which the author has worked extensively outside the scope of the analysis presented in this thesis.

The dataset used for the analysis presented here is no-bias data collected at LHCb during the 'Early Measurements' data taking period in July/August 2015. The data are split into the charged and neutral components, which are further segregated in four event classes - inclusive minimum-bias, hard scattering, diffractive enriched and non-diffractive enriched. The measurement is carried out over a 2D $(e \times \eta)$ space within the fiducial acceptance of $2.0 < p < 1000.0$ GeV/c and $2.0 < \eta < 5.0$ and full detector unfolding is performed.

The results are presented as 1D distributions of charged and neutral multiplicity and energy flow as a function of η and are compared with four sets of theoretical predictions - Pythia 8.212 LHCb, Pythia 8.212 Monash 2013, EPOS LHC and Sibyll 2.1.

It is found that, overall, the Pythia 8.212 LHCb set of theoretical predictions describes the real data best, while the EPOS LHC cosmic-ray MC generator performs, generally, the worst. Furthermore, it is found that, although Pythia 8.212 LHCb de-

scribes the entire inclusive minimum-bias and non-diffractive enriched data well, the description of the diffractive event class and the neutral component of the hard scattering event class is not excellent.

The analysis performed and the results obtained are important for improving our understanding of the hadron-hadron scattering process in full and, more specifically, for deeper understanding of the underlying event and multi-parton interactions. The study of a few aspects of the analysis, notably the understanding of the data and Monte-Carlo differences in the neutral reconstruction efficiency, is still ongoing and should further improve the quality of the results of this measurement.

When fully finalised, the study of particle multiplicity and energy flow presented here should considerably aid in tuning of the Monte-Carlo generators, which, in turn, will improve the ability of the whole community of experimental high-energy particle physics to produce excellent physics results in the future - through the lifespan of the LHC and beyond.

Appendices

APPENDIX

ONE

Event selection

This appendix provides supplementary information on the event selection cuts, discussed in section 5.4.

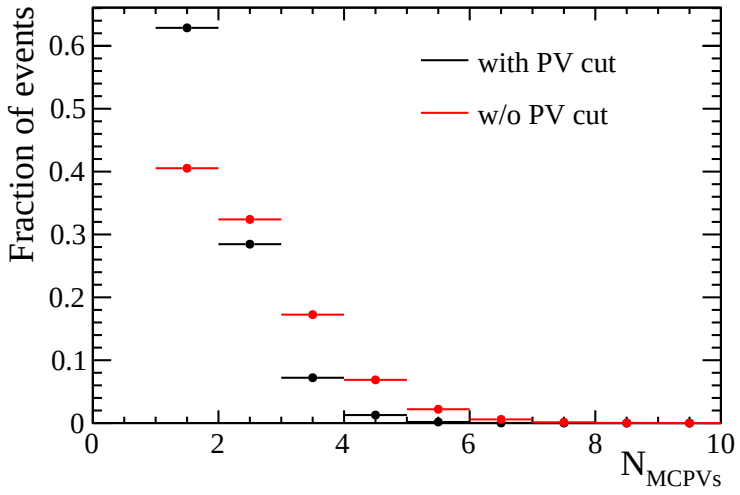

Figure A.1: The effect of the reconstructed primary vertex event selection cut - $N_{PV} < 2$. The number of Monte-Carlo primary vertices with the cut applied (black data points) is compared with the number of MCPVs without the application of this cut (red data points). The two distributions are normalised to unity.

First, in Figure A.1, the number of Monte-Carlo primary vertices is shown with and without the application of the $N_{PV} < 2$ cut. It can be seen that the application of the detector level $N_{PV} < 2$ cut, as one would intuitively assume, does enrich the sample with single MCPV events at the generator level.

Next, the three geometrical cuts used to select the inclusive minimum-bias events - d_0, z_0 and $\bar{z_0}$ - are examined. The two longitudinal cuts, z_0 and $\bar{z_0}$, are standard cuts at LHCb and are purposely designed to be loose. The third cut, d_0, is much tighter. The d_0 cut is optimised to the point where the background (unlinked) tracks begin to dominate over the signal (linked) tracks.

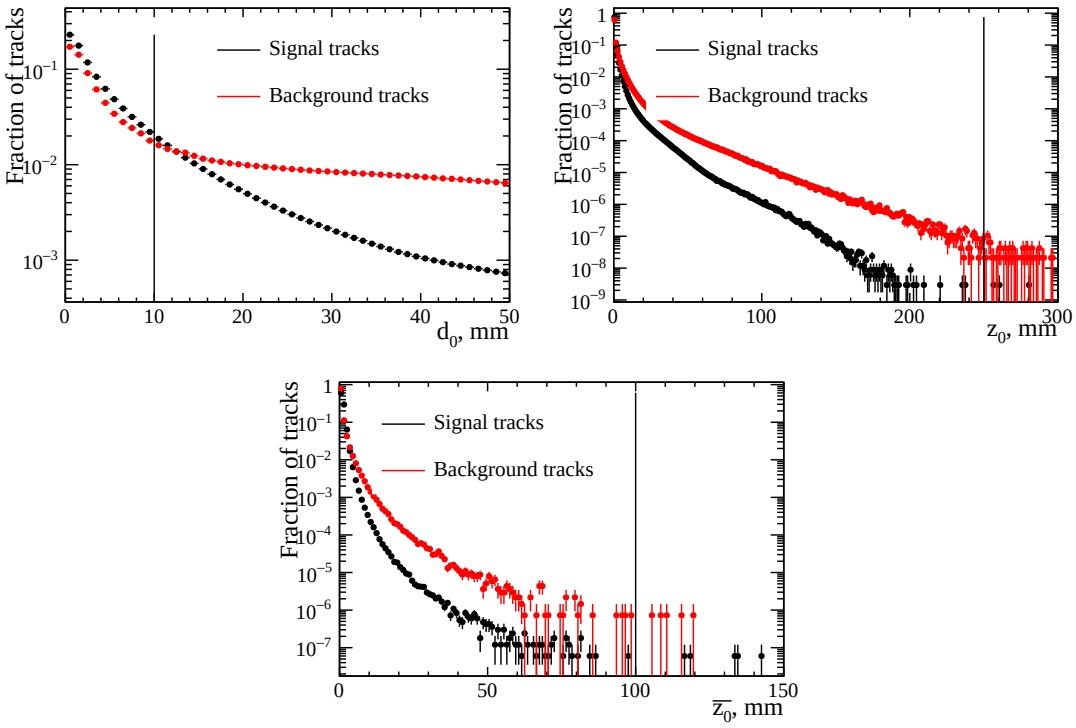

Figure A.2: The geometrical event selection cuts; top: z_0 (left), $\bar{z_0}$ (right); bottom: d_0. The distribution of the corresponding variable is shown for reconstructed tracks with a link to a generator level track (signal) and for tracks without such a link (background). The black vertical line shows the cut on the corresponding variable used in the analysis. All distributions are normalised to unity.

Finally, the selection cut used to separate the diffractive and non-diffractive events, N_{back}, is examined. As discussed in section 5.1.2, a key feature of diffractive events is the presence of the large rapidity gaps. Therefore the cut of $N_{back} = 0$ ($N_{back} > 0$) is used to select the diffractive (non-diffractive) events, where the tracks are counted within the VELO backwards coverage of -3.5 $< \eta <$ -1.5.

In order to test the validity of this cut, a central-exclusive production MC sample is used. The number of backwards tracks is examined in the CEP sample and compared with a sample of the MC used for the analysis, requiring that the number of long tracks in the forward acceptance of $2.0 < \eta < 5.0$, $N_{long} < 5$. The result of this study is shown in Figure A.3.

It can be seen that nearly a 100% of CEP events have exactly zero tracks in the backwards region, whereas the minimum-bias analysis sample contains a wide spectrum of the number of backwards tracks. Thus it is evident that the use of this cut indeed enriches the diffractive and non-diffractive event classes by either selecting or rejecting diffractive events.

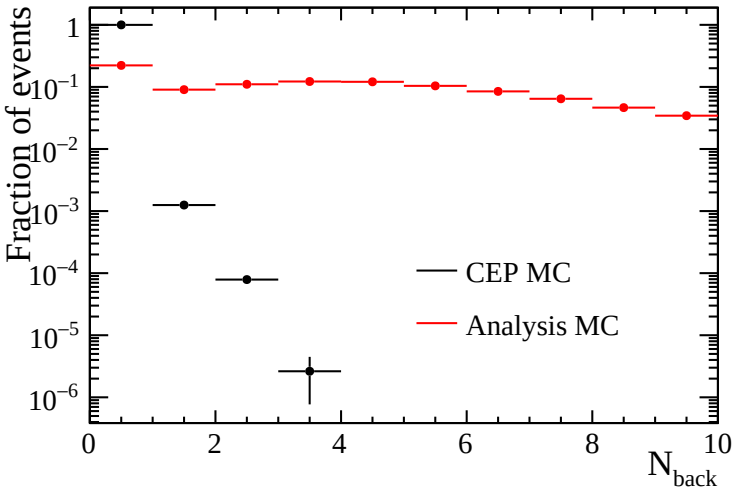

Figure A.3: The number of backwards tracks in a central-exclusive production Monte-Carlo sample (black data points) and the Monte-Carlo sample used in the analysis (red data points) for events with $N_{long} < 5$ in $1.5 < \eta < 3.5$. The two distributions are normalised to unity.

APPENDIX
TWO

Summary of results

In this appendix the final results of the measurement of the event-averaged particle multiplicity are summarised in a tabular form.

The charged event-averaged multiplicity for the inclusive minimum-bias event class is shown in Table B.1 with the total percentage errors on this distribution given in Table B.2. The same is shown for the neutral component of the inclusive minimum-bias event class in Tables B.3 and B.4 respectively. The same information is shown in Tables B.5 to B.8, B.9 to B.12 and B.13 to B.16 for the hard scattering, diffractive enriched and non-diffractive enriched event classes respectively.

The values given in the tables are not normalized to the bin widths in x and y.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.165	0.090	0.084	0.032	0.017	0.007	0.002	0.766×10^{-3}	0.174×10^{-3}	0.028×10^{-3}
2.150 - 2.300	0.181	0.108	0.106	0.043	0.025	0.010	0.004	0.001	0.318×10^{-3}	0.052×10^{-3}
2.300 - 2.450	0.187	0.119	0.124	0.054	0.034	0.015	0.006	0.002	0.543×10^{-3}	0.087×10^{-3}
2.450 - 2.600	0.192	0.128	0.142	0.067	0.045	0.022	0.009	0.004	0.930×10^{-3}	0.148×10^{-3}
2.600 - 2.750	0.189	0.133	0.158	0.081	0.058	0.030	0.013	0.006	0.002	0.256×10^{-3}
2.750 - 2.900	0.180	0.134	0.171	0.094	0.072	0.041	0.019	0.010	0.003	0.436×10^{-3}
2.900 - 3.050	0.168	0.132	0.179	0.105	0.086	0.053	0.027	0.015	0.004	0.738×10^{-3}
3.050 - 3.200	0.153	0.128	0.183	0.115	0.100	0.066	0.036	0.022	0.007	0.001
3.200 - 3.350	0.134	0.120	0.182	0.123	0.113	0.080	0.047	0.032	0.011	0.002
3.350 - 3.500	0.112	0.107	0.175	0.126	0.124	0.093	0.058	0.043	0.017	0.004
3.500 - 3.650	0.095	0.095	0.167	0.127	0.132	0.107	0.071	0.056	0.025	0.006
3.650 - 3.800	0.084	0.083	0.151	0.122	0.134	0.116	0.083	0.071	0.035	0.009
3.800 - 3.950	0.065	0.066	0.129	0.112	0.130	0.121	0.092	0.086	0.047	0.014
3.950 - 4.100	0.046	0.052	0.110	0.103	0.126	0.124	0.101	0.101	0.062	0.022
4.100 - 4.250	0.037	0.043	0.097	0.095	0.124	0.127	0.110	0.120	0.080	0.032
4.250 - 4.400	0.028	0.033	0.077	0.080	0.110	0.119	0.108	0.128	0.094	0.044
4.400 - 4.550	0.021	0.025	0.062	0.067	0.098	0.111	0.106	0.134	0.113	0.061
4.550 - 4.700	0.016	0.018	0.049	0.056	0.088	0.107	0.108	0.146	0.135	0.085
4.700 - 4.850	0.013	0.014	0.037	0.045	0.074	0.097	0.103	0.151	0.154	0.113
4.850 - 5.000	0.011	0.011	0.030	0.039	0.066	0.092	0.103	0.161	0.179	0.153

