WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

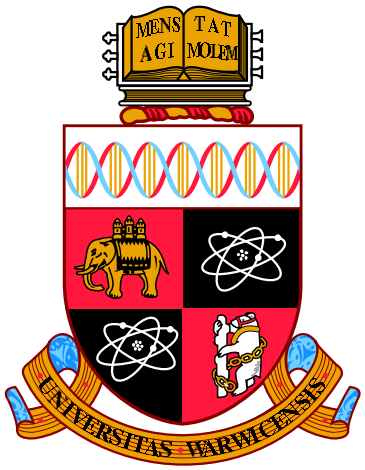
Permanent WRAP URL:

http://wrap.warwick.ac.uk/177364

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications



Precise measurements of the W boson's mass and lepton flavour universality, and trigger development with the LHCb experiment at CERN

by

Ross Hunter

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Physics

Department of Physics

September 2022

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Like any PhD, completing this one has certainly been a journey, and I am grateful for the help of many along the way. It's not often one is allowed to acknowledge all this help, so I hope the reader will allow me a brief indulgence.

Firstly, I'm grateful to Mark Hadley for the style template of this thesis. I'd like to thank my supervisor, Mika Vesterinen, for imparting huge amounts of wisdom and steering the ship so well. My trigger work would not have been possible without Olli Lupton, Rosen Matev and Sascha Stahl. The W mass measurement team is, in my opinion, a shining example of scientific collaboration and I count myself extremely lucky to have fallen in with such a formidable bunch. The Warwick EPP group is a great environment for research, and I am especially grateful to Flavia Cicala, Eleanor Jones, Andy Morris and Bryn Roberts for their emotional support, physics expertise and for many a wholesome and hilarious dinner party at CERN and various locations across the West Midlands. I could not forget another Warwick EPP alumnus and a key culinary influence, Arnau Brossa Gonzalo: we kept each other going and saw each other's PhDs through those dark days of the second national COVID-19 lockdown. When COVID struck Europe I was in danger of spending months doing very little work alone in my flat in Meyrin, so I thank the Hobson family for taking me in, for their love and for providing an excellent home-working atmosphere that led to a remarkably productive first few months of the pandemic.

The journey in fact began a long time before October 2018 however, and I am grateful to every inspiring teacher and every wonderful friend that has helped a council estate lad realize his professional dream. Nathan Ritson and Che Nabeta deserve special mention; you are truly tremendous people. Thank you to the Alexander Duckham Memorial Schools Trust for keeping me afloat financially during my undergraduate degree.

I'd like to thank my family for instilling good values in me, delivering a constant supply of Henderson's Relish in the flavourless South, and for providing a welcoming home every time I'm back in the city I love. Mum, you were and still are a fundamental part of any success I have. Dad, I'm very sad you couldn't be here to see this, as you deserve so much credit. I would've loved to see the smile on your face, although I think this one might be a bit heavy to put on the wall. Finally, I save the biggest thanks for Alice: you're the best partner a person could have and indeed, there were many times over the last eight years when the wheels could've come off were it not for you. To those I didn't mention, I apologize, but know that no matter how big or small your contribution, I am immensely grateful. In your own way, you're helping to further the noble pursuit of particle physics. I hope that isn't too shocking a revelation for you.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The data analyses were conducted on data collected by the LHCb experiment at the European Organization for Nuclear Research (CERN). The work presented has taken place in a team of collaborators, so although the majority was carried out by myself (under the direction of my supervisor, Dr Mika Vesterinen), it is necessary to present sections of work that have been completed by others. Such sections are indicated clearly at the beginning of each chapter. Parts of this thesis reiterate or expand upon published work, and in these cases this is stated - and the relevant publication is referenced - at the appropriate point in the text.

- Ross Hunter, September 2022

Abstract

In the electroweak sector of the Standard Model (SM), comparing precise measurements with predictions built on the SM's assumptions offers one of the principal avenues for indirect discoveries of new physics. The W boson mass, m_W, is a key SM parameter that is notoriously difficult to measure at hadron colliders, and the lack of high-precision measurements of it limits the sector's discovery power. Meanwhile, the SM's fundamental property of lepton flavour universality (LFU) has been questioned by hints of discrepancy in recent measurements of rare B-meson decays and legacy tests of W-boson decays. This thesis presents two measurements using LHCb's 2016 data that address these important issues: first, a proof-of-principle extraction of m_W that paves the way for a competitive legacy measurement; and second, a test of the W boson's LFU in decays to tau leptons and muons that, when completed, will validate and complement other recent measurements shedding light on previous LFU anomalies.

The value of m_W was measured to be

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

which is consistent with previous direct measurements and indirect SM predictions. It is not consistent with the very-recent CDF measurement, and therefore places LHCb in prime position to address this high-profile disagreement with a future measurement using all available data.

LHCb is currently undergoing commissioning for a fresh period of data-taking, which features a brand-new detector, a factor of five more collisions and a fully redesigned trigger system. The development of the trigger validation tool `HltEfficiencyChecker` is also presented, which plays a crucial role in facilitating trigger optimization that fully exploits the new detector, whilst also conforming to its constraints. This tool helped the collaboration decide that the new first-level trigger should be implemented with GPUs, and is now widely used in LHCb, as exemplified in the development of trigger selections for electroweak processes in Run 3 presented here.

“*Someone once told me that time was a predator that stalked us all our lives. But I rather believe that time is a companion who goes with us on the journey - reminds us to cherish every moment, because they'll never come again.*

What we leave behind is not as important as how we've lived.

After all, Number One, we're only mortal.”

- Captain Jean-Luc Picard

CHAPTER 1

Introduction

Particle physics is the field that studies the smallest building blocks of the universe and how they interact with one another via the fundamental forces. Although the objects of interest are infinitesimally small, almost everything else about particle physics is remarkably large. The field is a colossal endeavour that features tens (perhaps hundreds) of thousands of scientists around the globe - a field of cathedral-sized detectors, that surround atom-smashers of tens of kilometres in length which produce conditions similar to the moments after the Big Bang. The goals of the field are suitably lofty as well: particle physics is really trying to understand how *everything* works, at least on the most basic scale. What could be more grand a pursuit?

Although humans have always been curious about the universe and what it is made of, modern particle physics emerged at the turn of the 20th century. Experiments began to peer into the atom and to question the duality of waves and particles, and suddenly humanity's understanding of the microscopic world was revolutionized by the nascent theories of quantum mechanics. Particle physics since then has always been a highly predictive field; innovations in the mathematical theories give precise predictions of what particles should be seen and when they should be seen - it is then up to the experimentalists to devise an experiment to find them. Throughout the mid-late 20th century, this back-and-forth between theorists and experimentalists was extremely productive. Dozens of particles discoveries from the 1930s to 1950s lead theorists to the quark model, which says that the protons and neutrons that make up the nuclei of all the atoms in the human body are in fact made of *quarks*, and that these - along with the electron, a *lepton* - are the true fundamental particles. As the century came to a close more quarks and more leptons arrived on the scene at particle physics experiments, and the field's theorists were able to bookkeep them all together into one coherent and beautiful theory: the "Standard Model" of particle physics.

The Standard Model is arguably the biggest theoretical achievement of the field. It describes, in just a few lines of mathematics, how all the fundamental particles interact with each other. Its equations have had staggering success at predicting the outcomes of experiments. Besides a few hints of discrepancy - some of which this thesis will go into - every result from every particle physics experiment[1] over the last half-century has aligned with the expectations of the SM. The discovery of the Higgs boson [2, 3] in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) was the cherry on the cake - proof of the Higgs mechanism that holds the SM together and gives mass to the fundamental particles. This discovery - although expected - was spectacular, and can count the author of this thesis as among the many it inspired into particle physics.

Today, the SM stands unfazed by the ceaseless and painstaking examinations physicists have devised to unearth its flaws. However, for all its predictive power in atom-smashing experiments, there are several important features of the universe it cannot describe. Gravity - one of the four fundamental forces of the universe - is conspicuously absent. It has no answer for the origin or particle content of the "dark matter" that is theoretically necessary to explain both the formation of galaxies and their observed patterns of rotation. To the best of humanity's knowledge, the universe is dominated by *matter* rather than *antimatter*, yet the SM predicts that they should've been created in equal measure by the Big Bang, and gives no mechanism to yield the observed matter-antimatter asymmetry. There are further theoretical problems, all clearly showing that something is missing, and that the SM cannot be the ultimate theory of the universe's fundamental interactions.

So where to go next? A fine question, and if you get the correct answer you will surely get a call from the Royal Swedish Academy of Sciences that you do not want to miss. There are a plethora of theories around describing how to extend, generalize or modify the SM to solve the problems listed above, but to prove any of them requires experimental confirmation of their beyond-the-Standard-Model (BSM) phenomena. Another angle to take is to test the SM's predictions at ever higher precision, hoping to see a significant-enough deviation to suggest what theoretical direction should be followed - a foot in the door. A large portion of the field is dedicated to this angle of attack, and it is the trajectory that is taken here. The electroweak (EW) part of the SM (responsible for the everyday force of electromagnetism, and the nuclear weak force that keeps the stars shining) is a fertile ground for such high-precision tests: relatively-speaking, it is extremely well-understood theoretically; the physics processes involved are comparatively simple; and it is at the centre of the SM formalism. For these reasons however, there is a long

[1]The discovery of neutrino oscillations in 1998 [1] provided evidence that neutrinos have mass, which is not strictly predicted by the SM. However, it can be easily incorporated, and as such it is debatable whether this is truly a beyond-the-Standard-Model phenomenon.

history of electroweak precision measurements at particle colliders, and increasing the level of precision requires supreme effort in understanding the potential experimental biases, and constant innovation in theoretical predictions to match.

This thesis describes two high-precision tests of the EW sector of the SM: a measurement of the W boson mass, and a test of the W boson's property of "lepton flavour universality". The meaning of the latter property will be described in due course, as will the motivations for making these particular measurements. Both take place using data collected in 2016 by the LHCb [4] experiment at the LHC. LHCb was primarily designed as an experiment for studying the physics of hadrons containing charm and beauty quarks in the aim of shedding light on the aforementioned matter-antimatter asymmetry, but has broadened since its inception into a general-purpose physics experiment. The measurements here contribute to an increasingly impressive catalogue of EW precision measurements. At the time of writing, LHCb is emerging from a major upgrade designed to vastly increase its physics reach. A notable part of that upgrade is a complete redesign of its trigger system to fully exploit the new detector's capabilities and the larger rate of data that it will collect in the coming years. The author has played a part in this upgrade, first by developing tools to facilitate optimization of the new trigger system, and then in writing trigger "selections" to pick out those collision events that involve EW processes. The latter work ensures the data will be collected to allow further high-precision EW tests in the future.

The structure of the thesis is as follows. To fully understand the measurements introduced, a brief primer of the required theoretical background is presented in Chapter 2. It was mentioned that supreme understanding of the experimental apparatus is required, so Chapter 3 describes the LHCb experiment. The author's work on the upgraded trigger system is the subject of Chapter 4. The thesis then goes into further depth on the theoretical modelling aspects of precision EW physics in Chapter 5, followed by the detector modelling strategies in Chapter 6. The author's primary contribution to the W mass measurement was the study of muon reconstruction efficiencies, which is presented by Chapter 7. This is followed directly by description of the W mass measurement in its entirety in Chapter 8. Chapter 9 concerns the test of the W boson's lepton flavour universality. Finally, conclusions of the thesis are presented in Chapter 10.

1.1 Conventions and coordinates

Before diving deep into the theory of the SM, the reader should be aware of a number of conventions used throughout the thesis:

- In the electroweak process $q\overline{q} \to Z/\gamma^* \to \ell\ell$ (ℓ is a lepton, q and $\overline{q}$ are a quark and an antiquark respectively), where the interaction can be mediated by either a Z

boson or a photon, only Z will be used to denote both,

- In particle decays, the charge of the particles will usually be omitted, both for brevity and because the inclusion of both charges is usually implied, e.g. $W \to \mu\nu$ is used in place of both $W^+ \to \mu^+\nu_\mu$ and $W^- \to \mu^-\overline{\nu}_\mu$. The slightly more complex case of $W \to (\tau \to \mu\nu\nu)\,\nu$ corresponds to both $W^+ \to (\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$ and $W^- \to (\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau)\overline{\nu}_\tau$,
- Acronyms and abbreviations will generally be defined the first time they are used in each chapter, to save the reader having to look back through the whole document,
- The terms "cut" and "selection" will be used interchangeably; both are a colloquial term for a requirement that is asked of every member of a dataset. For example, the requirement that every muon have momentum p greater than 20 GeV could be described as a "$p > 20\,\text{GeV}$ cut", "$p > 20\,\text{GeV}$ selection" etc,
- Natural units $\hbar = c = 1$ are used throughout. Energies, momenta and masses therefore have the same unit, which is typically MeV, GeV or TeV.

In addition, there are many useful variables and coordinates that pop will up in various places, so are defined here for reference[2]. Beginning with coordinate systems, it is useful to note that collisions at hadron colliders have a cylindrical symmetry about their beam axis, which is typically labelled as the z axis, and therefore it is common to describe outgoing particles by their polar angle θ, and their azimuthal angle ϕ, both with respect to this axis.

Collisions of two hadrons are actually between one of the partons (quarks and gluons) from each hadron, and - as will be explained fully in the Section 2.4 - the longitudinal component of momentum of any given parton is *a priori* unknown, so momentum conservation cannot be utilized in this direction. In the transverse plane, neglecting any small intrinsic transverse momentum that the parton has, momentum conservation is obeyed. Therefore, rather than considering the momentum p, the transverse momentum $p_\text{T} = p\sin\theta$ is often the momentum variable of choice.

The lack of knowledge of how boosted the interaction is in the longitudinal direction also drives a preference for using Lorentz-invariant quantities in describing the collision's kinematics. One such quantity is Bjorken x, which - in the high-energy limit - is the fraction of the hadron momentum carried by the parton of interest, and so is bounded in $[0, 1]$. In a high-energy (such that hadron masses can be ignored) "s-channel" $2 \to 2$ process such as $q\overline{q} \to Z \to \ell\ell$, the dilepton Lorentz-invariant mass squared is approximately $M^2_{\ell\ell} = (x_1 p_1 + x_2 p_2)^2 \approx x_1 x_2 s$, where $s = 4E_p^2$ is the centre-of-mass energy squared, and p_1 and p_2 are the hadron momenta [6].

Another useful quantity is the *rapidity*, y, defined for a particle of energy E as

[2]This section was inspired by a similar one in Ref. [5].

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

where p_z is the component of the momentum parallel to the beam axis. In the aforementioned $q\overline{q} \to Z \to \ell\ell$ process, the Z rapidity is approximately

$$y_{\ell\ell} \approx \frac{1}{2} \ln \frac{x_1}{x_2} , \tag{1.2}$$

and can be simply related to the invariant mass m, $\sqrt{s}$ and the Bjorken x of the two incoming partons by

$$x_{1,2} = \frac{m}{\sqrt{s}} e^{\pm y} . \tag{1.3}$$

The rapidity roughly encodes "how far forward" the particle is travelling (higher values of rapidity correspond to further forward). It is also useful because it transforms linearly under Lorentz boosts, meaning that differences in rapidity are Lorentz-invariant. Cross sections differential in rapidity $\mathrm{d}\sigma/\mathrm{d}y$ are therefore also Lorentz-invariant [7].

For massless particles, the rapidity becomes the *pseudorapidity*

$$\eta \equiv \frac{1}{2} \ln(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}) = -\ln(\tan \frac{\theta}{2}) , \tag{1.4}$$

which depends only on the polar angle θ, and is therefore a useful quantity for describing flight direction and the angular acceptance of different detectors. The separation of two particle tracks coming from the same interaction point can be defined by differences in η and ϕ (denoted $\Delta\eta$ and $\Delta\phi$ respectively), and these can be combined to define a cone in (η, ϕ) space with the variable $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

A final quantity of interest is the *impact parameter* (IP), which is the scalar distance between a particle track and a *primary vertex* (PV) at their closest approach. The PV is the measured pp interaction point. If a particle is produced at the PV and decays some short time (and distance) later, the decay products coming from the secondary decay will typically have a non-zero IP with respect to the PV, as they have in fact originated from a secondary decay vertex (SV). The decay products of heavy-flavour hadrons and tau leptons typically have a significant IP that can be measured, whereas the final-state particles from e.g. $W \to \mu\nu$ or $Z \to \mu\mu$ should have an IP of zero (ignoring detector resolution effects), as the W and Z bosons decay effectively instantaneously. The impact parameter is therefore very important for separating all of these decays, and features throughout this thesis.

CHAPTER 2

Theoretical background

The aim of this chapter is to set a theoretical foundation for the physics measurements described in this thesis. The first few sections are a review of the salient points of a 3rd- or 4th-year undergraduate particle physics course: the Standard Model (SM) is introduced in Section 2.1, before delving into how the gauge bosons acquire their mass via spontaneous symmetry-breaking in Section 2.2. The chapter then focuses in on those topics most important for precision electroweak (EW) physics in proton-proton (pp) collisions at the Large Hadron Collider (LHC): higher-order corrections in perturbative Quantum Chromodynamics (pQCD) are described in Section 2.3; parton distribution functions (PDFs) are the subject of Section 2.4; and precise theoretical predictions of the Drell-Yan process - the process studied here - are tackled in Section 2.5.

Since this chapter is a theoretical review, it is compiled from a number of sources. The main sources for the first three sections are Refs. [8–13], while Refs. [14–17] were particularly useful for Sections 2.3, 2.4 and 2.5. Other references are cited in the text where appropriate.

2.1 The Standard Model of particle physics

The particle content of the SM is summarized in Figure 2.1: matter is made up of the twelve spin-1/2 fermions, plus an antiparticle associated with each fermion (e.g. the u - or "up" - quark and the corresponding antimatter $\overline{u}$ - or "anti-up" quark). Of the fermions, six are quarks, which are further split into three generations of ascending mass. Within a generation, there is an "up-type" quark (u, c, t) carrying an electric charge of $+2/3e$ and a "down-type" quark (d, s, b) carrying $-1/3e$. Here, $e \approx 1.6 \times 10^{-19}$ C is the electronic charge. The leptons also have a three-generation structure, each generation including a charged $(-e)$ lepton (e, μ, τ) and an electrically-neutral neutrino $(\nu_e, \nu_\mu, \nu_\tau)$.

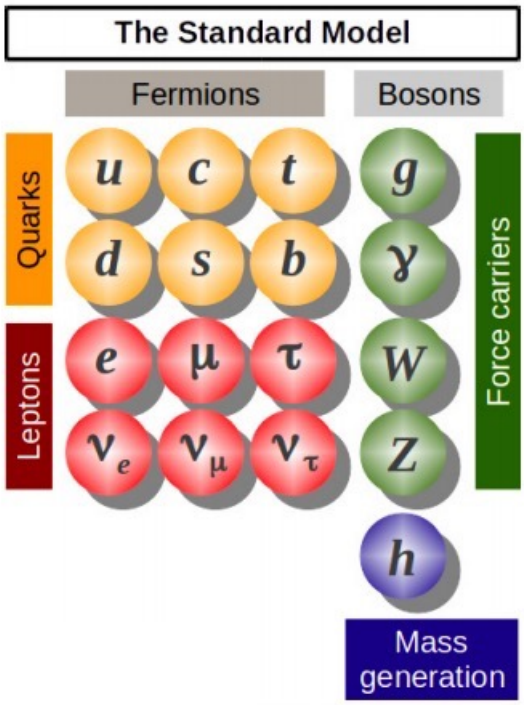


Figure 2.1: Schematic of the fundamental particles of nature. Taken from Ref. [18].

Three fundamental forces are also described, each mediated by the exchange of associated spin-1 vector bosons. Quarks are distinct from leptons because the former interact via the strong nuclear force (mediated by the massless gluon g) - whereas the latter do not. Electromagnetism is mediated by the massless photon (γ), whereas the massive W and Z bosons account for the weak interaction. Finally, the Higgs boson (H^0) is the only fundamental scalar (spin-0) particle observed in nature, and its presence is part of the mechanism that gives particles mass.

The SM's mathematical description of the interactions introduced above begins by representing each force with a quantum field theory. Like classical field theory, the dynamics and interactions of the SM are therefore encoded in the *Lagrangian*[1] $\mathcal{L}$. The equations of motion, which described the properties and interactions of the system, are found by extremizing the *action*, which is the integral of the Lagrangian over space and time. This yields the Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial \psi} - \partial^{\mu}\left(\frac{\partial \mathcal{L}}{\partial\left(\partial^{\mu}\psi\right)}\right) = 0\,, \tag{2.1}$$

where ψ is a matter field, ∂^{μ} is the partial derivative with respect to a spacetime index μ, and the Einstein summation convention has been used. From a given form of

[1]Lagrangian *density* to be more precise, but it is almost always referred to as just "the Lagrangian". The difference is unimportant for this discussion.

the Lagrangian, the Euler-Lagrange equations specify what interactions are allowed and provide a basis for calculation of quantities such as the probability or rate at which a particular interaction will occur. The Lagrangian sets down the rules, and thus the SM Lagrangian has been refined to reproduce the interactions between fundamental particles that have been observed in nature. The possible form of each force's Lagrangian is greatly restricted by the requirement of Lorentz invariance (properties and interactions shouldn't change if the frame of reference is moving at a different speed, or is in a different point in spacetime). Furthermore, the fundamental interactions and bosons that mediate them naturally arise after requiring that the Lagrangian is invariant under a *local gauge transformation*. This can be seen by applying the simplest version of the transformation to the fermions, which obey the Dirac equation and therefore are described by the Dirac Lagrangian:

$$\mathcal{L}_{\text{fermions}} = \overline{\psi}\,(i\gamma^{\mu}\partial_{\mu} - m)\,\psi\,, \tag{2.2}$$

where γ^{μ} is a set of four matrices, the form of which can be ignored here, and $\overline{\psi} = \psi^{\dagger}\gamma^{0}$. The simplest case of the local gauge transformation is

$$\psi \rightarrow \psi' = \psi e^{ie\alpha(\boldsymbol{x})}\,, \tag{2.3}$$

for a space-time coordinate $\boldsymbol{x}$, a charge e and a phase factor α, which is free to continuously vary in $[0, 2\pi]$. This is the quantum mechanical version of the gauge symmetry of Maxwell's equations. It is a "local" transformation in the sense that the phase α can vary with $\boldsymbol{x}$, and this freedom causes an extra derivative of α to emerge when the gauge transformation is applied to the $\partial_{\mu}\psi$ term, ruining the gauge invariance. The invariance can be restored by introducing a new gauge field A^{μ} that transforms in the complementary way $A^{\mu} \rightarrow A^{\mu} - \partial^{\mu}\alpha$. The derivative is then replaced by a covariant derivative D^{μ}:

$$\partial^{\mu} \rightarrow D^{\mu} = \partial^{\mu} + ieA^{\mu}\,, \tag{2.4}$$

and the extra terms incurred in the transformation of ψ and A^{μ} cancel out, restoring the gauge invariance. The new Lagrangian

$$\mathcal{L}_{\text{fermions}} = \overline{\psi}\,(i\gamma_{\mu}D^{\mu} - m)\,\psi\,, \tag{2.5}$$

can be expanded out, and a term proportional to $e\overline{\psi}\gamma^{\mu}\psi A_{\mu}$ will be found, which can be identified as an *interaction* term between ψ and A^{μ}, with a coupling strength e. In summary, applying the requirement of local gauge invariance, and taking the necessary steps to preserve it, has neatly conceived the photon and its fundamental interaction

with the fermions. Including the photon's kinetic terms, the full Lagrangian of Quantum Electrodynamics (QED) is therefore

$$\mathcal{L}_{\text{QED}} = \underbrace{i\overline{\psi}\gamma_\mu\partial^\mu\psi}_{\text{kinetic term of }\psi} - \underbrace{m\overline{\psi}\psi}_{\text{mass term of }\psi} + \underbrace{e\overline{\psi}\gamma_\mu A^\mu\psi}_{\text{interaction term}} - \underbrace{\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{kinetic term of }A^\mu} \quad . \tag{2.6}$$

Each of the fundamental forces arises in the same way, but with a slightly more complex gauge transformation. $U \equiv e^{ie\alpha}$ can be thought of a 1×1 unitary[2] matrix, and so symmetry under transformations of this form correspond to the U(1) symmetry group. Thus, the quantum field theory representing QED is a U(1) *gauge field theory*. That the transformation is unitary corresponds to it leaving probability densities unchanged under the transformation. The *degree* N of the group not only denotes the dimension of the matrices representing it, but is also related to the number of gauge fields (bosons) required to assert its gauge invariance: a unitary group U(N) has N^2 - hence one photon - while a "special unitary" group has $N^2 - 1$ bosons. The strong force is described by the SU(3) gauge theory Quantum chromodynamics (QCD), which has similar - but more complex - kinetic terms and contributions to the covariant derivative as QED. It has eight mediating gluons, and particles that interact via the strong force (quarks and gluons) carry *colour* charge, which is analogous to QED's electric charge and is conserved in strong interactions. The weak force is described by a $\text{SU}(2)_L$ theory that couples to particles carrying the *weak isospin*, and has two oppositely charged bosons W^1 and W^2, plus an electrically-neutral boson W^3.