Table B.1: Charged event-averaged multiplicity for the inclusive minimum-bias event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	6.842	6.564	4.133	5.644	5.019	3.223	3.024	2.298	4.334	7.169
2.150 - 2.300	6.861	6.137	4.152	5.234	4.744	3.305	2.740	2.234	3.589	18.400
2.300 - 2.450	7.060	5.864	4.370	5.236	4.422	3.411	3.036	2.079	3.460	5.957
2.450 - 2.600	7.271	5.859	4.503	5.382	4.146	3.224	3.308	2.372	2.039	5.964
2.600 - 2.750	7.475	5.955	4.424	5.565	4.124	2.477	3.339	2.896	2.564	5.676
2.750 - 2.900	7.695	6.069	4.180	5.691	4.130	1.623	2.853	2.724	1.749	3.907
2.900 - 3.050	7.875	6.053	4.280	5.642	4.288	1.725	2.338	2.900	2.347	4.966
3.050 - 3.200	7.953	6.287	4.321	5.595	4.444	1.827	1.662	2.649	2.296	2.539
3.200 - 3.350	8.044	6.350	4.403	5.594	4.440	3.100	2.594	2.111	2.580	2.751
3.350 - 3.500	8.393	6.545	4.573	5.785	4.282	3.415	2.837	1.672	2.689	2.235
3.500 - 3.650	8.685	6.522	4.892	5.947	4.281	3.533	3.537	1.675	2.687	1.881
3.650 - 3.800	9.145	6.433	4.965	5.984	4.367	3.117	3.516	2.009	1.901	2.039
3.800 - 3.950	9.795	6.527	4.906	6.213	4.445	3.066	3.377	2.527	1.613	1.911
3.950 - 4.100	10.652	6.494	4.808	6.528	4.721	3.436	3.412	2.883	1.651	1.981
4.100 - 4.250	13.556	7.286	4.807	6.429	4.874	3.507	3.296	2.762	2.276	1.554
4.250 - 4.400	14.939	8.969	5.124	6.533	5.367	4.030	3.783	3.292	3.100	2.156
4.400 - 4.550	15.441	8.867	5.191	6.802	6.131	5.339	5.194	4.431	4.205	3.054
4.550 - 4.700	16.092	9.851	5.219	6.252	4.991	4.239	4.068	3.240	3.199	1.836
4.700 - 4.850	19.232	11.988	6.461	6.572	4.725	4.341	4.317	3.527	3.180	2.121
4.850 - 5.000	26.699	14.288	8.886	8.876	6.422	5.390	5.773	5.046	4.653	4.874

Table B.2: Total uncerainty on the charged event-averaged multiplicity in percent for the inclusive minimum-bias event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.106	0.045	0.050	0.020	0.012	0.005	0.001	0.550×10^{-3}	0.114×10^{-3}	0.021×10^{-3}
2.150 - 2.300	0.145	0.065	0.061	0.025	0.015	0.007	0.002	0.851×10^{-3}	0.173×10^{-3}	0.032×10^{-3}
2.300 - 2.450	0.148	0.074	0.058	0.030	0.020	0.009	0.003	0.001	0.274×10^{-3}	0.056×10^{-3}
2.450 - 2.600	0.158	0.092	0.070	0.036	0.026	0.013	0.005	0.002	0.556×10^{-3}	0.097×10^{-3}
2.600 - 2.750	0.172	0.107	0.112	0.047	0.034	0.019	0.008	0.004	0.910×10^{-3}	0.141×10^{-3}
2.750 - 2.900	0.174	0.108	0.127	0.056	0.041	0.024	0.012	0.006	0.002	0.215×10^{-3}
2.900 - 3.050	0.174	0.109	0.127	0.066	0.048	0.030	0.016	0.009	0.003	0.396×10^{-3}
3.050 - 3.200	0.171	0.113	0.133	0.073	0.057	0.037	0.021	0.014	0.004	0.632×10^{-3}
3.200 - 3.350	0.164	0.114	0.141	0.084	0.070	0.046	0.027	0.020	0.007	0.001
3.350 - 3.500	0.154	0.114	0.146	0.092	0.085	0.056	0.034	0.027	0.011	0.002
3.500 - 3.650	0.139	0.106	0.143	0.091	0.091	0.066	0.038	0.032	0.015	0.003
3.650 - 3.800	0.123	0.097	0.139	0.090	0.087	0.072	0.046	0.037	0.019	0.005
3.800 - 3.950	0.106	0.087	0.129	0.083	0.079	0.064	0.046	0.041	0.023	0.006
3.950 - 4.100	0.093	0.078	0.117	0.083	0.078	0.063	0.046	0.044	0.027	0.009
4.100 - 4.250	0.083	0.073	0.118	0.087	0.086	0.068	0.051	0.052	0.036	0.016

Table B.3: Neutral event-averaged multiplicity for the inclusive minimum-bias event class.

η Bin	Energy Bin, GeV									
	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	24.474	23.005	23.988	24.371	21.659	22.701	25.402	37.187	28.052	31.524
2.150 - 2.300	21.645	21.285	21.700	23.206	21.324	20.304	20.513	26.941	23.521	48.412
2.300 - 2.450	20.921	21.143	21.837	23.317	22.154	20.243	20.975	22.016	22.444	27.106
2.450 - 2.600	20.454	20.646	21.452	21.882	22.397	20.274	23.419	20.328	23.174	36.341
2.600 - 2.750	20.298	20.487	20.848	21.154	21.805	20.181	23.539	25.169	20.523	23.747
2.750 - 2.900	20.416	20.218	20.494	20.899	21.315	20.998	21.827	30.681	21.149	27.974
2.900 - 3.050	20.719	20.249	20.362	20.485	20.643	21.109	20.094	29.353	23.273	26.715
3.050 - 3.200	21.302	20.463	20.207	20.318	20.469	20.432	20.137	26.185	31.080	25.404
3.200 - 3.350	21.493	20.819	20.201	20.236	20.328	20.390	20.380	21.292	34.141	22.735
3.350 - 3.500	20.456	21.297	20.381	20.172	20.282	20.489	20.329	20.121	29.600	20.311
3.500 - 3.650	20.317	21.591	20.620	20.209	20.268	20.320	20.399	20.227	24.684	21.494
3.650 - 3.800	21.287	20.225	21.085	20.275	20.171	20.197	20.390	20.407	20.988	26.627
3.800 - 3.950	21.715	20.332	21.697	20.475	20.251	20.230	20.324	20.372	20.068	27.498
3.950 - 4.100	22.150	20.890	20.999	21.123	20.399	20.167	20.306	20.598	20.881	23.653
4.100 - 4.250	22.315	22.049	20.361	22.737	21.175	20.627	20.380	21.029	21.424	21.158

Table B.4: Total uncerainty on the neutral event-averaged multiplicity in percent for the inclusive minimum-bias event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.338	0.208	0.222	0.097	0.061	0.056	0.027	0.010	0.002	0.385×10^{-3}
2.150 - 2.300	0.361	0.239	0.268	0.125	0.083	0.053	0.046	0.018	0.004	0.723×10^{-3}
2.300 - 2.450	0.367	0.254	0.299	0.150	0.107	0.057	0.053	0.031	0.007	0.001
2.450 - 2.600	0.369	0.264	0.326	0.177	0.135	0.077	0.047	0.051	0.013	0.002
2.600 - 2.750	0.359	0.265	0.345	0.200	0.162	0.099	0.051	0.063	0.021	0.004
2.750 - 2.900	0.339	0.260	0.355	0.218	0.187	0.123	0.067	0.064	0.036	0.006
2.900 - 3.050	0.317	0.252	0.360	0.232	0.211	0.148	0.086	0.063	0.056	0.010
3.050 - 3.200	0.290	0.242	0.356	0.242	0.232	0.172	0.107	0.078	0.065	0.017
3.200 - 3.350	0.256	0.224	0.346	0.245	0.247	0.196	0.130	0.104	0.072	0.029
3.350 - 3.500	0.216	0.202	0.327	0.244	0.257	0.215	0.150	0.129	0.076	0.048
3.500 - 3.650	0.187	0.178	0.307	0.237	0.262	0.230	0.171	0.158	0.088	0.065
3.650 - 3.800	0.167	0.157	0.279	0.224	0.254	0.238	0.186	0.184	0.112	0.076
3.800 - 3.950	0.131	0.127	0.238	0.201	0.239	0.234	0.195	0.205	0.138	0.090
3.950 - 4.100	0.096	0.100	0.204	0.185	0.227	0.230	0.204	0.227	0.167	0.111
4.100 - 4.250	0.077	0.084	0.180	0.170	0.218	0.228	0.209	0.251	0.200	0.137
4.250 - 4.400	0.059	0.066	0.144	0.144	0.192	0.207	0.196	0.252	0.217	0.156
4.400 - 4.550	0.045	0.050	0.119	0.121	0.172	0.191	0.182	0.250	0.240	0.198
4.550 - 4.700	0.035	0.035	0.092	0.102	0.154	0.181	0.183	0.257	0.270	0.244
4.700 - 4.850	0.027	0.027	0.070	0.082	0.128	0.166	0.174	0.260	0.290	0.287
4.850 - 5.000	0.022	0.022	0.058	0.072	0.119	0.159	0.171	0.270	0.317	0.344

Table B.5: Charged event-averaged multiplicity for the hard scattering event class.

η Bin	Energy Bin, GeV									
	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	7.019	6.122	4.153	5.830	5.062	2.068	2.098	1.812	4.436	9.453
2.150 - 2.300	7.032	5.810	3.929	5.396	4.978	2.747	2.021	2.287	3.208	17.540
2.300 - 2.450	7.197	5.711	3.956	5.261	4.680	3.505	2.232	2.274	2.841	6.960
2.450 - 2.600	7.428	5.907	4.063	5.335	4.402	3.761	2.865	1.810	2.847	6.072
2.600 - 2.750	7.636	5.951	4.178	5.303	4.073	3.015	3.774	2.022	2.389	5.377
2.750 - 2.900	7.792	5.967	4.180	5.390	3.965	2.221	3.175	2.385	2.188	4.299
2.900 - 3.050	7.886	6.058	4.165	5.443	3.994	2.022	2.786	2.903	2.284	5.130
3.050 - 3.200	8.001	6.087	4.281	5.559	4.031	1.809	2.492	3.142	2.136	3.743
3.200 - 3.350	8.115	5.990	4.382	5.706	4.058	2.668	2.802	2.874	2.269	3.678
3.350 - 3.500	8.905	6.202	4.258	5.741	4.205	2.728	2.817	2.359	2.854	3.398
3.500 - 3.650	9.087	6.437	4.322	5.968	4.280	2.896	2.889	2.149	3.400	3.095
3.650 - 3.800	10.495	6.487	4.377	5.907	4.301	2.842	2.880	2.014	2.889	3.396
3.800 - 3.950	11.723	6.818	4.229	5.880	4.225	2.933	2.682	1.823	2.144	3.329
3.950 - 4.100	12.380	7.110	4.384	5.877	4.534	3.289	3.014	1.789	1.829	2.654
4.100 - 4.250	15.932	8.925	4.756	6.095	4.495	2.997	2.989	1.800	1.806	2.608
4.250 - 4.400	20.677	12.777	6.398	6.743	5.368	3.275	3.516	2.404	2.394	2.943
4.400 - 4.550	18.287	9.178	4.822	6.571	5.648	4.650	4.864	3.854	3.461	2.935
4.550 - 4.700	20.152	9.446	5.231	5.952	4.613	3.715	3.449	2.649	2.019	1.321
4.700 - 4.850	32.605	14.003	6.993	6.796	4.470	4.173	4.030	2.433	2.106	1.430
4.850 - 5.000	41.869	22.197	11.457	9.014	6.184	6.474	6.344	5.101	3.795	4.659

Table B.6: Total uncertainty on the charged event-averaged multiplicity in percent for the hard scattering event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.267	0.129	0.138	0.063	0.042	0.019	0.008	0.004	0.001	0.204×10^{-3}
2.150 - 2.300	0.311	0.159	0.158	0.075	0.053	0.026	0.012	0.006	0.002	0.407×10^{-3}
2.300 - 2.450	0.313	0.172	0.160	0.085	0.064	0.035	0.015	0.009	0.002	0.653×10^{-3}
2.450 - 2.600	0.336	0.199	0.185	0.099	0.079	0.046	0.023	0.013	0.005	0.923×10^{-3}
2.600 - 2.750	0.359	0.224	0.248	0.117	0.095	0.060	0.031	0.019	0.007	0.001
2.750 - 2.900	0.359	0.225	0.269	0.133	0.109	0.072	0.041	0.026	0.009	0.002
2.900 - 3.050	0.359	0.226	0.269	0.148	0.121	0.086	0.051	0.035	0.015	0.003
3.050 - 3.200	0.349	0.231	0.278	0.160	0.138	0.099	0.063	0.049	0.021	0.005
3.200 - 3.350	0.327	0.230	0.286	0.173	0.155	0.114	0.077	0.064	0.030	0.008
3.350 - 3.500	0.304	0.226	0.293	0.188	0.176	0.129	0.090	0.079	0.042	0.012
3.500 - 3.650	0.267	0.207	0.283	0.182	0.185	0.142	0.093	0.089	0.052	0.016
3.650 - 3.800	0.232	0.187	0.272	0.178	0.177	0.148	0.103	0.095	0.061	0.022
3.800 - 3.950	0.199	0.166	0.251	0.163	0.159	0.135	0.101	0.101	0.068	0.026
3.950 - 4.100	0.172	0.146	0.224	0.160	0.154	0.126	0.098	0.100	0.075	0.034
4.100 - 4.250	0.153	0.137	0.224	0.167	0.167	0.136	0.105	0.116	0.094	0.057

Table B.7: Neutral event-averaged multiplicity for the hard scattering event class.

η Bin	Energy Bin, GeV									
	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	21.628	21.862	25.491	27.247	26.063	28.912	31.015	38.761	27.259	42.865
2.150 - 2.300	20.623	20.575	23.031	25.270	24.588	22.560	22.934	24.661	32.140	98.783
2.300 - 2.450	20.389	20.615	22.999	26.376	25.068	21.351	21.372	22.157	24.635	45.547
2.450 - 2.600	20.412	20.374	21.204	24.554	24.789	21.659	20.259	21.025	22.617	39.660
2.600 - 2.750	20.854	20.286	20.685	22.660	24.503	22.273	20.367	20.309	21.400	22.470
2.750 - 2.900	20.860	20.142	20.488	21.502	23.726	23.475	20.619	21.241	21.203	21.755
2.900 - 3.050	21.399	20.287	20.244	20.635	21.904	23.068	21.082	20.718	21.303	23.324
3.050 - 3.200	22.043	20.681	20.124	20.354	20.623	22.054	21.421	20.639	21.707	23.989
3.200 - 3.350	21.696	21.025	20.242	20.214	20.464	21.090	21.721	20.282	21.983	22.337
3.350 - 3.500	20.658	21.468	20.372	20.106	20.374	20.565	21.531	21.224	21.553	21.399
3.500 - 3.650	20.198	21.751	20.569	20.127	20.258	20.340	21.127	21.570	20.347	20.849
3.650 - 3.800	21.527	20.054	21.035	20.148	20.142	20.313	20.516	21.819	20.226	20.781
3.800 - 3.950	22.724	20.164	21.857	20.378	20.206	20.280	20.328	21.301	20.685	20.526
3.950 - 4.100	23.067	20.629	20.951	20.990	20.166	20.078	20.209	21.166	22.253	20.221
4.100 - 4.250	23.198	21.519	20.821	23.007	20.979	20.482	20.339	21.366	23.959	20.298

Table B.8: Total uncerainty on the neutral event-averaged multiplicity in percent for the hard scattering event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.054	0.025	0.019	0.005	0.002	0.811×10^{-3}	0.225×10^{-3}	0.076×10^{-3}	0.021×10^{-3}	0.003×10^{-3}
2.150 - 2.300	0.064	0.033	0.026	0.008	0.004	0.001	0.461×10^{-3}	0.152×10^{-3}	0.026×10^{-3}	0.002×10^{-3}
2.300 - 2.450	0.070	0.040	0.035	0.012	0.006	0.002	0.726×10^{-3}	0.251×10^{-3}	0.066×10^{-3}	0.006×10^{-3}
2.450 - 2.600	0.075	0.046	0.044	0.017	0.009	0.004	0.001	0.518×10^{-3}	0.106×10^{-3}	0.010×10^{-3}
2.600 - 2.750	0.076	0.052	0.054	0.023	0.014	0.006	0.002	0.869×10^{-3}	0.174×10^{-3}	0.022×10^{-3}
2.750 - 2.900	0.075	0.055	0.064	0.030	0.019	0.009	0.003	0.001	0.317×10^{-3}	0.059×10^{-3}
2.900 - 3.050	0.070	0.056	0.072	0.037	0.025	0.013	0.005	0.002	0.587×10^{-3}	0.078×10^{-3}
3.050 - 3.200	0.064	0.056	0.078	0.044	0.033	0.018	0.008	0.004	0.001	0.148×10^{-3}
3.200 - 3.350	0.056	0.053	0.081	0.051	0.042	0.025	0.012	0.007	0.002	0.254×10^{-3}
3.350 - 3.500	0.046	0.048	0.080	0.056	0.050	0.032	0.017	0.010	0.003	0.492×10^{-3}
3.500 - 3.650	0.038	0.042	0.077	0.059	0.057	0.041	0.023	0.015	0.005	0.875×10^{-3}
3.650 - 3.800	0.033	0.036	0.071	0.059	0.063	0.049	0.031	0.021	0.008	0.002
3.800 - 3.950	0.026	0.029	0.061	0.055	0.063	0.055	0.038	0.029	0.012	0.003
3.950 - 4.100	0.018	0.022	0.051	0.050	0.063	0.060	0.045	0.039	0.018	0.004
4.100 - 4.250	0.015	0.019	0.045	0.048	0.064	0.066	0.053	0.051	0.027	0.008
4.250 - 4.400	0.011	0.014	0.036	0.040	0.058	0.064	0.057	0.060	0.036	0.011
4.400 - 4.550	0.008	0.010	0.028	0.032	0.050	0.059	0.056	0.066	0.046	0.018
4.550 - 4.700	0.006	0.007	0.022	0.028	0.046	0.059	0.060	0.079	0.062	0.028
4.700 - 4.850	0.005	0.005	0.017	0.021	0.038	0.054	0.058	0.084	0.076	0.041
4.850 - 5.000	0.004	0.004	0.014	0.019	0.034	0.050	0.059	0.093	0.097	0.061

Table B.9: Charged event-averaged multiplicity for the diffractive enriched event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	22.093	31.056	41.836	56.819	75.123	94.621	99.044	110.406	186.692	137.419
2.150 - 2.300	24.177	30.613	40.846	55.481	72.025	92.108	104.606	113.605	158.117	604.128
2.300 - 2.450	23.280	27.108	35.684	49.920	64.574	81.912	98.090	114.105	106.379	200.787
2.450 - 2.600	21.268	23.340	28.799	39.629	52.828	68.786	82.068	93.449	130.998	229.370
2.600 - 2.750	20.302	21.289	25.872	33.896	45.484	58.462	72.230	97.982	108.443	144.217
2.750 - 2.900	21.487	20.328	22.827	29.381	38.438	50.554	66.930	76.800	104.419	136.271
2.900 - 3.050	21.087	20.036	20.771	25.415	32.511	42.421	56.679	73.007	97.993	207.302
3.050 - 3.200	21.239	19.034	19.028	21.981	27.264	35.056	46.967	62.214	81.304	121.641
3.200 - 3.350	22.276	19.016	17.444	18.992	22.409	29.381	39.841	51.345	70.379	113.593
3.350 - 3.500	21.452	17.407	16.250	16.713	19.276	23.961	31.992	42.866	63.733	89.683
3.500 - 3.650	22.069	17.471	15.253	15.135	16.857	20.536	25.793	36.635	53.236	78.059
3.650 - 3.800	28.646	19.976	16.420	15.421	15.603	17.859	22.498	31.433	43.750	69.673
3.800 - 3.950	32.537	20.946	16.067	15.261	13.967	15.370	19.463	25.617	38.922	62.175
3.950 - 4.100	33.512	19.457	13.848	12.842	11.055	11.820	14.518	19.597	29.655	50.152
4.100 - 4.250	44.070	28.256	19.431	15.820	13.670	13.274	14.730	18.417	27.380	46.681
4.250 - 4.400	52.617	36.456	24.168	21.542	19.217	18.050	19.161	22.268	30.721	48.501
4.400 - 4.550	51.074	27.511	11.990	7.620	4.442	3.749	2.604	2.496	6.787	20.807
4.550 - 4.700	60.082	30.546	19.220	14.049	11.348	9.049	8.092	9.079	13.848	26.424
4.700 - 4.850	80.898	42.855	24.644	17.676	12.179	10.915	9.882	10.179	13.325	24.504
4.850 - 5.000	115.516	64.862	39.710	31.349	24.695	20.755	19.249	19.145	22.162	33.772

Table B.10: Total uncerainty on the charged event-averaged multiplicity in percent for the diffractive enriched event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.019	0.005	0.008	0.003	0.002	0.550×10^{-3}	0.155×10^{-3}	0.073×10^{-3}	0.041×10^{-3}	0.000×10^{-9}
2.150 - 2.300	0.046	0.014	0.013	0.004	0.003	0.830×10^{-3}	0.272×10^{-3}	0.084×10^{-3}	0.019×10^{-3}	0.000×10^{-9}
2.300 - 2.450	0.049	0.019	0.010	0.006	0.004	0.001	0.493×10^{-3}	0.169×10^{-3}	0.035×10^{-3}	0.014×10^{-3}
2.450 - 2.600	0.056	0.030	0.014	0.008	0.005	0.002	0.827×10^{-3}	0.314×10^{-3}	0.057×10^{-3}	0.010×10^{-3}
2.600 - 2.750	0.063	0.040	0.037	0.013	0.008	0.003	0.001	0.518×10^{-3}	0.098×10^{-3}	0.000×10^{-9}
2.750 - 2.900	0.065	0.040	0.048	0.017	0.011	0.005	0.002	0.931×10^{-3}	0.146×10^{-3}	0.022×10^{-3}
2.900 - 3.050	0.066	0.041	0.048	0.023	0.014	0.007	0.003	0.001	0.368×10^{-3}	0.042×10^{-3}
3.050 - 3.200	0.066	0.043	0.049	0.027	0.018	0.010	0.005	0.002	0.538×10^{-3}	0.076×10^{-3}
3.200 - 3.350	0.066	0.044	0.056	0.033	0.026	0.014	0.007	0.004	0.001	0.135×10^{-3}
3.350 - 3.500	0.065	0.047	0.060	0.038	0.034	0.020	0.010	0.007	0.002	0.195×10^{-3}
3.500 - 3.650	0.061	0.046	0.060	0.038	0.038	0.026	0.012	0.009	0.003	0.553×10^{-3}
3.650 - 3.800	0.054	0.042	0.059	0.038	0.037	0.029	0.017	0.011	0.004	0.752×10^{-3}
3.800 - 3.950	0.048	0.038	0.055	0.035	0.033	0.027	0.018	0.014	0.006	0.992×10^{-3}
3.950 - 4.100	0.043	0.034	0.051	0.036	0.033	0.026	0.019	0.016	0.007	0.002
4.100 - 4.250	0.040	0.032	0.053	0.039	0.037	0.028	0.021	0.019	0.011	0.004

Table B.11: Neutral event-averaged multiplicity for the diffractive enriched event class.

η Bin	Energy Bin, GeV									
	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	33.295	55.004	78.667	99.400	98.444	153.560	144.267	181.542	165.542	200.998
2.150 - 2.300	31.591	35.882	60.828	76.152	88.580	109.620	140.623	153.029	89.546	200.998
2.300 - 2.450	31.471	31.201	53.368	66.801	81.981	57.333	130.851	136.869	295.853	157.022
2.450 - 2.600	30.060	28.338	36.898	56.513	73.118	84.026	149.120	101.244	173.166	410.747
2.600 - 2.750	28.882	27.749	28.091	44.544	55.972	67.365	75.478	141.141	109.831	200.998
2.750 - 2.900	29.496	28.766	28.679	38.023	50.840	63.058	79.065	110.515	103.446	245.414
2.900 - 3.050	28.570	27.706	28.660	33.537	41.926	50.612	65.505	86.989	105.926	91.129
3.050 - 3.200	28.324	27.529	28.173	31.196	39.085	49.120	60.928	71.225	100.485	114.353
3.200 - 3.350	28.415	28.561	27.008	29.191	34.516	44.831	51.350	64.141	87.588	156.214
3.350 - 3.500	25.317	26.851	26.678	27.247	31.493	38.079	42.097	53.963	74.921	96.591
3.500 - 3.650	24.760	25.977	25.602	26.817	28.739	32.302	40.862	46.713	66.042	77.175
3.650 - 3.800	25.945	24.135	25.272	25.696	26.844	30.104	35.713	45.085	51.428	77.526
3.800 - 3.950	26.372	25.114	24.073	23.877	26.123	28.810	31.146	39.254	52.486	65.927
3.950 - 4.100	27.200	25.348	23.076	22.577	22.069	23.041	26.604	29.538	39.024	70.693
4.100 - 4.250	29.125	27.535	23.307	22.894	23.500	23.887	28.197	31.839	37.722	49.956