The subscript L in $\text{SU}(2)_L$ reflects that the weak force only acts upon particles with left-handed *chirality*. In the ultra-relativistic limit relevant at hadron colliders, chirality is approximately the same as helicity, which is the direction of a particle's spin relative to its velocity: a particle with right (left)-handed helicity has spin parallel (antiparallel) to the direction of motion. That the weak force couples only to left-handed chiral particles was a theoretical solution proposed to explain the discovery of parity violation in charged-current weak decays [19]. Making this requirement has significant further consequences. Whilst the charged weak bosons were observed to act on left-handed states, the neutral boson was found to couple to either handedness, suggesting that the observed weak bosons couldn't be produced by a straightforward $\text{SU}(2)_L$ theory. To explain this new development and also preserve the parity violating structure, theorists were forced to conclude that the observed Z boson is in fact a mix of the neutral weak boson W^3 and a neutral U(1) boson (call it B). This also means that the photon must be a mix of W^3 and B orthogonal to the Z, and that the weak force and electromagnetism are in fact not distinct phenomena, but

[2]Since $U^\dagger U = e^{-ie\alpha}e^{ie\alpha} = 1$.

rather different aspects of a common $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)$ *electroweak* force. The U(1) charge is the weak hypercharge Y, but by construction the observed photon and its coupling to electric charge is preserved after this process of electroweak unification [20–22]. The mixing of the bosons is governed by the weak mixing angle θ_W, and the observable W, Z and photon are:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp W_\mu^2 \right) , \tag{2.7}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W , \tag{2.8}$$

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W . \tag{2.9}$$

2.2 Electroweak symmetry-breaking and the W boson mass

The final piece of the SM's mathematical formalism is the Higgs mechanism of spontaneous electroweak symmetry breaking, which gives a viable mechanism for the generation of mass. Without it, the experimental fact that the weak bosons have mass cannot be reconciled with gauge invariance. It also resolves the issue that the naive, gauge-invariant fermion mass terms present in Eqn. 2.6 must be replaced by terms that are not gauge invariant once the handedness of the weak force is introduced.

As Eqn. 2.6 illustrates, mass terms are quadratic in the fields, so a massive photon field would have a term proportional to $A^\mu A_\mu$. Applying the transformations introduced above would show that such a term is not gauge invariant, but this is not a problem for the photon, since it is known empirically to be massless. However, the weak force is short-range, so must have massive bosons, and by now the (non-zero) masses of the W and Z are well-known. The problem with the fermion mass terms is as follows. To capture mathematically that the charged-current weak force couples to left-handed particles, the left-handed quarks and leptons are placed into $\mathrm{SU}(2)_L$ doublets L, whereas their right-handed counterparts are placed in singlets R:

$$L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u_L \\ d_L \end{pmatrix} , \quad R = e_R \quad \text{and} \quad u_R \quad \text{and} \quad d_R . \tag{2.10}$$

There is one such set of fields for each of the three generations. Note that since neutrinos only interact via the weak force and the weak force only interacts with left-handed particles, right-handed neutrinos either do not exist, or are completely sterile in

the SM. The first-generation SU(2) doublets and singlets transform as

$$\begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \rightarrow e^{\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{\alpha}(\boldsymbol{x})} \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} , \quad e_R \rightarrow e_R \, , \tag{2.11}$$

which mixes up the left-handed electron and neutrino fields[3], as is to be expected from a theory that couples charged leptons and neutrinos at a vertex. The fermion mass term $m\overline{\psi}\psi$ therefore must become $m\overline{L}L$ and $m\overline{R}R$, and the mixing cross terms of the left-handed fields brought about by the gauge transformation do not cancel out, and so these mass terms are not gauge invariant. Without any further addition to the SM Lagrangian, the only solution is to have massless quarks and leptons.

The Higgs mechanism [23–25] provides a way around these issues by proposing that the gauge symmetry of the SM was spontaneously broken in the very early universe (around 10^{-12} seconds after the Big Bang); that is to say that the theory exhibits the invariance, yet the ground states and fluctuations about them do not. The mechanism introduces a complex scalar Higgs doublet ϕ, which has a contribution to the Lagrangian,

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 , \tag{2.12}$$

which can be made invariant under $\text{SU}(2)_L \otimes \text{U}(1)$ gauge transformations with appropriate weak hypercharge and isospin. The final two terms correspond to $-V(\phi)$ - the Higgs potential. This potential has the familiar "Mexican hat" shape, which features minima at non-zero values of the field ϕ (the recesses of the hat) if $\mu^2 < 0$ and $\lambda > 0$, defined by $|\phi_0| = v = \sqrt{-\mu^2/2\lambda}$, also known as the *vacuum expectation value*. Fluctuations of the field (read: Higgs bosons) will be about the ground state of the field, which will be one of these minimum points. The consequences of this are as follows. Without a loss of generality, the minima can be chosen to be

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{2.13}$$

If the covariant derivatives in Eqn. 2.12 are expanded out to show the contributions due to the U(1) B_μ and SU(2) W_μ fields, the Higgs Lagrangian is

$$\mathcal{L}_{\text{Higgs}} = \left| \left(i\partial_\mu - ig\frac{\tau_a W_\mu^a}{2} - ig'\frac{B_\mu}{2} \right) \phi \right|^2 - V(\phi) \, . \tag{2.14}$$

Now, introducing the ground state of the form given by Eqn. 2.13, the first term

[3] $\boldsymbol{\tau}$ is a vector of three 2×2 non-diagonal matrices - one for each boson - and the exponential can be Taylor-expanded to show that the matrices will mix the electron and neutrino in their doublets.

expands as

$$\left| \left(-ig\frac{\tau_a W_\mu^a}{2} - i\frac{g' B_\mu}{2} \right) \phi_0 \right|^2 = \left(\frac{1}{2} vg \right)^2 W^+{}_\mu W^{-\mu} + \frac{1}{8} v^2 \left(g^2 + g'^2 \right) Z_\mu Z^\mu + 0 \cdot \left(g^2 + g'^2 \right) A_\mu A^\mu \,, \tag{2.15}$$

which is written in terms of the observable electroweak states for the $W^\pm$, Z and photon. Since a gauge boson mass term has the general form $\frac{1}{2}m^2 A^\mu A_\mu$, the mass of the physical W boson can be identified as $m_W = \frac{1}{2}vg$, where g is the SU(2) coupling strength (g' is the U(1) coupling strength). This Lagrangian is no longer gauge-invariant, because the symmetry has been broken by introducing the non-zero ground state, yet the theory as a whole, before introducing the ground state, is still gauge-invariant. A gauge-invariant term can also be added that describes interactions between the Higgs and the left- and right-handed fermions as

$$\mathcal{L}_{\text{fermion masses}} = -g_e \left(\overline{L} \phi R + \overline{R} \phi^\dagger L \right) \,, \tag{2.16}$$

which, plugging in the ground state of Eqn. 2.13, expands to

$$\mathcal{L}_{\text{fermion masses}} = -g_e \left(\overline{e_L} v e_R + \overline{e_R} v e_L \right) \,, \tag{2.17}$$

which can be manipulated to give $-g_e v \bar{e} e$: a fermion mass term, with mass proportional to the coupling strength g_e and the vacuum expectation value. The coupling g_e is known as the Yukawa coupling, and encodes the different masses of the different fermions.

This completes the most salient parts of the SM Lagrangian. The Yukawa couplings of the fermions and the coupling strengths of the forces are not predicted by the theory - it only gives relations between them. Measuring these properties and testing their consistency with the SM is therefore a primary goal of the field. The discussion above was general and placed no preference on one of the three generations of quarks and leptons. This assumption in the SM that each generation is treated equally must also be tested with experiment. Such tests as introduced here are the topics of Chapters 8 and 9. Making such analyses is not simple however, particularly at high precision and with a hadron collider. The following sections therefore build on the SM framework laid down here, and examine the consequences of applying it to high-precision electroweak measurements.

2.3 Higher-order corrections in perturbative QED/QCD

From the SM Lagrangian and the Euler-Lagrange equations, the mathematical crank can be turned to derive the "Feynman rules" of the SM gauge interactions. The Feynman rules are the link between the SM Lagrangian (and its fundamental parameters), and observable quantities such as decay rates and cross sections. They instruct how to draw interactions pictorially, as so-called Feynman diagrams, and then provide rules for mapping diagrams to "amplitudes" $\mathcal{M}_{fi}$ between the final and initial states f and i. Then all that remains to be done is to sum the amplitudes from all possible diagrams between f and i, and put this result into a formula that takes care of the kinematics. For example, in the centre-of-mass (COM) frame, the cross section of *any* "$2 \to 2$" process such as $e^+e^- \to q\bar{q}$ is given by

$$\sigma = \frac{1}{64\pi^2 s}\frac{p_f}{p_i}\int |\mathcal{M}_{fi}|^2 d\Omega\,, \tag{2.18}$$

where p_f and p_i are the momenta of the initial and final states[4], $\sqrt{s} = (E_1 + E_2)$ is the centre-of-mass energy, $d\Omega$ is an element of the solid angle and $\mathcal{M}_{fi}$ is the total complex amplitude. The cross section is a Lorentz-invariant quantity, so can be calculated in the (mathematically convenient) COM frame and will apply to any other frame, such as the lab frame of an LHC experiment.

Reiterating the above, to get a perfect SM prediction, *all* Feynman diagrams must be summed, but this is impractical - in principle, Feynman diagrams can be drawn with infinite complexity. Fortunately, for QED and - at high collision energies such as those at the LHC - QCD, the coupling strength of the interaction is much smaller than 1, such that diagrams of higher complexity, which will have more interaction vertices and more factors of the coupling, contribute less to the total amplitude. Thus, perturbative methods can be used to make the prediction, and only the first one or two orders in coupling strength may be sufficient depending on the process under study. In high-precision EW physics, next-to-leading order (NLO) predictions - corresponding to two powers of the strong coupling α_s - are the benchmark, and in some cases even higher orders are commonplace.

Going from leading order (LO) to NLO has a number of consequences, as will be initially illustrated with $e^+e^- \to q\bar{q}$. The quarks themselves are not observed in experiment, but rather a shower or *jet* of hadrons confining quarks created by the hadronization process and a number of gluon emissions from the high-energy quark. It is therefore relevant to consider adding a gluon emission to $e^+e^- \to q\bar{q}$, giving the diagram shown in Figure 2.2 (left). This is LO in QCD, but even so, after applying Feynman rules, the amplitude for

[4]In the centre-of-mass frame, $p_i = p_{i,1} = p_{i,2}$ is the same for each incoming particle. The same can be said for the final-state particles.

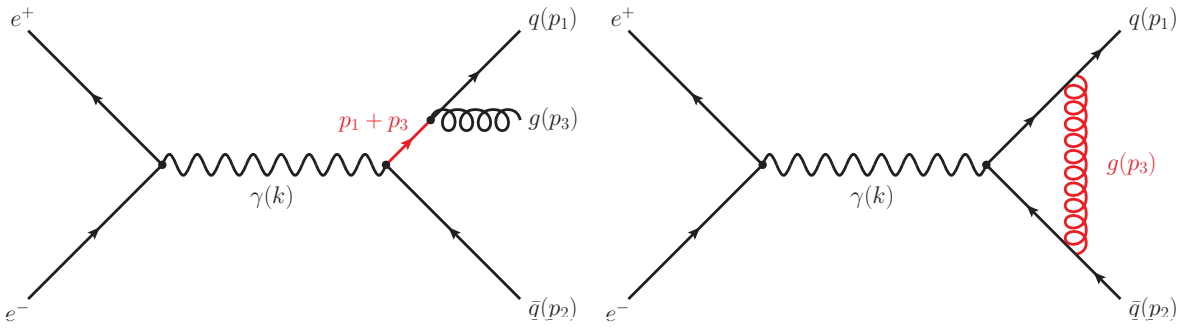


Figure 2.2: Feynman diagram for the $e^+e^- \to q\bar{q}g$ process (left) and $e^+e^- \to q\bar{q}$ process with one gluon loop (right). Produced with the JaxoDraw tool [26].

this diagram will contain a factor $1/(p_1 + p_3)^2 = 1/2E_1E_3(1 - \cos\theta_{13})$ due to the internal fermion line, or *propagator*. If the gluon becomes collinear to the quark ($\theta_{13} \to 0$), or goes to infinitesimally low energy ($E_3 \to 0$), this amplitude will diverge. This is known as an *infrared (IR) singularity*. However, it is known that this process has a finite probability, so such divergences are clearly unphysical. The solution comes at NLO, with diagrams such as Figure 2.2 (right). This is typically referred to as "virtual" gluon emission, while the previous diagram is called "real" emission. Since the gluon momentum is unknown, it can be infinitesimally small, which also leads to a divergence. Thankfully, the two diagrams interfere in such a way as to cancel the divergences in a complete NLO amplitude. In fact, this appearance and resolution of these divergences is general: the "KLN" theorem [27, 28] states that, for any physical variable - e.g. a differential or total cross section - IR singularities should cancel in this way. This property of physical variables is called *infrared safety*.

Diagrams such as Figure 2.2 (right) are not only infrared-divergent, but also unphysically *ultraviolet*-divergent, i.e. when the gluon momentum becomes large. This manifests because the amplitude for such diagrams will contain an integral over all possible gluon momenta, with typically more powers of the gluon momentum in the numerator than in the denominator, leading to a divergence as the gluon momenta is sent to infinity. The solution to this divergence is to *renormalize* the strong coupling, by adding so-called "counter terms" to the Lagrangian, which themselves may be infinite, but which cancel the divergences, leaving the overall calculation finite. Another effect of these counter terms is to shift the values of the parameters such as couplings and masses. Physically, this corresponds to "dressing" the propagators with interactions with the vacuum, which produce a screening effect on the strength of the interaction. What is measured in nature is the resultant finite couplings and masses, while the "bare" charges and masses that the theory started from were in fact infinite and unphysical.

A theory is renormalizable if a finite number of these counter terms can be added

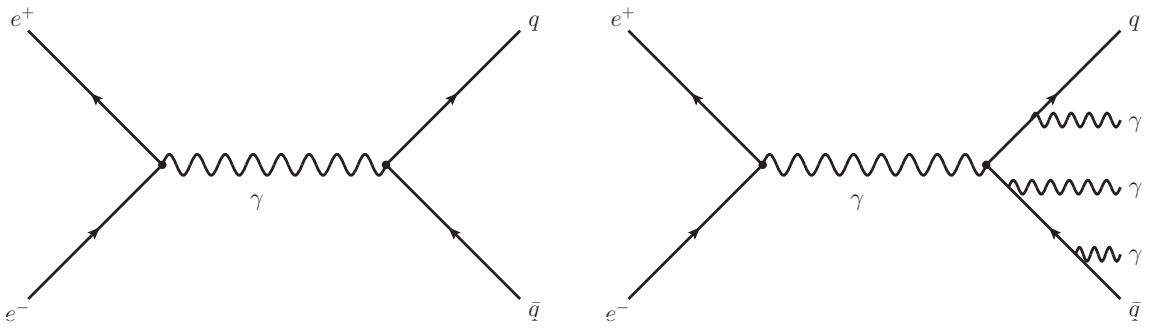


Figure 2.3: Feynman diagrams for the $e^+e^- \to q\bar{q}$ process without (left) and with (right) final-state radiation. Inspired by a similar figure in Ref. [5]. Produced with the JaxoDraw tool [26].

- at every order of perturbation theory - that will cancel all the ultraviolet divergences. All the QFTs in the SM are renormalizable, and the method for working out these counter terms is called the renormalization *scheme*, a common one being the $\overline{\text{MS}}$ scheme. Renormalization introduces a scale to SM calculations that must be specified, and is typically picked to be the highest relevant scale in the process, e.g. m_Z for $Z \to \mu\mu$, or the p_{T} of the leading hadron in a jet. As the perturbative order of the calculation increases the renormalization scale dependence reduces, and the choice of scale would have no effect on an all-orders calculation. Varying the renormalization scales in a calculation therefore gives an estimate of its uncertainty due to missing higher orders. Renormalization also causes the couplings to "run" (i.e. change) with the energy scale Q. In QED, the coupling α increases with energy, but for QCD the opposite behaviour is observed. It evolves as

$$\alpha_s(Q^2) = \frac{\alpha_s(Q_0^2)}{(1 + B\alpha_s(Q_0^2)\ln(\frac{Q^2}{Q_0^2}))}\,, \tag{2.19}$$

with $B = \frac{11N_C - 2N_f}{12\pi}$. N_C is the number of colours, and $N_f = 6$ is the number of quark flavours. $N_C = 3$, and $N_f = 6$, so $B > 0$ and α_s decreases with Q^2. At Q^2 of 1 GeV, $\alpha_s \sim 1$, making perturbation theory intractable and encapsulating why QCD at low energies is so difficult - for example there has been no success in "calculating" hadronization. However, at collider energies e.g. $Q^2 \sim M_Z^2$, $\alpha_s \approx 0.118$, so perturbation theory is valid, although perturbative series do not converge as quickly as e.g. QED, where $\alpha_{\text{QED}} \sim 1/128$ at m_Z. This property of QCD is called *asymptotic freedom*: at high-enough energies the quarks and gluons in the proton can be treated as quasi-free particles that are only weakly bound within hadrons. The value of α_s, the renormalization scale and missing higher orders are deeply connected, such that varying α_s can also help to account for the missing higher orders in a QCD calculation that are so important in a hadron collider process.

The final type of higher-order correction that is relevant here is QED final-state radiation (FSR). The final states of interest here are charged leptons, which - being electrically charged - radiate photons. Predictions for a process can be made with or without this dressing of the final-state charged leptons: without is known as the "Born" level, and with dressing of final-state radiation is known as the "Bare" level. The latter is what is measurable in a real detector. The distinction is shown in Figure 2.3. Since this dressing of the final-state particles factorizes from the rest of the event, dedicated computational programmes can be used to model just the final-state radiation.

2.4 Parton distribution functions and QCD factorization

At a proton-proton collider, the fundamental interaction that occurs is the hard (i.e. high-momentum) scattering of two partons (quarks and gluons) - one from each proton. The proton is however a dynamic sea of confined partons strongly-interacting with one another, which has a number of consequences for the interactions that are observed. The first consequence is that, although it is possible to prepare a beam of protons such that each one has almost exactly e.g. 6.5 TeV of energy at the LHC, how this energy is shared around between the partons is inherently probabilistic, and quantified in terms of the Bjorken x. The probability of a parton of flavour i having a given value of x is given by the Parton Distribution Function (PDF) $f_i(x)$. At high energies, the property of asymptotic freedom allows the PDFs to mathematically factorize from the partonic hard-scattering process in a principle known as QCD factorization. Therefore, any observable, for example a cross section, can be written in terms of the partonic cross section for two partons of a given flavour and given momentum fractions x_1 and x_2, and then integrated over all possible flavours and momentum fractions, weighted appropriately by the PDFs:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, {\mu_\mathrm{F}}^2) f_j(x_2, {\mu_\mathrm{F}}^2) \sigma_{ij}(p_1, p_2, \alpha_s({\mu_\mathrm{F}}^2), Q^2, {\mu_\mathrm{F}}^2) + \mathcal{O}({\Lambda_\mathrm{QCD}}^2/Q^2)\,, \tag{2.20}$$

with higher-order corrections arising as a function of $O({\Lambda_\mathrm{QCD}}^2/Q^2)$ [29]. Λ_QCD is the scale at which QCD becomes non-perturbative and the factorization breaks down, and μ_F is the *factorization scale*, defined shortly.

The principle of QCD factorization hinges on the separation between the long-timescale, low-energy physics within the proton and the short-timescale, high-energy physics of the hard scattering. This separation is clear for a process like $e^+e^- \to q\bar{q}$, where the quarks hadronize over some comparatively long timescale after the interaction.

However, in a process like deep inelastic scattering (DIS: $e^- p \to e^- X$, which is actually the elastic scattering $e^- q \to e^- q$), the factorization is spoiled by the different weighting of the incoming and final-state quarks via the PDFs. The incoming quark could also radiate an arbitrarily soft (low-momentum) gluon, which is in conflict with QCD factorization's assertion that low-energy physics is confined to the PDFs, not the hard scattering.

The solution to this is to define the factorization scale μ_{F}, below which all gluon emissions are packed into the definition of the PDF, and therefore any long-distance collinear emission would have happened long before the collision. Calculations of the hard process can then integrate from the factorization scale to infinite momentum, with the softest and most collinear emissions (and their IR divergences) removed. Like the renormalization scale, this introduces a factorization scale dependence to the calculations, as well as the PDFs themselves. PDFs therefore evolve with energy, and this evolution is controlled by the DGLAP evolution equations [30–32]. A final point here is that while QCD factorization is process-independent, it will be ruined by any initial-final-state interference.

Since PDFs arise from low-energy interactions within the proton, they are not calculable in pQCD - α_s at this scale is too large - and they must be inferred from data. Historically, PDF determination was done in DIS experiments (e.g. at HERA at DESY), where a high-energy electron is used as a probe of the proton's structure. Neutrino scattering from nucleons gives complementary information. Today, PDFs are extracted from fits to a large number of datasets from different experiments, by multiple fitting groups, each group yielding a different PDF "set".

The proton PDFs as recently extracted by the NNPDF group [33] at two different energy scales are shown in Figure 2.4 as a function of Bjorken x. The "valence quark" (u and d in the proton) PDFs dominate at most values of x, but as $\sqrt{s}$ increases the gluon PDF becomes stronger, and consequently the LHC is essentially a gluon collider most of the time. "Sea quarks" arise from $g \to q\bar{q}$ processes in the proton, whose amplitude has a $1/q^2$ propagator and therefore the sea quarks PDFs fall off steeply with momentum. The sea however dominates the quark contribution at low x, while it is small at high x. In the end, since the PDFs are probability densities that describe how momentum is shared, they must sum appropriately to give the total proton momentum. This constraint, known as a momentum sum rule, is given by

$$\sum_i \int_{x=0}^{x=1} x f_i(x) \mathrm{d}x = 1\,. \tag{2.21}$$

That the momentum sum rules (others can be written, giving different information [34]) constrain the PDFs is extremely useful when it comes to combining the results from experiments that have different kinematic acceptances. The LHC and its main

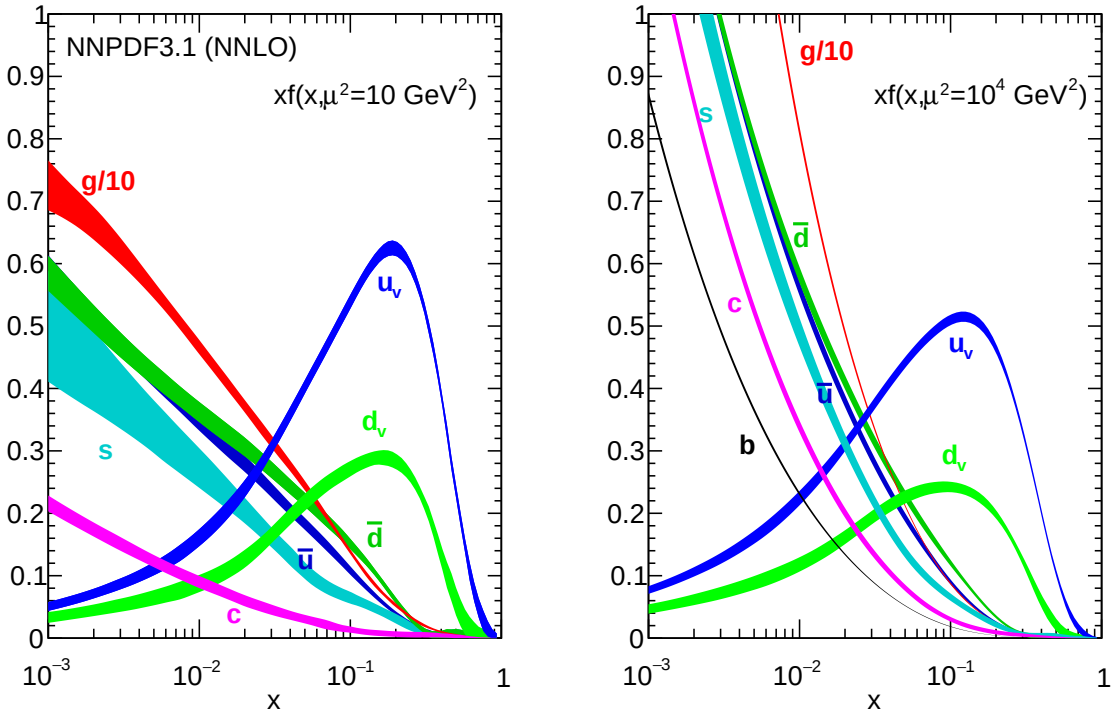


Figure 2.4: The NNPDF3.1 NNLO PDFs, evaluated at scales ${\mu_F}^2 = 10$ GeV^2 (left) and ${\mu_F}^2 = 10^4$ GeV^2 (right). The width of the bands shows the uncertainty. Taken from Ref. [33].