Table B.12: Total uncertainty on the neutral event-averaged multiplicity in percent for the diffractive enriched event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.177	0.098	0.092	0.035	0.018	0.007	0.002	0.838×10^{-3}	0.189×10^{-3}	0.031×10^{-3}
2.150 - 2.300	0.193	0.116	0.115	0.046	0.027	0.011	0.004	0.002	0.349×10^{-3}	0.057×10^{-3}
2.300 - 2.450	0.200	0.128	0.134	0.059	0.037	0.017	0.006	0.003	0.597×10^{-3}	0.096×10^{-3}
2.450 - 2.600	0.205	0.137	0.153	0.073	0.049	0.024	0.010	0.004	0.001	0.163×10^{-3}
2.600 - 2.750	0.201	0.142	0.170	0.087	0.063	0.033	0.014	0.007	0.002	0.282×10^{-3}
2.750 - 2.900	0.192	0.143	0.182	0.101	0.077	0.044	0.021	0.011	0.003	0.478×10^{-3}
2.900 - 3.050	0.178	0.141	0.191	0.113	0.093	0.057	0.029	0.016	0.005	0.810×10^{-3}
3.050 - 3.200	0.162	0.136	0.195	0.123	0.107	0.071	0.039	0.024	0.008	0.001
3.200 - 3.350	0.143	0.127	0.193	0.130	0.121	0.086	0.050	0.034	0.012	0.002
3.350 - 3.500	0.119	0.114	0.186	0.133	0.132	0.100	0.063	0.046	0.019	0.004
3.500 - 3.650	0.101	0.101	0.176	0.134	0.141	0.114	0.076	0.061	0.027	0.006
3.650 - 3.800	0.089	0.088	0.160	0.129	0.142	0.124	0.088	0.076	0.038	0.010
3.800 - 3.950	0.069	0.070	0.137	0.118	0.137	0.128	0.098	0.092	0.051	0.016
3.950 - 4.100	0.050	0.055	0.117	0.109	0.133	0.131	0.108	0.108	0.067	0.024
4.100 - 4.250	0.039	0.046	0.102	0.100	0.130	0.134	0.117	0.127	0.086	0.035
4.250 - 4.400	0.030	0.035	0.081	0.084	0.115	0.125	0.114	0.135	0.100	0.047
4.400 - 4.550	0.023	0.027	0.066	0.071	0.104	0.117	0.111	0.142	0.120	0.066
4.550 - 4.700	0.017	0.019	0.052	0.059	0.093	0.113	0.113	0.153	0.143	0.092
4.700 - 4.850	0.014	0.014	0.039	0.047	0.077	0.102	0.108	0.158	0.162	0.121
4.850 - 5.000	0.011	0.012	0.032	0.041	0.070	0.096	0.108	0.168	0.188	0.162

Table B.13: Charged event-averaged multiplicity for the non-diffractive enriched event class.

η Bin	Energy Bin, GeV									
	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	6.525	6.297	3.839	5.489	4.910	2.820	2.966	1.527	4.130	7.889
2.150 - 2.300	6.711	6.019	3.989	5.258	4.831	3.249	2.849	1.770	3.346	18.756
2.300 - 2.450	6.961	5.721	4.233	5.246	4.421	3.439	3.050	2.359	3.304	6.278
2.450 - 2.600	7.132	5.692	4.312	5.289	4.116	3.154	3.200	2.100	2.692	6.203
2.600 - 2.750	7.379	5.814	4.248	5.481	4.041	2.420	3.420	2.715	2.382	5.678
2.750 - 2.900	7.639	6.004	4.084	5.624	4.103	1.875	3.110	2.906	2.171	3.697
2.900 - 3.050	7.788	5.989	4.187	5.593	4.236	1.771	2.345	2.940	2.420	4.520
3.050 - 3.200	7.898	6.200	4.215	5.521	4.379	1.823	1.744	2.728	2.320	2.843
3.200 - 3.350	8.032	6.270	4.280	5.519	4.351	3.024	2.688	2.341	2.748	3.106
3.350 - 3.500	8.299	6.420	4.401	5.666	4.134	3.264	2.745	1.768	2.856	2.616
3.500 - 3.650	8.597	6.355	4.689	5.787	4.103	3.340	3.347	1.658	2.697	2.235
3.650 - 3.800	9.204	6.371	4.789	5.855	4.220	2.953	3.387	1.883	1.969	2.070
3.800 - 3.950	9.863	6.541	4.765	6.120	4.321	2.920	3.274	2.455	1.778	2.075
3.950 - 4.100	10.824	6.362	4.517	6.297	4.443	3.067	3.082	2.567	1.406	1.716
4.100 - 4.250	13.977	7.400	4.827	6.377	4.792	3.367	3.220	2.718	2.320	1.881
4.250 - 4.400	15.628	9.151	5.058	6.457	5.208	3.940	3.755	3.340	3.351	2.727
4.400 - 4.550	16.114	9.058	4.745	6.197	5.372	4.448	4.351	3.424	3.129	1.660
4.550 - 4.700	16.968	10.240	5.399	6.185	4.804	4.000	3.784	2.874	2.924	1.633
4.700 - 4.850	19.644	12.662	6.765	6.460	4.571	4.107	4.111	3.333	2.930	2.041
4.850 - 5.000	26.915	14.998	9.193	8.879	6.264	5.255	5.554	4.804	4.455	4.901

Table B.14: Total uncertainty on the charged event-averaged multiplicity in percent for the non-diffractive enriched event class.

	Energy Bin, GeV									
η Bin	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	0.116	0.050	0.054	0.022	0.013	0.005	0.002	0.604×10^{-3}	0.125×10^{-3}	0.023×10^{-3}
2.150 - 2.300	0.156	0.071	0.066	0.028	0.017	0.007	0.003	0.936×10^{-3}	0.190×10^{-3}	0.035×10^{-3}
2.300 - 2.450	0.158	0.080	0.064	0.032	0.022	0.010	0.004	0.001	0.300×10^{-3}	0.061×10^{-3}
2.450 - 2.600	0.170	0.099	0.077	0.039	0.028	0.015	0.006	0.003	0.610×10^{-3}	0.107×10^{-3}
2.600 - 2.750	0.184	0.115	0.120	0.051	0.037	0.020	0.009	0.004	0.998×10^{-3}	0.156×10^{-3}
2.750 - 2.900	0.186	0.115	0.135	0.060	0.044	0.026	0.013	0.007	0.002	0.235×10^{-3}
2.900 - 3.050	0.186	0.117	0.136	0.070	0.052	0.032	0.017	0.010	0.003	0.434×10^{-3}
3.050 - 3.200	0.182	0.120	0.142	0.078	0.062	0.040	0.023	0.015	0.005	0.693×10^{-3}
3.200 - 3.350	0.175	0.121	0.150	0.089	0.075	0.050	0.029	0.021	0.008	0.001
3.350 - 3.500	0.164	0.121	0.155	0.098	0.091	0.060	0.037	0.029	0.012	0.002
3.500 - 3.650	0.147	0.113	0.152	0.097	0.096	0.071	0.041	0.035	0.016	0.004
3.650 - 3.800	0.130	0.103	0.147	0.095	0.093	0.076	0.049	0.040	0.021	0.005
3.800 - 3.950	0.112	0.092	0.137	0.088	0.084	0.069	0.049	0.044	0.025	0.007
3.950 - 4.100	0.098	0.082	0.124	0.088	0.083	0.067	0.049	0.047	0.029	0.010
4.100 - 4.250	0.088	0.077	0.124	0.092	0.091	0.072	0.054	0.056	0.039	0.018

Table B.15: Neutral event-averaged multiplicity for the non-diffractive enriched event class.

η Bin	Energy Bin, GeV									
	2 - 3	3 - 4	4 - 6	6 - 8	8 - 11	11 - 15	15 - 20	20 - 30	30 - 50	50 - 1000
2.000 - 2.150	24.603	22.910	24.200	24.602	21.907	22.850	25.680	37.720	28.384	29.956
2.150 - 2.300	21.715	21.278	21.832	23.341	21.498	20.414	20.692	27.061	23.869	47.825
2.300 - 2.450	20.975	21.208	21.823	23.545	22.352	20.271	21.075	22.150	22.629	28.218
2.450 - 2.600	20.419	20.621	21.342	22.070	22.586	20.367	23.490	20.469	23.258	36.999
2.600 - 2.750	20.195	20.436	20.808	21.155	21.968	20.295	23.722	25.198	20.596	23.494
2.750 - 2.900	20.332	20.185	20.509	20.893	21.490	21.127	21.956	30.927	21.450	28.568
2.900 - 3.050	20.609	20.165	20.335	20.516	20.673	21.233	20.146	29.494	23.452	27.145
3.050 - 3.200	21.192	20.377	20.160	20.330	20.483	20.538	20.233	26.273	31.309	25.366
3.200 - 3.350	21.432	20.761	20.186	20.265	20.376	20.465	20.507	21.403	34.355	22.944
3.350 - 3.500	20.416	21.278	20.351	20.176	20.329	20.552	20.488	20.283	29.822	20.390
3.500 - 3.650	20.276	21.589	20.606	20.196	20.243	20.364	20.468	20.344	24.794	21.673
3.650 - 3.800	21.269	20.199	21.101	20.267	20.168	20.218	20.454	20.514	21.119	26.905
3.800 - 3.950	21.718	20.294	21.763	20.474	20.235	20.238	20.358	20.436	20.123	27.772
3.950 - 4.100	22.138	20.832	21.005	21.129	20.344	20.119	20.257	20.581	20.904	23.744
4.100 - 4.250	22.258	21.973	20.331	22.904	21.250	20.617	20.409	21.004	21.445	21.153

Table B.16: Total uncertainty on the neutral event-averaged multiplicity in percent for the non-diffractive enriched event class.

APPENDIX
THREE

Two dimensional multiplicity distributions

In this appendix the full two-dimensional event-averaged particle multiplicity distributions are shown.

In C.1, Figures C.1 and C.2 show the charged and neutral reconstructed multiplicity distributions, respectively. The data and MC reconstructed multiplicity is shown side-by-side for comparison for all four event classes.

In C.2, Figures C.3 and C.4 show the charged and neutral unfolded and truth level multiplicity distributions, respectively. The unfolded data and truth level MC are shown side-by-side for comparison for all four event classes.

In C.3, Figures C.5 to C.12 show the charged and neutral multiplicity distributions from the four sets of theoretical predictions for the four event classes. The figures are grouped such as to show each model's predicted multiplicity for a corresponding component and event class.

All bins in all distributions are normalized by their area in the $(e \times \eta)$ space.

C.1 Reconstructed multiplicity

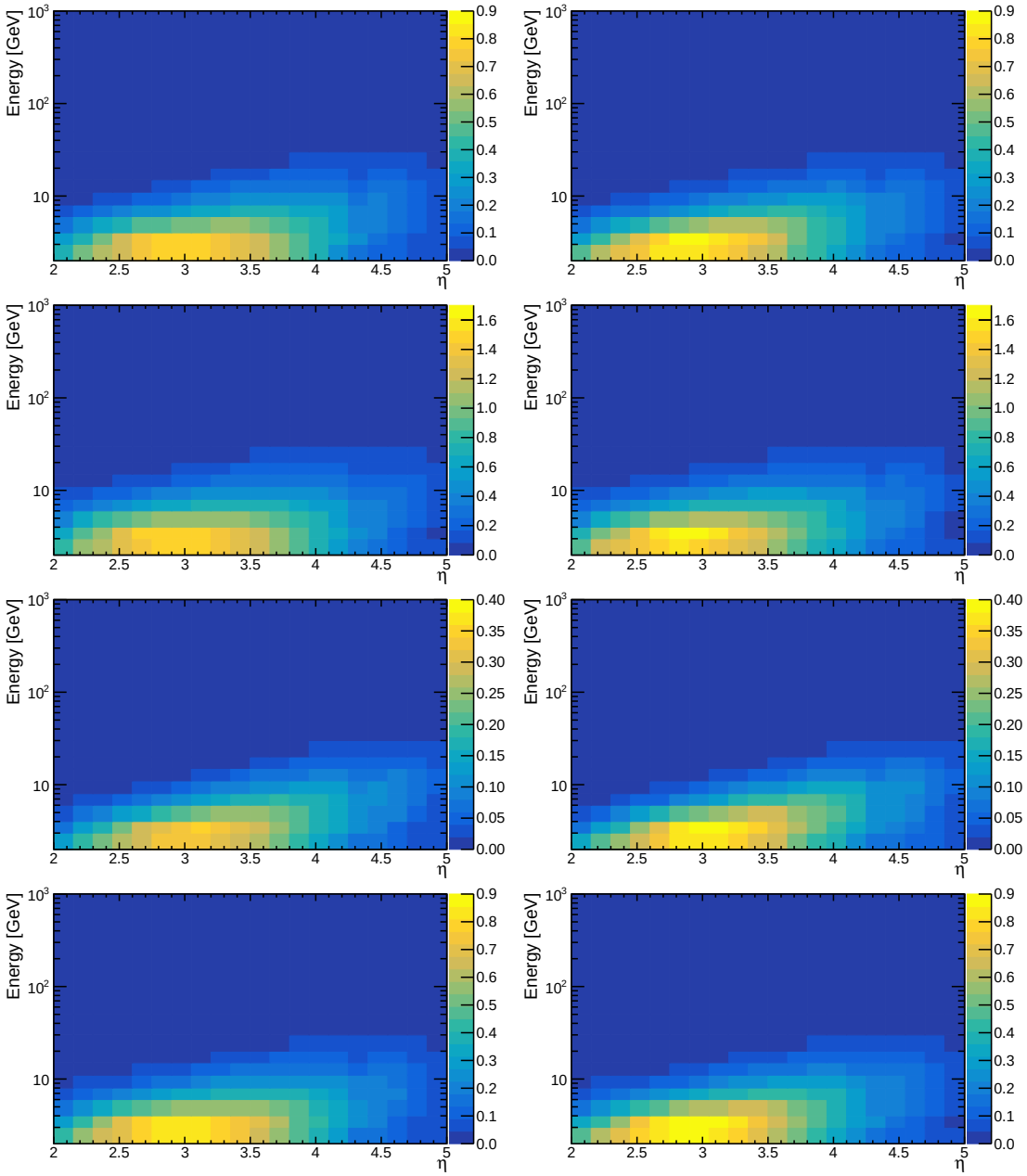

Figure C.1: Event averaged reconstructed charged multiplicity in data (left) and Monte-Carlo (right). Top to bottom: inclusive minimum-bias, hard scattering, diffractive enriched, non-diffractive enriched.