experiments will be introduced in the next chapter, but for now it is necessary to know that the LHCb experiment [4] is instrumented in the forward region ($2 < \eta < 5$), while ATLAS [35] and CMS [36] cover the central angles ($|\eta| \lesssim 2.5$) around the pp collision point. Collisions with decay products boosted forward into LHCb tend to be highly asymmetric, with one low-x and one high-x parton, while ATLAS and CMS have more equitable collisions on average, and so the different detectors probe different areas in the PDF kinematic space, as shown in Figure 2.5 as a function of x and the squared momentum transfer Q^2. Comparable, PDF-sensitive measurements in LHCb will therefore (to some degree) anti-correlate with those made in ATLAS or CMS - if the PDFs are stronger at low and high x, the sum rule constrains that they must be weaker at moderate x. The degree of anti-correlation will depend on factors such as how orthogonal and how broad the η acceptances are, and how important the PDFs are in the measurement, but in general the combination of anti-correlating measurements gives a smaller total uncertainty than the combination of two similar, positively-correlating measurements. These principles were part of the inspiration for a hitherto-unconsidered measurement of the W boson mass at LHCb, as will be described fully in Chapter 8.

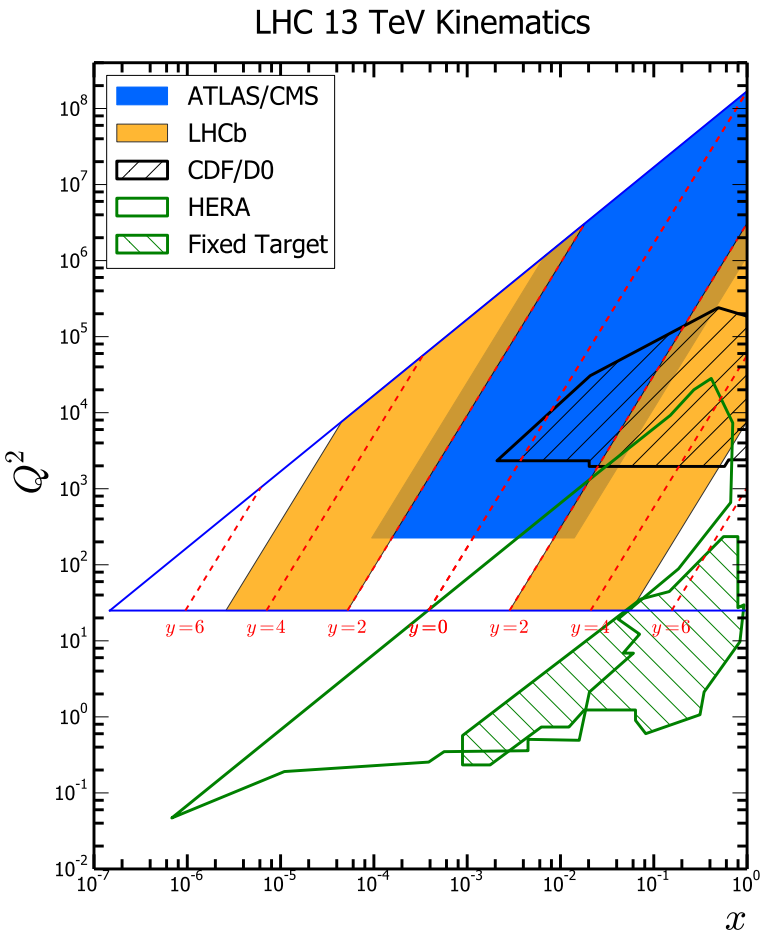


Figure 2.5: Deep inelastic scattering kinematic acceptance in the space of Bjorken x and momentum transfer Q^2 for the main modern experiments. Adapted from [37], based upon [38].

2.4.1 PDF uncertainties

Since PDFs are extracted from fits to data, there is uncertainty in them coming from the extraction procedure. This uncertainty will propagate as a PDF-related uncertainty to the measurement of any observable (call it F) which subsequently uses the PDFs. The PDF uncertainty on F can be assessed crudely by simply changing the PDF set, but if greater control is required, there are methods to assess the PDF uncertainty coming from a single PDF set. In this thesis, two such methods are relevant: the *Hessian* [38, 39] method and a method based upon using PDF "replicas" [40]. In the former case, the PDF uncertainty on F is evaluated by re-performing measurements using a set of PDFs associated with the eigenvectors of the Hessian matrix (the inverse covariance matrix of the PDF extraction), and combining them using a *master formula* [39]. The eigenvector PDFs characterize the PDF parameter space and its uncertainties, and reflect a conventional PDF fit where a functional form is proposed and its many parameters are fitted to the data. The latter

approach reflects a PDF extraction where a statistical ensemble of PDF "replicas" has been trained by machine learning methods, and the variance in the ensemble represents the PDF extraction uncertainty. In this case, the PDF uncertainty on F is simply extracted by re-performing the measurement a large number of times with different PDF replicas, and taking the RMS of the resulting values of F [40].

The value of α_s is also an input parameter to the PDF fits, but rather than include these as part of the parametric PDF uncertainty, the PDF collaborations provide PDF sets with different, preferred values of α_s used in the extraction. The PDF4LHC group provides recommendations for how these α_s-varied PDF sets should be used to calculate a PDF uncertainty. In the case of a single-parameter measurement like a total cross section (again using F), fits should be performed using PDFs with values of α_s varied up and down by their 1σ uncertainty, and the PDF-α_s uncertainty is then half the absolute difference between the upper and lower fitted values of F. The total PDF uncertainty is then the quadrature sum of the α_s-related uncertainty and the parametric uncertainty [41].

2.5 Precise theoretical predictions for the Drell-Yan process

The high precision EW physics studies detailed in this thesis are all with Drell-Yan process [43] $pp \to V + X \to \ell\ell + X$, where V is a vector boson (Z/γ^* - hereafter just Z - or $W^{\pm}$), $\ell\ell$ is a lepton pair (charged lepton and a neutrino if W, two charged leptons if Z) and X is the remnants of the rest of the pp collision. A leading-order Feynman diagram of the hard scattering process between the two quarks, along with a schematic of the whole pp collision, is shown in Figure 2.6.

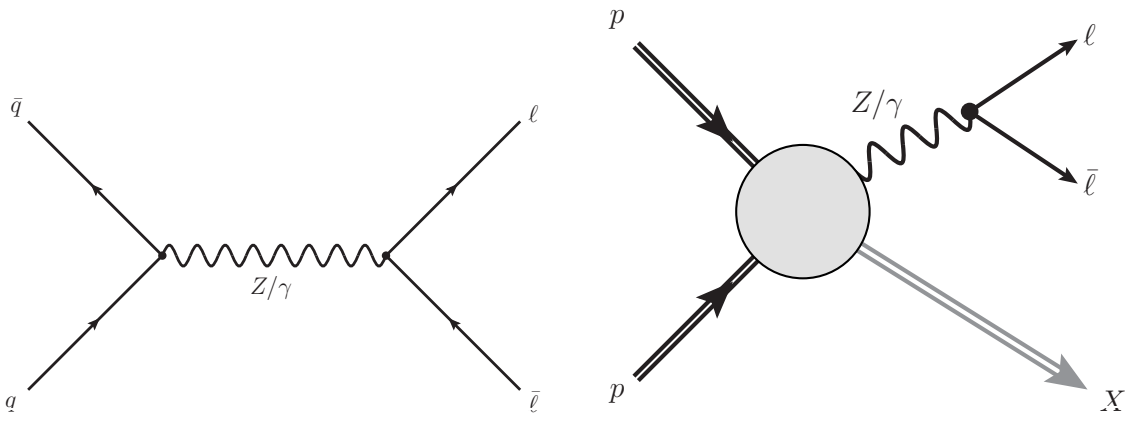


Figure 2.6: The leading-order Feynman diagram for Drell-Yan hard scattering process (left) and a schematic of the whole pp collision in Drell-Yan events (right). The latter is inspired by Ref. [42]. Produced with the JaxoDraw tool [26].

In terms of QCD, Drell-Yan is a "$2 \to 0$" process, in that the initial state has two coloured partons but the final-state particles are colourless. It is therefore the simplest process involving QCD at hadron colliders, and is correspondingly theoretically very well understood. Charged leptons produce clean final states, and the cross section is also very high on the scale of SM processes, meaning there are large data samples available for analysis. This combination of factors makes Drell-Yan the perfect process to stringently test the SM.

These tests typically proceed via *template fits*. It is typically not possible to derive analytic forms for the distributions that are being fitted to make precision EW measurements, so instead these distributions are simulated with numerical methods, using the best knowledge of the underlying physics, and fixed values of the physics parameters of interest. The simulated distributions are called *templates*, and in a template fit for a hypothetical parameter x, many different templates with differing values of x are produced, and the best-fitting template to the data corresponds to the best-fitting value of x. The accuracy of the predictions used to make the templates is therefore of paramount importance. Both physics measurements presented in this thesis are template fits to the kinematic variables of leptons in Drell-Yan, so this section covers the various techniques and theoretical considerations necessary to make accurate and precise predictions of this process.

The relevant numerical calculations here heavily utilize Monte-Carlo integration: a technique that estimates an integral by statistical sampling of the integration variables. Monte-Carlo integration has the useful property that the error on the prediction (which relates directly to the number of samplings and hence CPU time cost) does not scale with the number of integration dimensions, which is ideal in HEP setting where there are often many degrees of freedom in the problem.