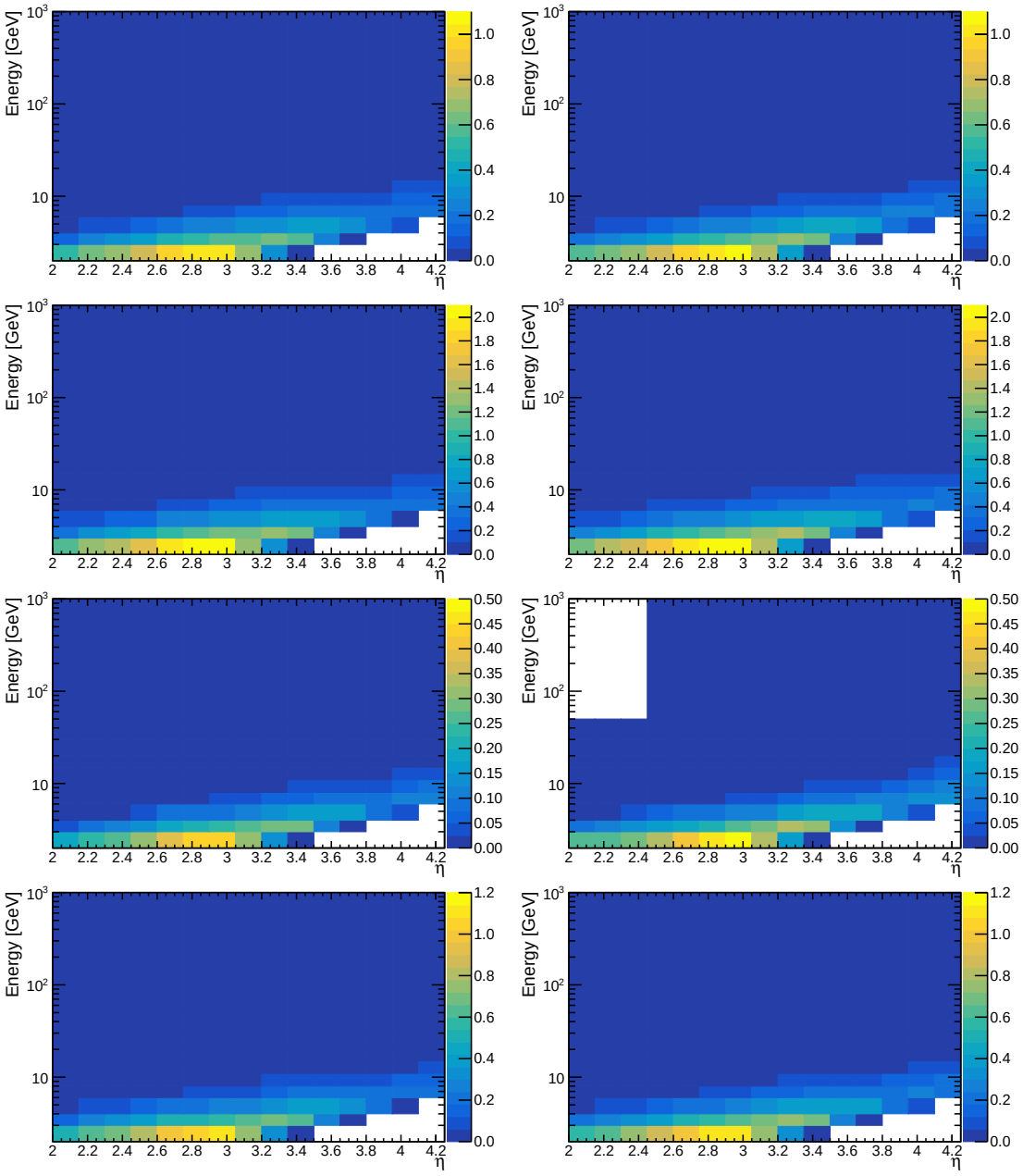

Figure C.2: Event averaged reconstructed neutral multiplicity in data (left) and Monte-Carlo (right). Top to bottom: inclusive minimum-bias, hard scattering, diffractive enriched, non-diffractive enriched.

C.2 Unfolded and truth level multiplicity

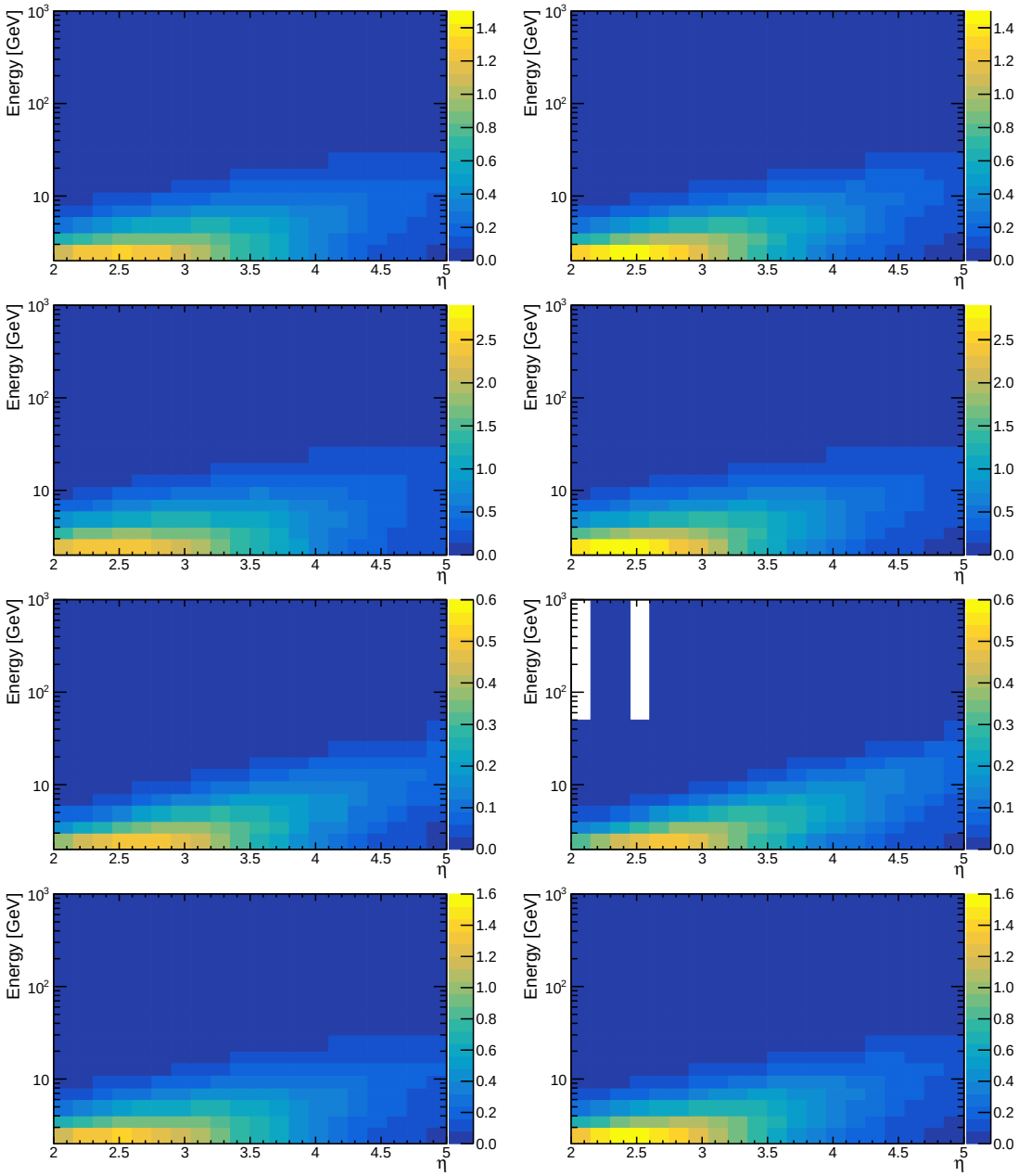

Figure C.3: Event averaged unfolded charged multiplicity from data (left) and Monte-Carlo truth level multiplicity (right). Top to bottom: inclusive minimum-bias, hard scattering, diffractive enriched, non-diffractive enriched.

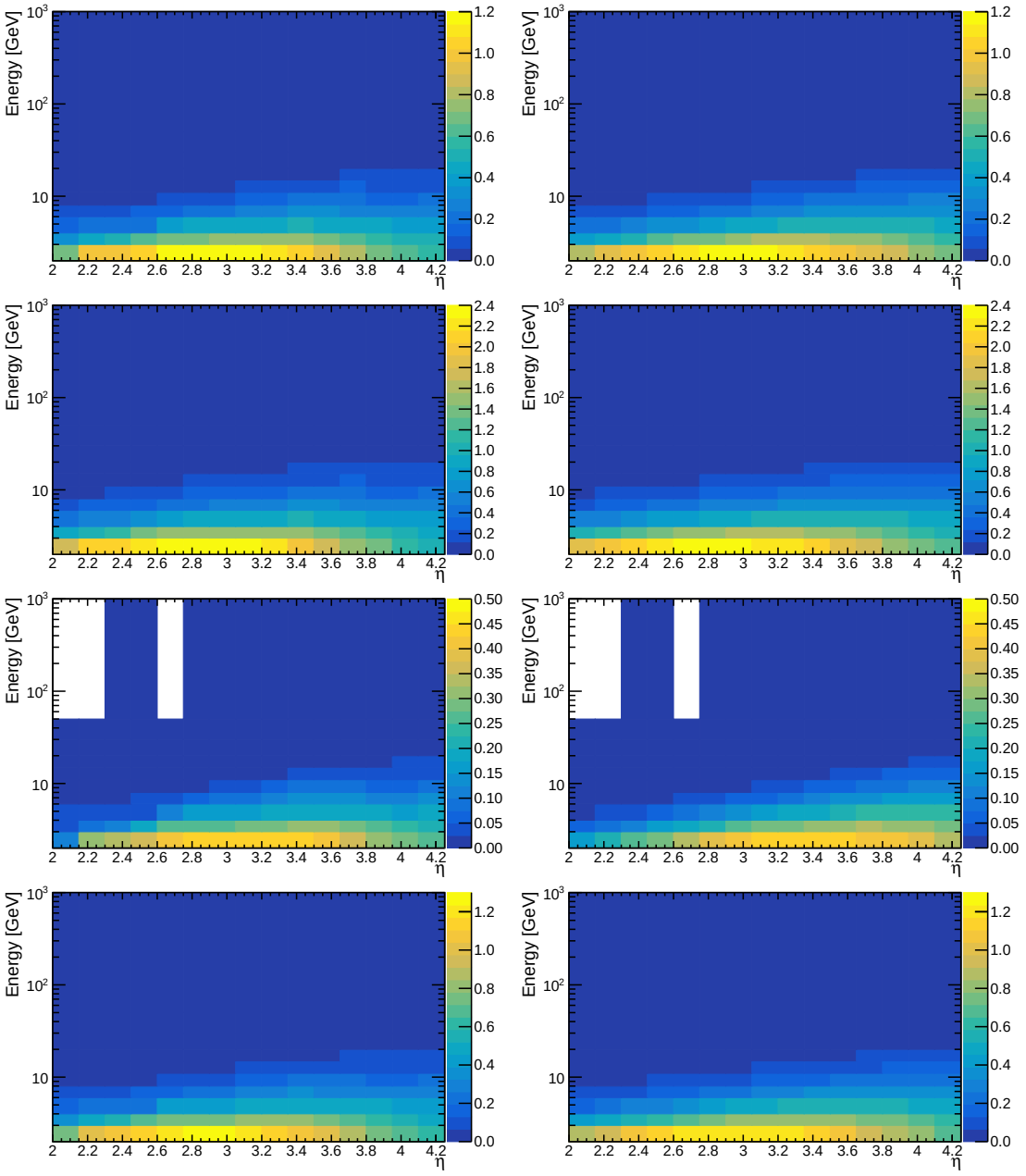

Figure C.4: Event averaged unfolded neutral multiplicity from data (left) and Monte-Carlo truth level multiplicity (right). Top to bottom: inclusive minimum-bias, hard scattering, diffractive enriched, non-diffractive enriched.

C.3 Theoretical predictions

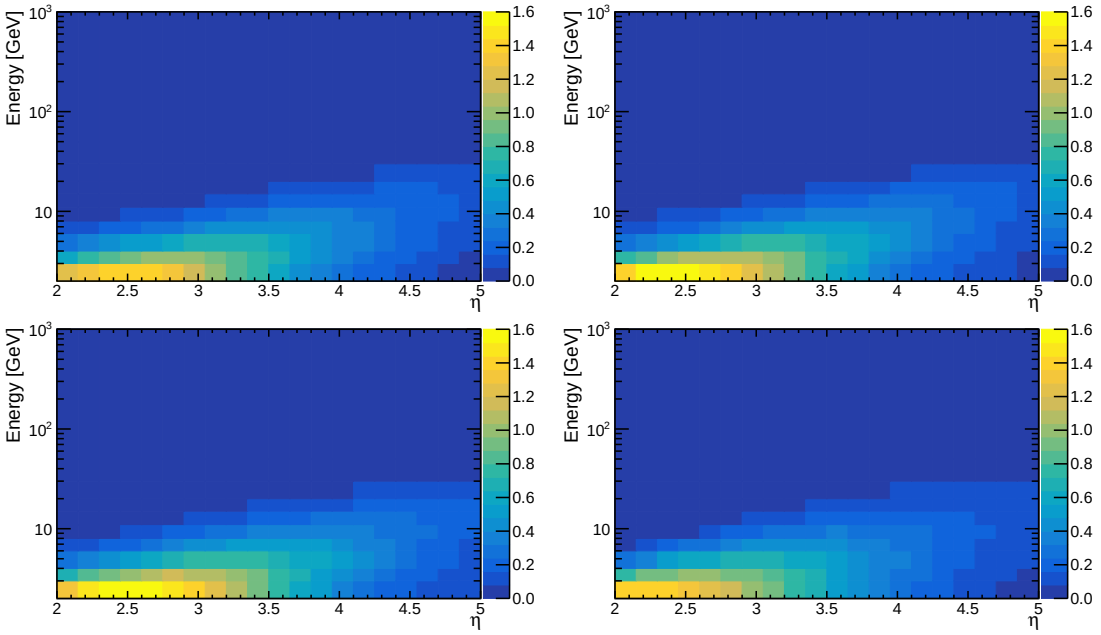

Figure C.5: Theoretical predictions of event averaged charged multiplicity for the inclusive minimum-bias event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

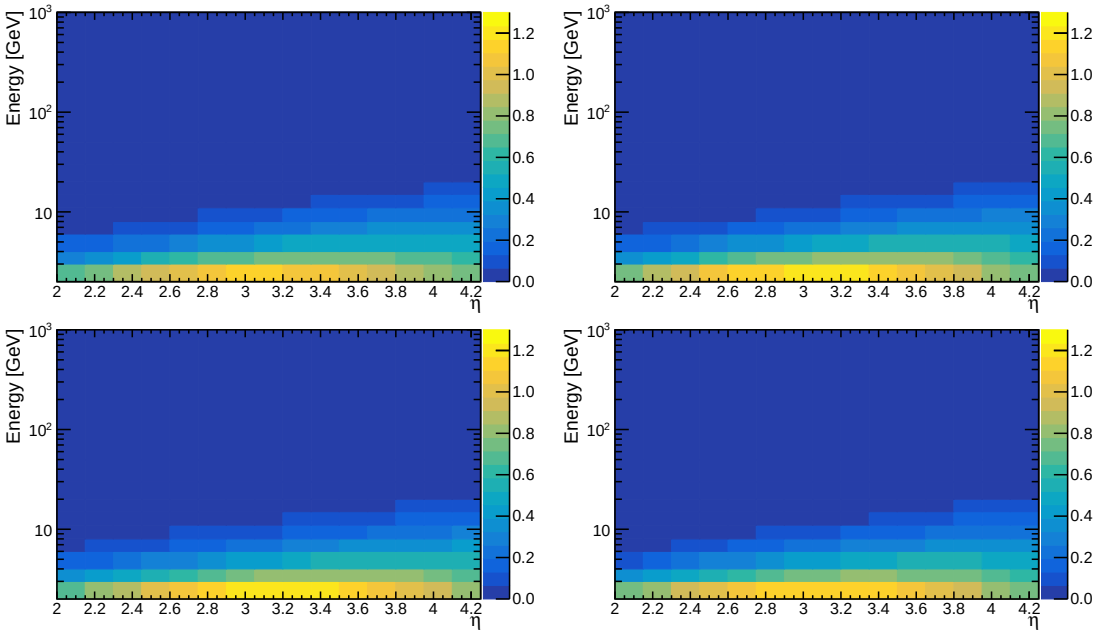

Figure C.6: Theoretical predictions of event averaged neutral multiplicity for the inclusive minimum-bias event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

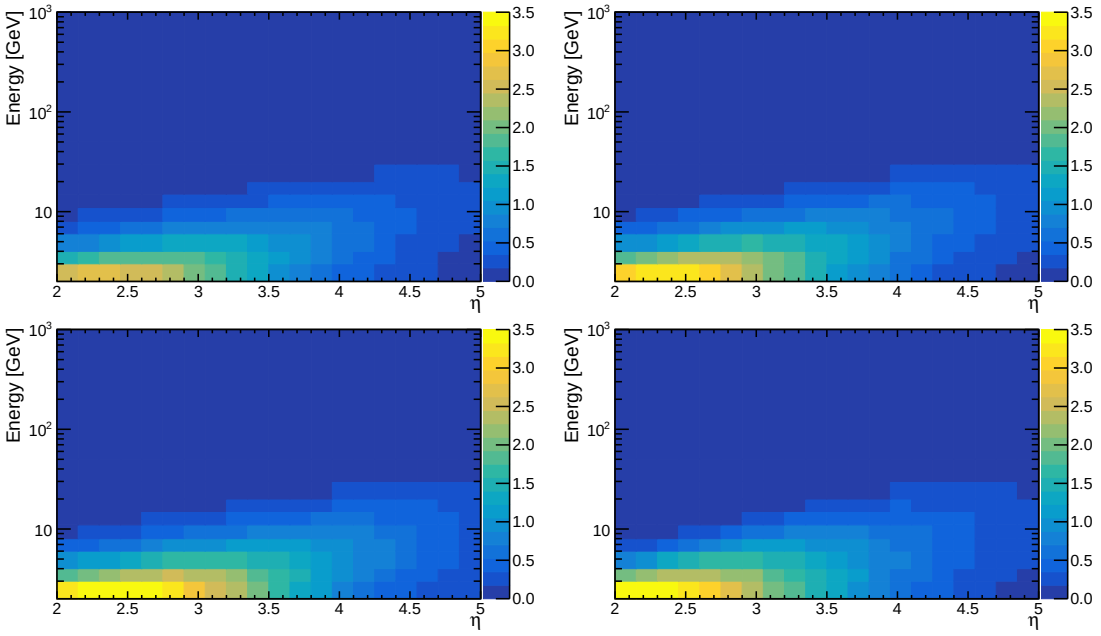

Figure C.7: Theoretical predictions of event averaged charged multiplicity for the hard scattering event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

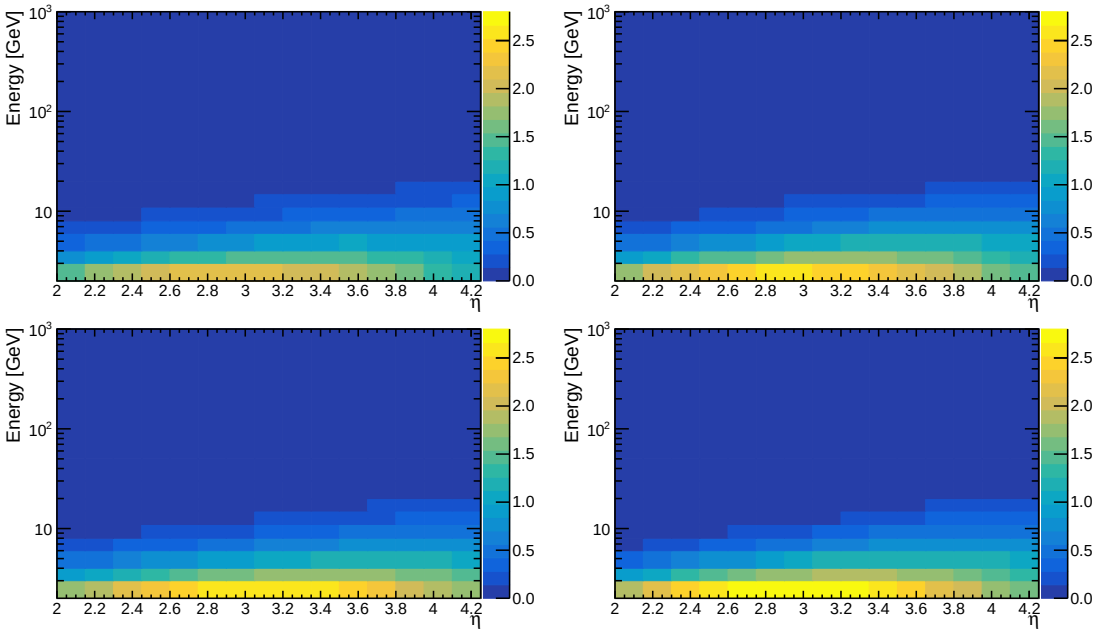

Figure C.8: Theoretical predictions of event averaged neutral multiplicity for the hard scattering event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

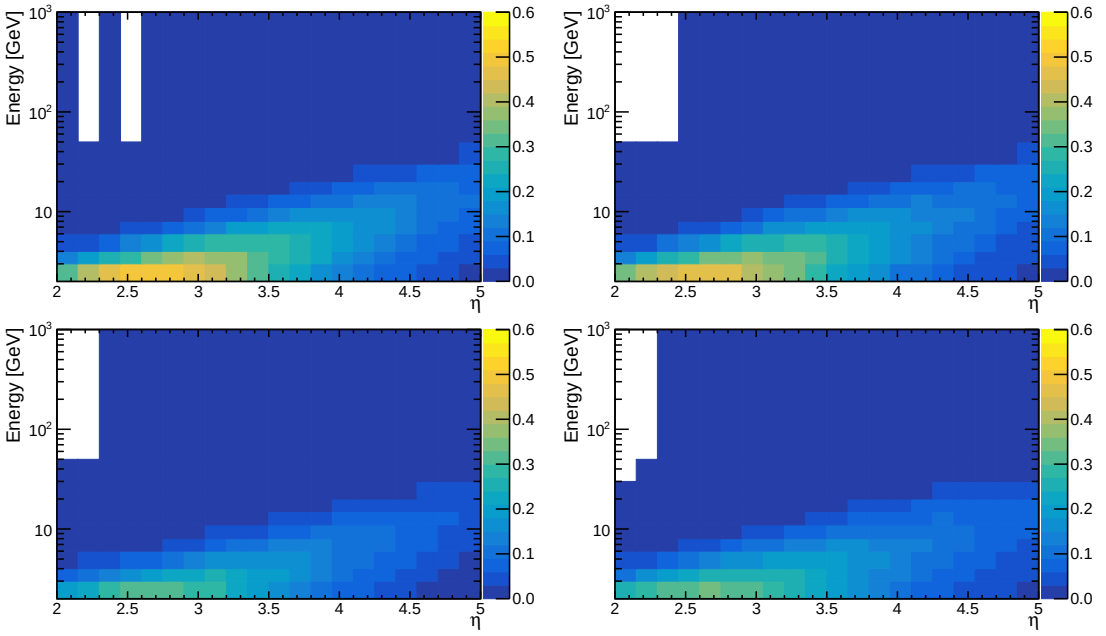

Figure C.9: Theoretical predictions of event averaged charged multiplicity for the diffractive enriched event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