Fixed-order (e.g. LO, NLO in α_s) predictions of templates for Born-level distributions such as differential cross sections can be made in this way. The phase space degrees of freedom are sampled and at each point the transition probability/amplitude can be calculated by summing up Feynman diagrams to a certain fixed order, eventually giving an observable such as a differential cross section after enough sampling. This is the approach taken by generators such as POWHEG [44, 45], PYTHIA [46] and HERWIG [47]. In the Drell-Yan process, fixed-order generators at NLO describe the inclusive cross section and the rapidity distribution (differential cross section binned in boson rapidity) well, but have difficulty with the boson p_{T} distribution at low p_{T}. At LO, as shown on the left of Figure 2.6, the vector boson has zero p_{T}, but at higher orders diagrams including e.g. initial-state gluon emission enter, which the vector boson recoils against, thereby giving a non-zero p_{T}. Similar to the breakdown of QCD factorization in DIS, the low-p_{T} region of the p_{T} distribution includes diagrams with soft gluon emission, which

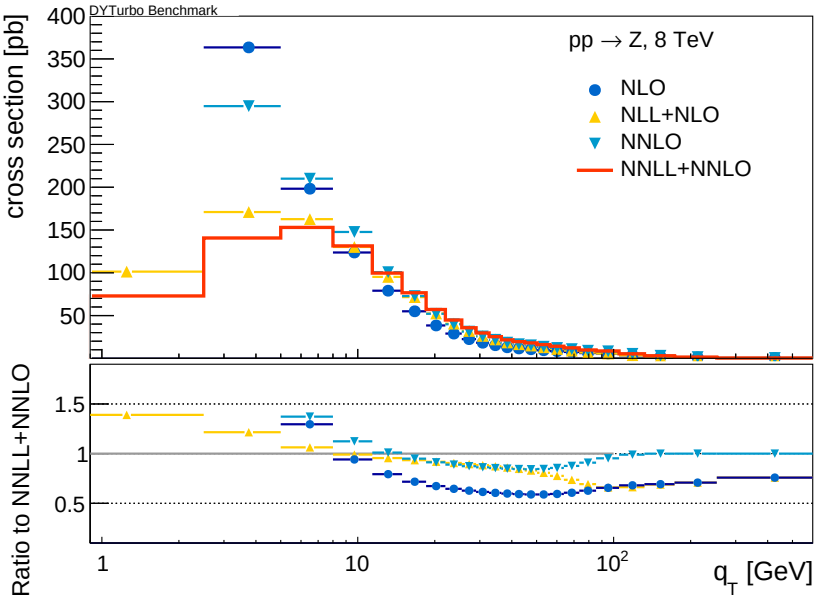


Figure 2.7: Predictions from DYTurbo of the p_{T}^{Z} distribution at $\sqrt{s} = 8\,\mathrm{TeV}$ at various orders of the calculation (NLO, NLL+NLO, NNLO, and NNLL+NNLO). The bottom panel shows ratios of results at various orders to the NNLL+NNLO result. The fixed-order results (NLO and NNLO) diverge at low p_{T}^{Z}, which is resolved by resummation. Taken from Ref. [48].

by QCD factorization should be regulated by the PDFs. This breakdown can be seen mathematically in the 3D N^{p}LO differential cross section far below the factorization scale:

$$\lim_{p_{\mathrm{T}}^{V} << \mu_{\mathrm{F}}} \frac{d\sigma_{pp\to\ell\bar{\ell}+X}}{dQ^2 dy dp_{\mathrm{T}}^{V}} \propto \sum_{n=1}^{p+1} \alpha_s^n(Q) \sum_{k=0}^{2n-1} I^{[n,k]} \frac{1}{p_{\mathrm{T}}^{V\,2}} \ln^k\left(\frac{Q}{p_{\mathrm{T}}^{V}}\right) + \mathcal{O}\left(\frac{p_{\mathrm{T}}^{V\,2}}{Q^2}\right). \tag{2.22}$$

This equation shows that, at each order of perturbation theory, there is a series of logarithms that diverge at low boson p_{T}. The first log in the series is called the "leading-log" (LL), the next "next-to-leading-log" (NLL) etc. If this expression is truncated to any fixed order of perturbation theory, it will give inaccurate predictions in the low p_{T} region, as these logarithms enhance the higher-order corrections there. This problem can be circumnavigated by instead using the technique of *analytic resummation*. At successive orders of perturbation theory, the LL terms can be summed via a Taylor series, as can all the NLL terms and so on. These resummations can be matched to a fixed-order calculation, e.g. at NLO, removing the divergence at low p_{T}, and giving a "NLO+NLL" calculation. For example, the DYTurbo program [48] can provide such predictions up to NNLO+NNLL accuracy. The effect of resummation at low p_{T} is showcased by DYTurbo in Figure 2.7.

A useful feature of Drell-Yan in terms of making predictions is how its differential cross section can be factorized into angular and unpolarized terms, allowing for the possibility of using different programs to predict either part. The vector boson production

and decay has five degrees of freedom: six initially (three components of momentum for each parton), but one is removed due to the cylindrical symmetry of the interaction about the beam axis. The remaining five are typically parametrized in terms of the vector boson's mass M, rapidity y and transverse momentum p_{T}^{V}, and two decay angles defining the direction of the final-state leptons. The angular distribution of the lepton system is influenced by the polarization of the vector boson, which can be factorized from the unpolarized part of the differential cross section σ^{unpol}, giving the full 5D differential cross section as

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}^{V}\mathrm{d}y\mathrm{d}M\mathrm{d}\cos\vartheta\mathrm{d}\varphi} = \frac{3}{16\pi}\frac{\mathrm{d}\sigma^{\mathrm{unpol}}}{\mathrm{d}p_{\mathrm{T}}^{V}\mathrm{d}y\mathrm{d}M} \\ \big\{(1+\cos^2\vartheta) + A_0\frac{1}{2}(1-3\cos^2\vartheta) + A_1\sin 2\vartheta\cos\varphi \\ + A_2\frac{1}{2}\sin^2\vartheta\cos 2\varphi + A_3\sin\vartheta\cos\varphi + A_4\cos\vartheta \\ + A_5\sin^2\vartheta\sin 2\varphi + A_6\sin 2\vartheta\sin\varphi + A_7\sin\vartheta\sin\varphi\big\},\end{aligned} \tag{2.23}$$

where ϑ and φ are the lepton decay angles defined in the Collins-Soper [49] rest frame of the lepton pair [42]. The angular distribution has been written in a basis of spherical harmonics, each term having an *angular coefficient* A_i, which themselves are ratios of helicity cross sections and depend on p_{T}^{V}, y and M. At LO only A_4 is present, which corresponds to the forward-backward asymmetry A_{FB}. At NLO, diagrams arise that give $A_0 - A_3$ non-zero values. The coefficients $A_5 - A_7$ are numerically small because they only arise at NNLO or higher, and all except A_4 tend to zero at low p_{T}^{V}. This equation is only valid at the Born level (neglecting QED FSR), and if there are no kinematic cuts on the leptons.

The preceding paragraphs describe how predictions can be made for observables such as partonic cross sections at the Born level, but not a detector-level observable that includes the effects of FSR or experimental resolutions. To get full events, with which templates can be constructed in any observable to compare directly with experiment, *parton shower* algorithms are typically used. These begin with Monte-Carlo simulation of the hard process, which produces coloured partons, then each parton is probabilistically allowed to split and radiate via QCD processes, the products of which themselves split and radiate and so on, developing a shower of soft emissions. The shower resembles the collimated jets that are observed in experiment, with the dominant emissions happening earlier and preferentially in the direction of the hard parton, and ending with the softest emissions before hadronization. Since any number of splittings is possible, a parton shower is therefore an "all-orders" prediction for the event, and their final-state decay products can be propagated through a detector simulation to also account for resolution

effects. Large samples of these events are used to build templates. However, many choices are written into parton shower algorithms, and they begin with a hard process that is restricted to a certain perturbative order, so there are many sources of uncertainty that can propagate to the results of a template fit. To mitigate this, parton showers are matched to fixed-order calculations of high-level observables like cross sections. For example, matching a Pythia parton shower to a NLO calculation from Powheg gives a "NLO" event. Alternatively, since the hard process and parton shower factorize, it is possible to mimic the presence of a higher-order event generator by reweighting the hard process kinematics to better predictions. This technique will be used in the analyses here, as detailed in Chapter 5.

CHAPTER 3

The Large Hadron Collider Beauty experiment

The physics measurements presented in this thesis are all performed using data collected by the Large Hadron Collider Beauty experiment, LHCb. A supreme understanding of the behaviour and potential biases of the experiment are required to conduct a robust high-precision measurement. This chapter therefore describes the LHCb experiment, with particular focus on the those aspects that are most salient to the rest of the thesis. The chapter begins with a brief introduction (taken mostly from Refs. [50] and [7]) to the Large Hadron Collider (LHC), which provides the high-energy beams of protons and collides them within LHCb. Section 3.2 introduces the LHCb experiment. The detector can be broadly categorized into a tracking system and a particle identification system, which are the subjects of Sections 3.3 and 3.4. Ref. [4] was the key source here. This is followed by Section 3.5, which details the LHCb trigger and data processing, mostly taken from Ref. [51]. A few highlights of the detector's performance relevant to electroweak (EW) analyses is given in Section 3.6. In Section 3.7 there is a short overview of the LHCb simulation chain. Although the majority of the chapter focuses on LHCb as it was during 2016 (when the data used here was taken), the chapter finishes with a brief outlook on the upgrade of the LHCb detector, which sets the scene for Chapter 4.

3.1 The Large Hadron Collider

The Large Hadron Collider is a two-ring, approximately circular, superconducting hadron accelerator and collider with a circumference of 26.7 km. Straddling the French-Swiss border near Geneva, it lies underground at a depth varying between 45-175 m. It primarily collides protons, and in proton-proton (pp) collisions is designed to provide a centre-of-mass energy ($\sqrt{s}$) of 14 TeV, making it the largest and highest energy proton-proton collider ever constructed. It also has an unprecedented design luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$,

enabling the collection of vast amounts of collision data over a lifetime of around 30 years.

The history of the LHC began in 1994, when the project was approved by the CERN council as a successor to the Large Electron-Positron Collider (LEP). Construction began after the shutdown of LEP in November 2000, and LEP's tunnels were used to house the new ring in order to reduce costs. Commissioning of the brand-new LHC began in September 2008, marking the start of the period known as "Run 1", but was quickly derailed by a catastrophic electrical fault which set commissioning back by 14 months. Run 1 restarted in November 2009, with collisions at 7 TeV, pushing up to 8 TeV by 2012. Physics measurements then began to flow, including the famous discovery of the Higgs boson by the ATLAS and CMS experiments, which was announced in July 2012 [2, 3]. Run 1 ended in February 2013, followed by "Long Shutdown 1" (LS1), where preparations to hit the design collision energy and luminosity were made, as well as necessary improvements highlighted by the 2008 electrical fault [52]. "Run 2" then began in April 2015 with $\sqrt{s} = 13\,\text{TeV}$, and in June 2016 the LHC reached its design luminosity of $1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$. 14 TeV collisions were postponed to Run 3, due to longer-than-envisaged magnet preparation times. Run 2 then concluded in December 2018 [53], upon which the LHC went into a second Long Shutdown to prepare for the high-luminosity upgrade of the LHC (HL-LHC) from Run 4. The current shutdown was originally planned to last until 2021, but was delayed into 2022 by the COVID-19 pandemic. At the time of writing, the LHC is providing collisions at $\sqrt{s} = 13.6\,\text{TeV}$, and commissioning of the collider and the experiments housed on its ring at this new energy is ongoing.

Protons had a complex journey to undergo before they eventually ended up in the LHC with 6.5 TeV of energy. Hydrogen atoms in a gas bottle were stripped down to bare protons via an electric field, at which point they entered the LHC "injector chain": a series of old accelerators (once discovery machines in their own right) that increase the proton energy in steps from accelerator to accelerator. The first in the chain was Linac2: a linear accelerator that brings the protons up to 450 MeV of energy before passing them on to a series of circular synchotron accelerators. These start with the Proton Synchrotron Booster (PSB), taking the protons to 1.4 GeV; the Proton Synchrotron takes them to 25 GeV; and finally the Super Proton Synchrotron (SPS) to 450 GeV. At the PS, the proton beam is shaped into a train of bunches - each containing roughly 1.5×10^{11} protons - with a minimum of 25 ns of spacing between each bunch. The LHC is designed to accommodate 2808 such bunches per beam. At 450 GeV, the proton bunches are injected into the LHC.

The layout of the LHC ring is shown in Figure 3.1. Injection takes place at "Points" 2 and 8. The ring - despite its circular moniker - has eight straight and eight curved sections, the latter known as "arcs". In these arcs the beams are bent around the ring by 1232 superconducting dipole magnets. Since the maximum beam energy that can be

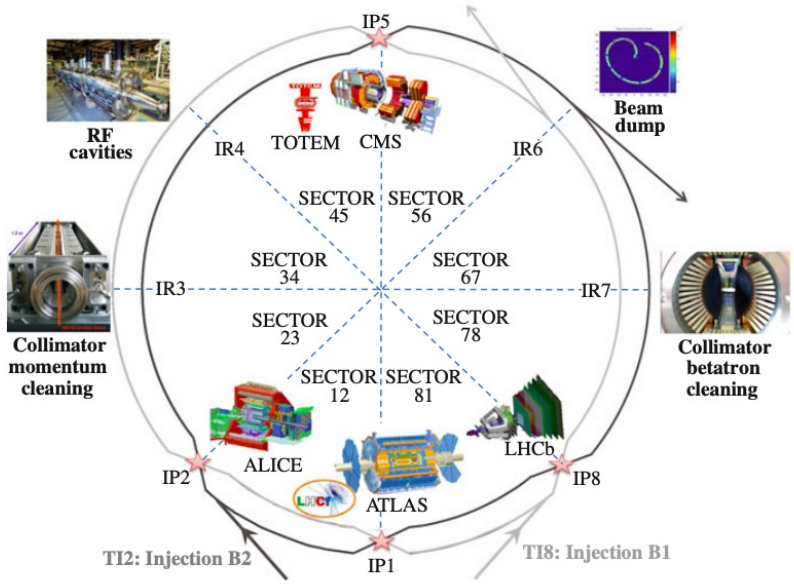


Figure 3.1: Schematic layout of the LHC during Run 2. Taken from Ref. [7].

achieved is determined by the field of the dipoles, an extensive R&D campaign generated dipoles at the frontier of magnet technology, with an achievable peak field of more than 8 T. Such high field requires that they be cooled to 1.9 K using superfluid helium. Also housed in the arcs are quadrupole magnets that focus the beam in the horizontal and vertical directions. The eight straight sections of the ring - each approximately 530 m long - serve as an "insertion region" (IR) for either a particle detector or necessary hardware for the ring itself. The radio-frequency superconducting cavity system is located in IR 4, which focuses the beam longitudinally and provides acceleration that ramps the beam from 450 GeV to its final energy. Collimators at IR3 and IR7 remove protons deviating significantly from the nominal bunch kinematics, and IR6 houses the beam dump system, where the beam can be safely absorbed in stainless-steel-jacketed concrete in the case of operational problems.