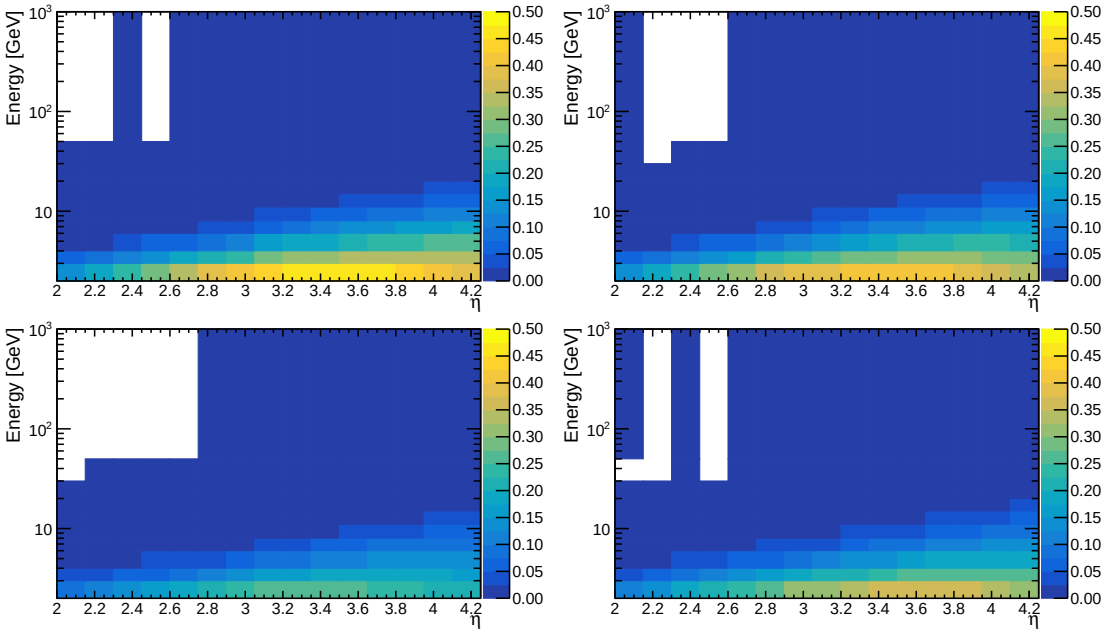

Figure C.10: Theoretical predictions of event averaged neutral multiplicity for the diffractive enriched event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

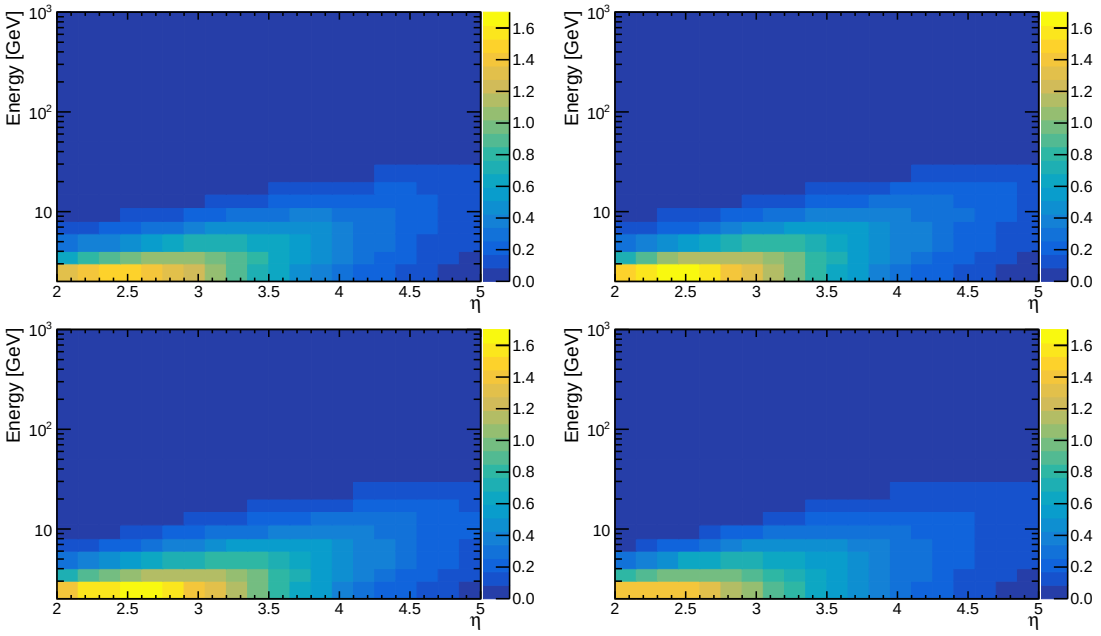

Figure C.11: Theoretical predictions of event averaged charged multiplicity for the non-diffractive enriched event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

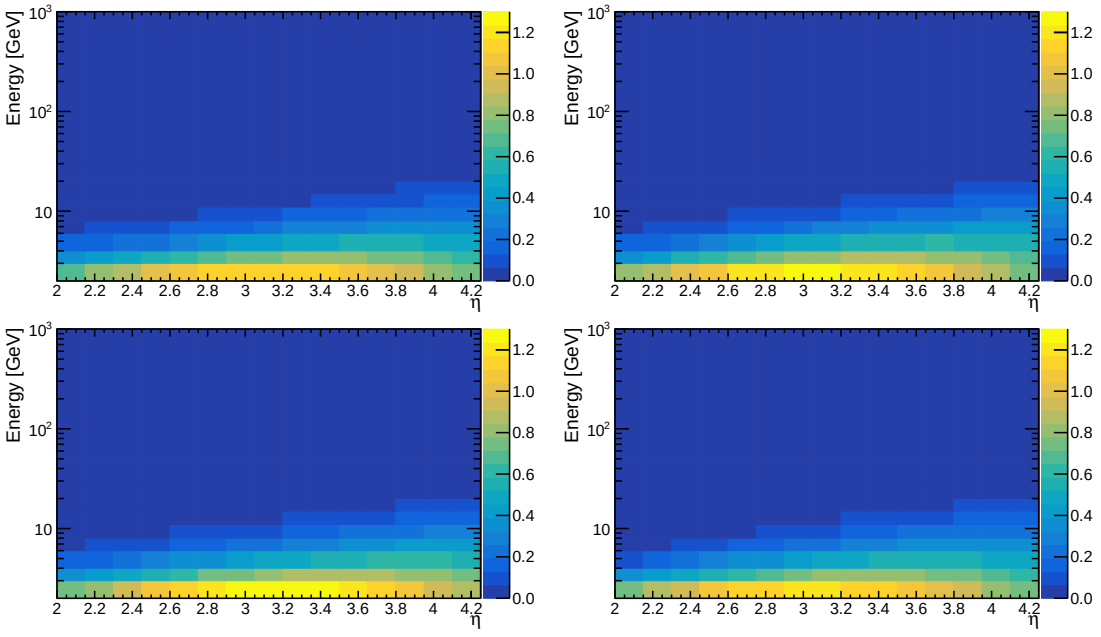

Figure C.12: Theoretical predictions of event averaged neutral multiplicity for the non-diffractive enriched event class. Clockwise from top left: Pythia 8.212 - LHCb, Pythia 8.212 - Monash, Sibyll 2.1 and EPOS LHC generators.

BIBLIOGRAPHY

[1] P. A. M. Dirac, *The quantum theory of the electron*, Proc. Roy. Soc. Lond. **A117** (1928) 610.

[2] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[3] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[6] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[7] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44** (1972) 189.

[8] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory; 1995 ed.*, Westview, Boulder, CO, 1995. Includes exercises.

[9] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8** (1996) 1.

[10] F. Mandl and G. Shaw, *Quantum Field Theory*, Wiley, 2013.

[11] T. Lancaster and S. J. Blundell, *Quantum field theory for the gifted amateur*, Oxford University Press, Oxford, Apr, 2014.

[12] A. Sirlin, *Role of sin2w(mz) at the Z peak*, Physics Letters B **232** (1989), no. 1 123 .

[13] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10 100001.

[14] S. A. Thomas, F. B. Abdalla, and O. Lahav, *Upper Bound of 0.28eV on the Neutrino Masses from the Largest Photometric Redshift Survey*, Phys. Rev. Lett. **105** (2010) 031301, arXiv:0911.5291.

[15] 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** (2012) 111807, arXiv:1205.5368.

[16] D. Hanneke, S. Fogwell, and G. Gabrielse, *New measurement of the electron magnetic moment and the fine structure constant*, Phys. Rev. Lett. **100** (2008) 120801.

[17] ATLAS, 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** (2012) 1, arXiv:1207.7214.

[18] CMS, 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** (2012) 30, arXiv:1207.7235.

[19] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the* 2π *decay of the* k_2^0 *meson*, Phys. Rev. Lett. **13** (1964) 138.

[20] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[21] Super-Kamiokande, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003.

[22] J. Lesgourgues and S. Pastor, *Massive neutrinos and cosmology*, Phys. Rept. **429** (2006) 307, arXiv:astro-ph/0603494.

[23] D. K. Ghosh and R. S. Hundi, *Neutrino masses along with fermion mass hierarchy*, Phys. Rev. **D85** (2012) 013005, arXiv:1108.3428.

[24] T2K, K. Abe *et al.*, *Upper bound on neutrino mass based on T2K neutrino timing measurements*, Phys. Rev. **D93** (2016), no. 1 012006, arXiv:1502.06605.

[25] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, Sov. Phys. JETP **26** (1968) 984, [Zh. Eksp. Teor. Fiz.53,1717(1967)].

[26] A. Einstein, *Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes*, Ann. Phys. **35** (1911) 898.

[27] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helv. Phys. Acta **6** (1933) 110.

[28] A. Y. Kniazev *et al.*, *A Search for Planetary Nebulae With the SDSS: the outer regions of M31*, Tech. Rep. arXiv:1311.1951, Nov, 2013. Comments: 58 pages, 17 figures, 2 tables, Accepted to Astronomical Journal.

[29] P. J. E. Peebles and B. Ratra, *The Cosmological constant and dark energy*, Rev. Mod. Phys. **75** (2003) 559, `arXiv:astro-ph/0207347`.

[30] R. P. Feynman, *QED: The Strange Theory of Light and Matter*, Princeton University Press, 1986.

[31] JabberWok, *Electron positron annihilation*, `https://commons.wikimedia.org/wiki/File:Electron-positron-annihilation.%svg`, 2006.

[32] I. R. Kenyon, *The Drell-Yan Process*, Rept. Prog. Phys. **45** (1982) 1261.

[33] C. Elsasseri, *Feynman diagram library (FDL)*, `http://www.physik.uzh.ch/~che/FeynDiag/`, 2017.

[34] J. Pumplin *et al.*, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, `arXiv:hep-ph/0201195`.

[35] 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** (1977) 641, [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[36] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438, [Yad. Fiz.15,781(1972)].

[37] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298.

[38] R. Gupta, *Introduction to lattice QCD: Course*, in *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2*, pp. 83–219, 1997. `arXiv:hep-lat/9807028`.

[39] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622.

[40] J. C. Collins, *Sudakov form-factors*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 573, arXiv:hep-ph/0312336.

[41] X. Artru, *Classical String Phenomenology. 1. How Strings Work*, Phys. Rept. **97** (1983) 147.

[42] G. Dissertori, I. Knowles, and M. Schmelling, *Quantum Chromodynamics: High Energy Experiments and Theory (International Series of Monographs on Physics)*, Clarendon Press, 2003.

[43] T. Pierog *et al.*, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. **C92** (2015), no. 3 034906, arXiv:1306.0121.

[44] B. Andersson and F. Soderberg, *The Diagonalization of the Lund fragmentation model. 1.*, Eur. Phys. J. **C16** (2000) 303, arXiv:hep-ph/9910374.

[45] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. **B238** (1984) 492.

[46] D. Amati and G. Veneziano, *Preconfinement as a Property of Perturbative QCD*, Phys. Lett. **83B** (1979) 87.

[47] P. Bartalini *et al.*, *Multi-Parton Interactions at the LHC*, 2011. arXiv:1111.0469.

[48] T. Regge, *Introduction to complex orbital momenta*, Il Nuovo Cimento (1955-1965) **14** (1959), no. 5 951.

[49] T. Regge, *Bound states, shadow states and mandelstam representation*, Il Nuovo Cimento (1955-1965) **18** (1960), no. 5 947.

[50] D. M. Austin, W. H. Greiman, and W. Rarita, *Regge pole model for p p and anti-p p elastic scattering at high energies*, Phys. Rev. **D2** (1970) 2613.

[51] A. B. Kaidalov, *Regge Poles in QCD*, Tech. Rep. hep-ph/0103011, Mar, 2001. Comments: 31 pages,17 figures.

[52] V. Schatz, *High-energy scattering in the Regge limit*, PhD thesis, Heidelberg U., 2003, arXiv:hep-ph/0307326.

[53] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[54] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[55] T. Pierog and K. Werner, *EPOS Model and Ultra High Energy Cosmic Rays*, Nucl. Phys. Proc. Suppl. **196** (2009) 102, arXiv:0905.1198.

[56] M. T. Dova, *Ultra-High Energy Cosmic Rays*, in *Proceedings, 7th CERN–Latin-American School of High-Energy Physics (CLASHEP2013): Arequipa, Peru, March 6-19, 2013*, pp. 169–190, 2015. arXiv:1604.07584. doi: 10.5170/CERN-2015-001.169.

[57] K. Werner, *Core-corona separation in ultra-relativistic heavy ion collisions*, Phys. Rev. Lett. **98** (2007) 152301, arXiv:0704.1270.

[58] E.-J. Ahn *et al.*, *Cosmic ray interaction event generator SIBYLL 2.1*, Phys. Rev. **D80** (2009) 094003, arXiv:0906.4113.

[59] T. Sjostrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, Phys. Rev. **D36** (1987) 2019.