The beams circulate in an ultra-high vacuum of below 10^{-13} atmospheres (necessary for their stability and lifetime), with one bunch train moving clockwise around the ring, and the other anticlockwise. The two beam-pipes are separated by 194 mm for most of the circumference, with the bending magnets encasing both beam-pipes in a "twin-bore" design, as shown in Figure 3.2. To maintain an equal length for each beam's path around the ring, the two beams cross over in IRs 1, 2, 5 and 8. It is here that the beams are brought into collision at the four interaction points (IPs). Each IP is enveloped by a particle detector to study the products of the collisions. ATLAS [35] and CMS [36], housed at IP1 and IP5 respectively, are general-purpose detectors mainly designed for direct observation of new particles with large masses. ALICE [54] studies the quark-gluon

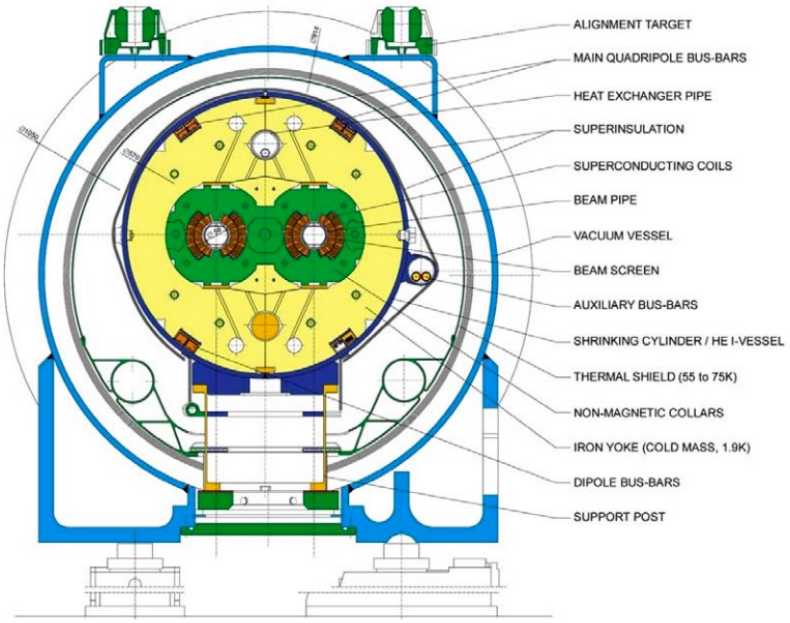


Figure 3.2: Schematic of the cross-section of the LHC beam-pipe in one of the arc sections, surrounded by the superconducting dipole magnet. Lengths are in mm. Taken from Ref. [50].

plasma via lead ion collisions at IP2, and finally LHCb [4] is situated at IP8.

As the beams complete circuits of the ring, the bunches deplete due to collimation and collisions, which leads to an exponential decay in luminosity across an LHC "fill"[1] if the bunches collide head-on, as they did in ATLAS and CMS. Contrastingly, LHCb was able to achieve a stable instantaneous luminosity for the majority of a fill due to the application of the "luminosity levelling technique", as showcased in Figure 3.3. LHCb was designed to operate at the lower luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-12}$ [55], which was achieved by introducing a transverse offset between the two beams at the interaction point, thereby controlling the number of collisions per bunch crossing. As the bunches depleted, this offset was reduced, thereby keeping the number of collisions constant. This technique facilitated the collection of three times more data than originally foreseen in Run 1. As well as enabling luminosity levelling, the lower luminosity had other benefits: it gave stable data-taking conditions for LHCb much earlier than ATLAS and CMS; less radiation damage was inflicted on the detector; and events were easier to reconstruct and analyse - an average LHCb event in Runs 1 and 2 featured only a single pp interaction [56].

[1] A single period of collisions delimited by the announcement of stable beam conditions and the dumping of the beam by the LHC [51].
[2] Although it actually ran at twice this in Run 2 [53].

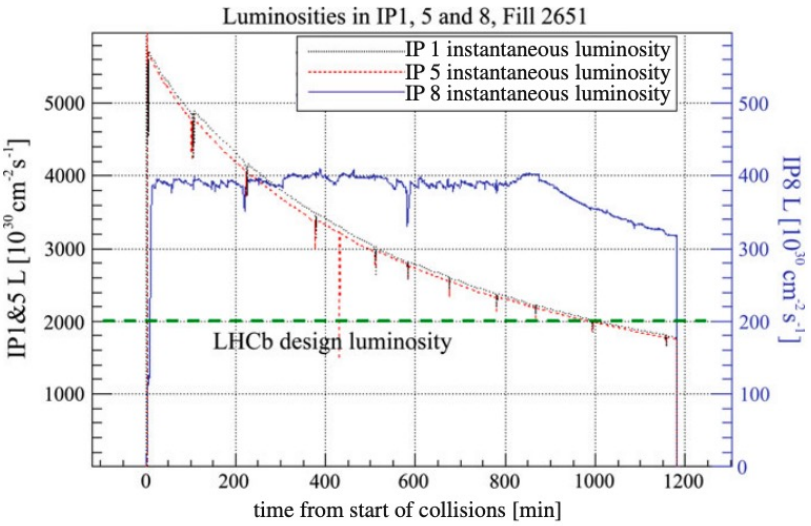


Figure 3.3: Instantaneous luminosity during a typical fill for ATLAS (IP1), CMS (IP5) and LHCb (IP8). When the transverse offset between the beams is fully removed in IP8 the beams collide head-on and the luminosity decays exponentially as in IP1 and IP5. Taken from Ref. [7].

3.2 Overview of LHCb

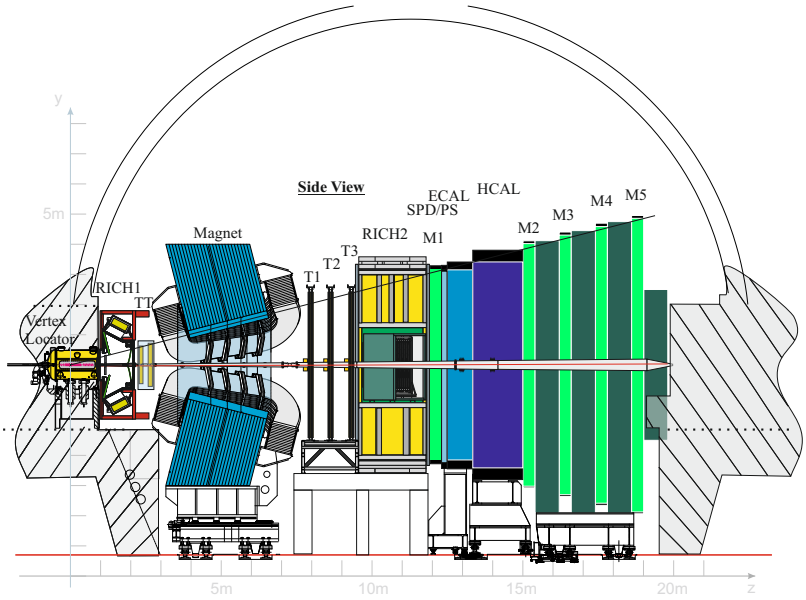


Figure 3.4: Schematic of the LHCb detector. Taken from Ref. [57].

The Large Hadron Collider Beauty (LHCb) detector [4, 58] is a single-arm forward spectrometer covering angles from approximately 15 mrad to 300 (250) mrad in the bending (non-bending) plane (approximately $2 < \eta < 5$). It was designed as a dedicated experiment for the study of particles containing b or c quarks. The layout of the detector is shown

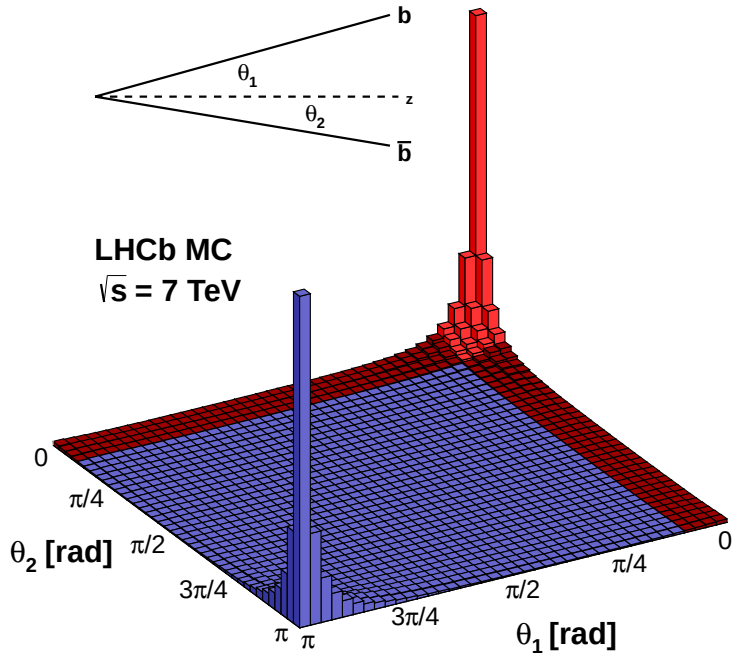


Figure 3.5: Production angles of $b\bar{b}$ pairs at the LHC during Run 1 conditions. The red region shows the acceptance of the LHCb detector, with the light red region highlighting the events where both b-hadrons are produced in the LHCb acceptance. The change to $\sqrt{s} = 13\,\text{TeV}$ for Run 2 makes no qualitative difference to the figure. Taken from Ref. [61].

in Figure 3.4. It consists of many detector subsystems, or *sub-detectors*, most of which are assembled in two halves, which can be moved out horizontally for assembly and maintenance purposes, as well as to provide access to the beam-pipe. They are referred to as the detector A- and C-sides (left and right respectively looking into the detector from the interaction point). The detector geometry is defined with reference to a right-handed coordinate system, with z along the beam axis into the detector (from left to right in the figure), y vertical and x horizontal. Cylindrical polar coordinates (r, ϕ, z) are also used. The forward acceptance exploits that, at high-energy pp collisions, both the b and $\bar{b}$-hadrons are predominantly produced in the same forward or backward cone, as shown in Figure 3.5. The size of the cavern that LHCb was built in - excavated for the DELPHI experiment [59] on LEP - prevents LHCb being instrumented in the backward region as well. Despite covering only 1.8% of the solid angle, approximately 27% of $b\bar{b}$ production falls in LHCb [60].

Broadly, LHCb consists of a high-precision tracking system and a series of sub-detectors primarily responsible for identification of particle species. The tracking system features - in the order that a particle produced in the collision would encounter them

- a silicon-strip vertex detector surrounding the pp interaction region known as the Vertex Locator (VELO) [60], a large-area silicon-strip detector (the Tracker Turicensis, or TT), a dipole magnet [62] and finally three stations (T1-3) of silicon-strip detectors and straw drift tubes [63]. The identification system consists of two Ring-Imaging Cherenkov (RICH) detectors, RICH1 and RICH2 [64], which lie before and after the magnet respectively. These are responsible for differentiating between hadrons of different species. Power to discriminate between photons, electrons and hadrons is given by a four-component calorimeter system: the Scintillating Pad Detector (SPD), PreShower (PS), Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). Finally, there is a muon identification system with alternating layers of iron absorber and multiwire proportional chambers [65].