[60] A. Capella, J. Kwiecinski, and J. Tran Thanh Van, *High-energy Nucleus-nucleus Inelastic Interactions in the Dual Parton Model*, Phys. Lett. **B108** (1982) 347.

[61] A. Accardi and D. Treleani, *Hard Parton Rescatterings and Minijets in Nuclear Collisions at LHC*, Phys. Rev. D **63** (2001), no. hep-ph/0009234. 11 116002/1.

[62] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[63] V. Frigo, *LHC map in 3D. Schéma du LHC en 3D*, https://cds.cern.ch/record/842700, Mar, 1997. AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[64] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, https://cds.cern.ch/record/1260465, Dec, 2008.

[65] C. E. Hill, A. M. Lombardi, E. Tanke, and M. Vretenar, *Present performance of the CERN proton linac*, https://cds.cern.ch/record/366818, Aug, 1998.

[66] B. R. Martin and G. Shaw, *Particle Physics*, Manchester Physics Series, Wiley, 2008.

[67] R. Bailey and P. Collier, *Standard Filling Schemes for Various LHC Operation Modes*, Tech. Rep. LHC-PROJECT-NOTE-323, CERN, Geneva, Sep, 2003.

[68] ATLAS Collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, J. Instrum. **3** (2008) S08003. 437 p.

[69] CMS Collaboration, S. Chatrchyan, G. Hmayakyan, V. Khachatryan, and A. M. Sirunyan, *The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment*, J. Instrum. **3** (2008) S08004. 361 p.

[70] ALICE Collaboration, K. Aamodt and A. Abrahantes Quintana, *The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment*, J. Instrum. **3** (2008) S08002. 259 p.

[71] LHCb, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[72] R. Veness *et al.*, *Design of beampipes for LHC experiments*, Vacuum **64** (2002) 467.

[73] LHCb, *LHCb VELO TDR: Vertex locator. Technical design report*, , CERN-LHCC-2001-011.

[74] J. Gassner, F. Lehner, and F. Steiner, *The mechanical design of the lhcb silicon trigger tracker*, , CERN-LHCB-2004-110.

[75] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030. LHCb-TDR-009.

[76] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029. LHCb-TDR-008.

[77] LHCb Outer Tracker Group, R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014), no. 01 P01002, `arXiv:1311.3893`.

[78] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037. LHCb-TDR-003.

[79] C. W. Fabjan and F. Gianotti, *Calorimetry for Particle Physics*, Review of Modern Physics **75** (2003) 1243.

[80] P. Collins, *High rapidity shower counters for LHCb: The HeRSCheL Project*, https://indico.cern.ch/event/295567/contributions/673792/attachments/55%4024/763299/la_biodola_paula_collins_may_2014.pdf, 2014.

[81] J. R. Forshaw, *Central Exclusive Production at the LHC*, Nucl. Phys. Proc. Suppl. **191** (2009) 247, arXiv:0901.3040.

[82] McNulty, R. for the LHCb collaboration, *Central Exclusive Production at LHCb*, Acta Phys. Polon. Supp. **8** (2015) 861, arXiv:1511.07302.

[83] E. Nurse and S. Sen, *Large Rapidity Gap Method to Select Soft Diffraction Dissociation at the LHC*, Adv. High Energy Phys. **2016** (2016) 5082847.

[84] M. R. Salzgeber, *The HeRSCheL detector & related physics result*, https://cds.cern.ch/record/2235872/files/LHCb-TALK-2016-385.pdf, 2016.

[85] Arduino, *Arduino Yún*, https://www.arduino.cc/en/Main/ArduinoBoardYun, 2017.

[86] M. T. Inc. *Atmega32u4*, http://www.microchip.com/wwwproducts/en/ATmega32U4, 1998-2017.

[87] Dog hunter Ilc. , *Linio.org*, http://www.linino.org/, 2017.

[88] T. O. Group, *The open group*, http://www.opengroup.org/, 1995-2017.

[89] W. Herr and B. Muratori, *Concept of luminosity*, https://cds.cern.ch/record/941318, 2006.

[90] CERN, *Accelerator performance and statistics*, https://acc-stats.web.cern.ch/acc-stats/#lhc/, 2017.

[91] J. Boyd, *LHC programme coordination web pages*, http://lpc.web.cern.ch/, 2016.

[92] R. Alemany-Fernandez, F. Follin, and R. Jacobsson, *The LHCB Online Luminosity Control and Monitoring*, .

[93] F. Alessio, R. Jaconbsson, and R. Le Gac, *LHCb operations plots webpage*, `http://lhcb-operationsplots.web.cern.ch/lhcb-operationsplots/index.htm`, 2016.

[94] E. V. Shuryak, *Quantum Chromodynamics and the Theory of Superdense Matter*, Phys. Rept. **61** (1980) 71.

[95] ALICE, J. Adam *et al.*, *Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions*, Nature Phys. (2017) , `arXiv:1606.07424`.

[96] LHCb, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state*, Phys. Rev. Lett. **112** (2014), no. 22 222002, `arXiv:1404.1903`.

[97] LHCb, R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001, `arXiv:1507.03414`.

[98] R. Aaij and et al. , *Angular analysis of the b0 k*0μ+μ decay using 3 fb1 of integrated luminosity*, Journal of High Energy Physics **2016** (2016), no. 2 104.

[99] G. Hiller and F. Krüger, *More model-independent analysis of b→s processes*, Phys. Rev. D **69** (2004) 074020.

[100] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, `arXiv:1705.05802`.

[101] E. Michielin, *The LHCb trigger in Run II*, PoS **ICHEP2016** (2016) 996.

[102] LHCb Trigger group, *Trigger schemes*, `https://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html`, 2017.

[103] LHCb, B. Sciascia, *LHCb Run 2 trigger performance*, PoS **BEAUTY2016** (2016) 029.

[104] G. Dujany and B. Storaci, *Real-time alignment and calibration of the LHCb Detector in Run II*, J. Phys. : Conf. Ser. **664** (2015) 082010. 8 p.

[105] R. Aaij *et al.*, *Tesla : an application for real-time data analysis in High Energy Physics*, Comput. Phys. Commun. **208** (2016) 35, `arXiv:1604.05596`.

[106] LHCb Trigger, A. Puig, *The LHCb Turbo stream*, Nucl. Instrum. Meth. **A824** (2016) 38.

[107] LHCb Tracking group, *Tracking and alignment plots for conference*, https://twiki.cern.ch/twiki/bin/view/LHCb/ConferencePlots, 2017.

[108] J. Van Tilburg and M. Merk, *Track simulation and reconstruction in LHCb*, PhD thesis, Vrije U. Amsterdam, Amsterdam, 2005, Presented on 01 Sep 2005.

[109] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.

[110] LHCb, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015), no. 02 P02007, arXiv:1408.1251.

[111] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015), no. 07 1530022, arXiv:1412.6352.

[112] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, Tech. Rep. LHCb-PUB-2014-044. CERN-LHCb-PUB-2014-044, CERN, Geneva, Sep, 2014.

[113] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley-Interscience, 2005.

[114] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007, arXiv:1405.7808.

[115] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015), no. 07 1530022, arXiv:1412.6352.

[116] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[117] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series)*, Wiley, 1989.

[118] M. A. Thomson, *Particle flow calorimetry and the pandorapfa algorithm*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **611** (2009), no. 1 25 .

[119] O. Augusto, *LHCb; LHCb Jet Reconstruction*, https://cds.cern.ch/record/1499646, Dec, 2012.

[120] M. Adinolfi *et al.*, *Performance of the RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[121] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[122] I. Belyaev *et al.*, *Simulation application for the LHCb experiment*, eConf **C0303241** (2003) TUMT003, arXiv:physics/0306035.

[123] S. Agostinelli *et al.*, *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003), no. 3 250 .

[124] LHCb Computing, *The BOOLE project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/, 2017.

[125] LHCb Computing, *The BRUNEL project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/, 2017.

[126] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[127] D. Britton and S. L. Lloyd, *How to deal with petabytes of data: the LHC Grid project*, Reports on Progress in Physics **77** (2014), no. 6 065902.

[128] The GridPP Collaboration *et al.*, *GridPP: development of the UK computing Grid for particle physics*, Journal of Physics G: Nuclear and Particle Physics **32** (2006), no. 1 N1.

[129] F. Stagni *et al.*, *LHCbDirac: Distributed computing in LHCb*, J. Phys. Conf. Ser. **396** (2012) 032104.

[130] J. T. Moscicki, F. Brochu, J. Ebke, and U. Egede, *Ganga: A tool for computational-task management and easy access to grid resources*, Computer Physics Communications **180** (2009), no. 11 2303.

[131] R. R. Wilson, *The Tevatron*, Phys. Today **30N10** (1977) 23.

[132] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, Rept. Prog. Phys. **70** (2007) 89, arXiv:hep-ph/0611148.

[133] A. Buckley, *Soft QCD in ATLAS: Measurements and modelling of multi-parton interactions*, Acta Phys. Polon. **B42** (2011) 2669, arXiv:1112.5477.

[134] CMS, P. Bartalini and L. Fano, *Multiple Parton Interactions Studies at CMS*, 2011. arXiv:1103.6201.

[135] LHCb, R. Aaij *et al.*, *Measurement of the forward energy flow in pp collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. **C73** (2013) 2421, arXiv:1212.4755.

[136] E. Gotsman *et al.*, *Survival probability of large rapidity gaps*, in *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*, pp. 221–241, 2005. arXiv:hep-ph/0511060.

[137] LHCb, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, in *Proceedings, 2010 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC 2010): Knoxville, Tennessee, October 30-November 6, 2010*, pp. 1155–1161, 2010. doi: 10.1109/NSSMIC.2010.5873949.

[138] T. Sjöstrand *et al.*, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[139] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, Eur. Phys. J. **C74** (2014), no. 8 3024, arXiv:1404.5630.

[140] T. Sjöstrand and M. van Zijl, *A multiple-interaction model for the event structure in hadron collisions*, Phys. Rev. D **36** (1987) 2019.

[141] C. Bierlich *et al.*, *PYTHIA 8*, http://home.thep.lu.se/~torbjorn/pythia82html/Welcome.html, 2017.

[142] G. Ingelman and P. E. Schlein, *Jet structure in high mass diffractive scattering*, Physics Letters B **152** (1985), no. 3 256 .

[143] L. Cornell U. *The CLEO-III detector: Design and physics goals*, pp. CLNS–94–1277, 1994.

[144] BaBar, B. Aubert *et al.*, *The BaBar detector*, Nucl. Instrum. Meth. **A479** (2002) 1, arXiv:hep-ex/0105044.

[145] N. Doble, L. Gatignon, K. Hübner, and E. Wilson, *The Super Proton Synchrotron (SPS): A Tale of Two Lives*, Adv. Ser. Direct. High Energy Phys. **27** (2017) 135.

[146] G. A. Schuler and T. Sjöstrand, *Towards a complete description of high-energy photoproduction*, Nuclear Physics B **407** (1993), no. 3 539 .

[147] G. A. Schuler and T. Sjostrand, *Hadronic diffractive cross-sections and the rise of the total cross-section*, Phys. Rev. **D49** (1994) 2257.

[148] ZEUS, H1, D. Britzger, *QCD Results from HERA*, in *52nd Rencontres de Moriond on QCD and High Energy Interactions (Moriond QCD 2017) La Thuile, Italy, March 25-April 1, 2017*, 2017. arXiv:1705.05204.

[149] E.-J. Ahn *et al.*, *LHC Update of the Hadronic Interaction Model Sibyll 2.1*, http://www.cbpf.br/{\%}7Eicrc2013/papers/icrc2013-0803.pdf, 2013.

[150] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **A559** (2006) 190.

[151] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, pp. CERN–LHCB–2008–026, LPHE–2008–04, 2008.

[152] LHCb Tracking group, *Tracking efficiencies*, https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingEfficiencies, 2017.

[153] *Private communication with the LHCb Tracking group*, 2017.

[154] E. Govorkova, *Study of π^0/γ efficiency using B meson decays in the LHCb experiment*, Phys. Atom. Nucl. **79** (2016), no. 11-12 1474, arXiv:1505.02960.

[155] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Meth. **A362** (1995) 487.

[156] R. J. Barlow, *Statistics: a guide to the use of statistical methods in the physical sciences*, Manchester physics series, Wiley, Chichester, 1989.

[157] T. Adye, *Unfolding algorithms and tests using RooUnfold*, in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, (Geneva), pp. 313–318, CERN, CERN, 2011. arXiv:1105.1160. doi: 10.5170/CERN-2011-006.313.

[158] G. D'Agostini, *Improved iterative Bayesian unfolding*, ArXiv e-prints (2010) arXiv:1010.0632.

[159] S. Farry, *Data-driven ghost rate estimate*, https://indico.cern.ch/event/600189/contributions/2518034/attachments/1%428716/2193383/sfarry_datadrivenghosts_170316.pdf, 2017.