Interesting collision events are sifted out by the means of a trigger [66], which in Run 2 consisted of a hardware stage and a two-level software stage. The former used only information from the calorimeter and muon systems, whereas the latter performs a full event reconstruction to enable complex selections before saving to permanent storage. Together these perform a data reduction of 99.7% [51], with the remaining data saved for use in physics analysis and detector calibration.

LHCb's original physics goals were to precisely study charge-parity (CP) violation in the "heavy flavour" (b and c quarks) sector, and search for rare or even forbidden heavy-flavour decays. Both are aimed at finding beyond-the-Standard-Model (BSM) physics: the former for missing sources of CP violation that would help to explain the origin of the matter-antimatter asymmetry in the universe, and the latter as enhancement/presence of these rare/forbidden decays would necessitate new methods of production. Excellent tracking and particle identification performance is necessitated by both of these goals, and how they influenced the design will be the topic of the next two sections. Although not on the original menu for LHCb, its design is also fortuitous for precision EW physics too, as will also be highlighted where appropriate.

3.3 The LHCb tracking system

"Tracking" is the reconstruction of the trajectories of charged particles in a detector. The passage of an energetic charged particle through a sensor material (often a semiconductor or a gas) causes the creation of electron-hole pairs or ionization of gas molecules. An electric field is applied across the sensor, causing the resulting electrons to drift to an anode and a current is measured. The exact intersection point in one sensor is typically difficult to ascertain, so detectors will be segmented with many such sensors, and the precision at which the hit position is known is set by how finely the detector is segmented. Reconstruction software then "join-the-dots" to work out the trajectory through adjacent

sensor layers. LHCb's tracking system, like any other, works on these principles.

LHCb's tracking system was designed with excellent vertex and momentum resolution in mind, as well as high tracking efficiency. Precise measurements of heavy-flavour decays require precise reconstruction of the production ("primary") and decay ("secondary") vertices (PV and SV, respectively) of the heavy-flavour hadron. The length between these two positions is a proxy for the decay lifetime, which enables efficient triggering on heavy-flavour decays and removal of backgrounds, and allows measurements such as the benchmark process of B_s^0 meson oscillations (see e.g. Ref. [67]). The VELO, surrounding the interaction point, is finely segmented to achieve this. LHCb measures particle momentum with tracking stations before and after the magnet; the track's curvature in response to the magnetic field is inversely proportional to the track's momentum. The magnet also identifies the sign of the charge; oppositely-charged particles curve in opposite directions under the same B field. Excellent momentum resolution naturally gives excellent invariant mass resolution, which is key in discriminating between different heavy-flavour decays and their various backgrounds, which all have differing mass spectra. EW physics also benefits from these characteristics: the excellent vertex resolution helps to distinguish between muons from $W \to \mu\nu$ or $W \to (\tau \to \mu\nu\nu)\,\nu$, where muons from the latter are produced when the tau lepton decays, away from the PV; and since LHCb has been designed to pick out heavy-flavour decays, the same characteristics can be used to suppress them in an EW analysis. The excellent momentum resolution is crucial for precise measurement of the p_{T}^{μ} distribution, which is the foundation of the measurements in Chapters 8 and 9. The presence of many tracking stations ensures high tracking efficiency, which is necessary when you need to reconstruct every piece of a multi-body heavy-flavour decay. A final common point is that the stations are typically segmented more finely closer to the beam-pipe, as this is where the flux of particles is larger. Greater segmentation keeps the *occupancy* - a measure of how frequently each segment will be hit - low enough to control radiation dose (and damage), and to keep track reconstruction efficiency high while keeping the rate of fake tracks small.

3.3.1 Dipole magnet

LHCb's dipole magnet provided the bending force necessary for charged-particle momentum measurement and charge assignment. Magnets with greater field strength deflect particles of a given momentum more strongly, and more deflection gives better momentum resolution. However, the nearby VELO and the RICH detectors required as low a field as possible. With these constraints, plus cost and the size of the LHCb cavern, a "warm" (rather than superconducting) dipole magnet was chosen with a peak field of 1.1 T - enough for a momentum resolution $\Delta p/p$ of 0.3% for momenta between 5-200 GeV [62]. A sketch

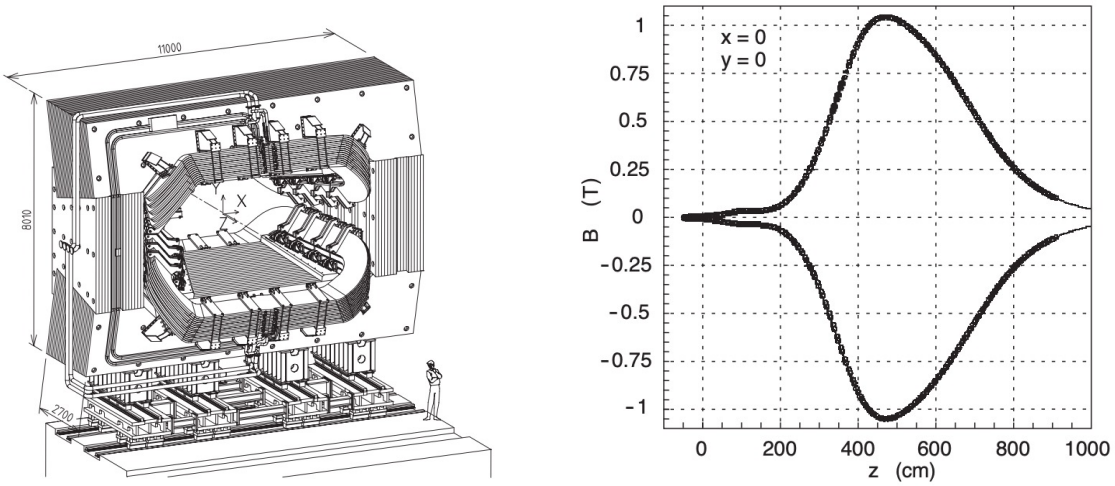


Figure 3.6: Technical sketch of the LHCb dipole magnet (left) and measurement of the vertical component of its field as a function of z (right). The latter is overlaid with a model calculation, which has excellent agreement with the true field strength. Taken from Ref. [4].

of the magnet's geometry and its magnetic field as a function of z are shown in Figure 3.6. The magnetic field lines are aligned vertically, deflecting charged particles in the $x - z$ plane. As it is not a superconducting magnet, it was comparatively straightforward to reverse the field polarity at regular intervals. When measuring *CP* asymmetries, this is extremely useful since many potential sources of systematic error due to detection asymmetries can be removed by averaging over both field polarities.

3.3.2 Vertex Locator

The layout of the VELO's sensors is shown in Figure 3.7. Excellent vertex resolution is achieved by putting the sensors as close to the beam as possible (less extrapolation required), segmenting them very finely and minimizing the amount of material (also known as the *material budget*) between the VELO sensors and the *pp* interaction - more material means more scattering off the material, and therefore more uncertainty on its trajectory. All tracks produced in $1.6 < \eta < 4.9$ will pass within at least 3 sets of sensors, and should therefore provide at least 3 distinct clusters of hits to reconstruct. As can be seen from the figure, the VELO has two halves (an A- and C-side). When the VELO is "closed" for data-taking, the two halves overlap slightly to provide full angular coverage and to help with alignment. At this point, the VELO's sensors lay only 7 mm from the beams. Each VELO half contains 21 "modules" spaced in z, as well as a further two modules forming the *pile-up veto system*, which in principle allows for removing hard-to-reconstruct events with multiple *pp* interactions, although this system was unused in Run 2. The modules carry the sensors and their support structure. The active area, shown in Figure 3.8 of

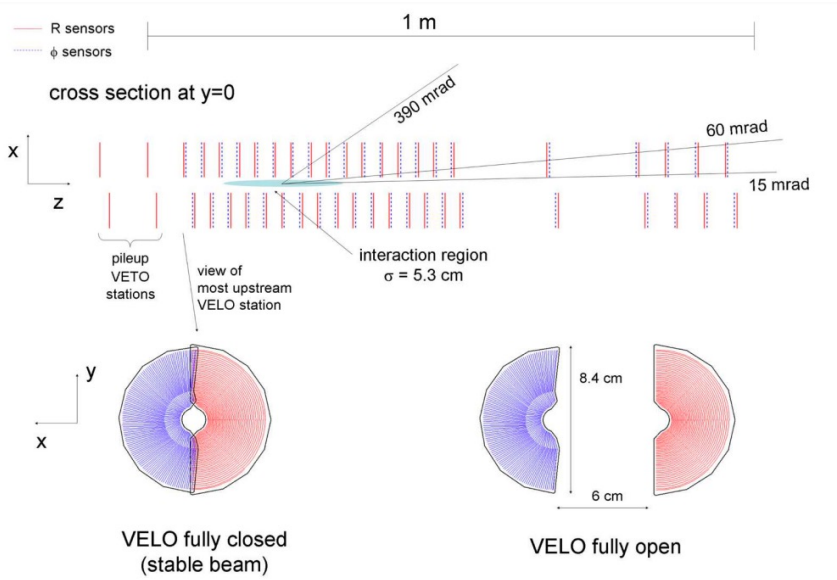


Figure 3.7: Schematic layout of the VELO modules with respect to the interaction point (viewed from above) and of the two modules (one from each half) when in their open and closed positions. Taken from Ref. [4].

each is approximately semicircular and double-sided: one side measures the particle's radial coordinate, and the other its azimuthal, or ϕ coordinate. They are therefore called R and ϕ sensors respectively. The R sensor has its silicon semiconductor strips arranged in concentric circles centred on the nominal beam axis, with a strip pitch[3] of 38 µm at the edge closest to the beam, and increasing linearly with radius. The ϕ sensors have strips that point approximately radially outwards from the beam axis, although they have a kinked, "dog-leg" shape to aid the pattern recognition and keep strip pitch low at the farthest point from the beam. Separation into two regions (which are individually read-out) helps to minimize the occupancy. Low occupancy is achieved in the R sensors by dividing the semicircular strips into four separately read-out 45° regions. The readout electronics are placed at the outer circumference of the modules. This arrangement of strips, with a pitch varying from 40-100 µm, gave a single-hit position resolution of around 5-25 µm, as measured from data taken in 2010 [58].

To limit the effects of radiation damage, the sensors were cooled to below -5°C by pumping two-phase CO_2 through the modules. The VELO operates in vacuum, but a less stringent vacuum than the LHC beam vacuum, so the two are isolated by an aluminium vacuum vessel known as the "RF-box", which also shields the sensors from radio-frequency interference from the beam. The side of the box facing the beam is corrugated to allow the VELO halves to overlap, and is just 0.5 mm thick to keep the material budget low. It

[3]Distance between the centre of adjacent active areas. The position resolution of the tracker is dependent on the pitch and the angle of incidence on the sensor.

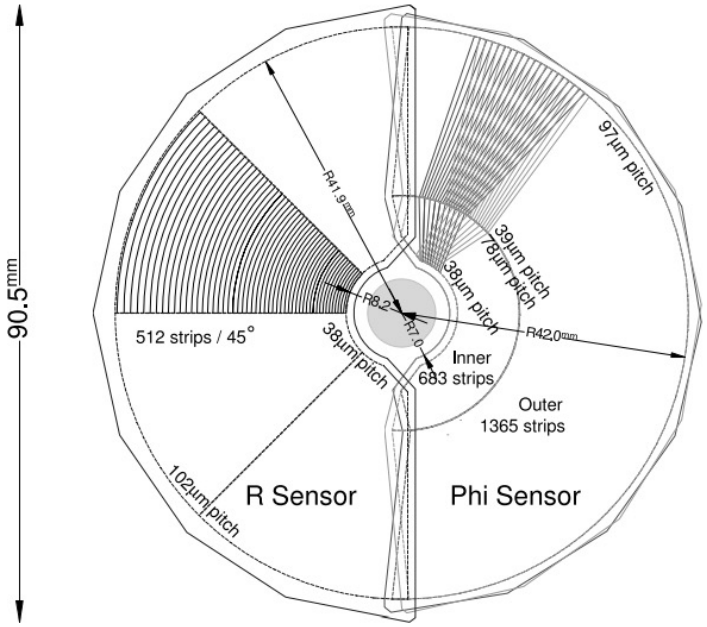


Figure 3.8: Technical drawing of the VELO's R and ϕ sensors. Taken from Ref. [4].

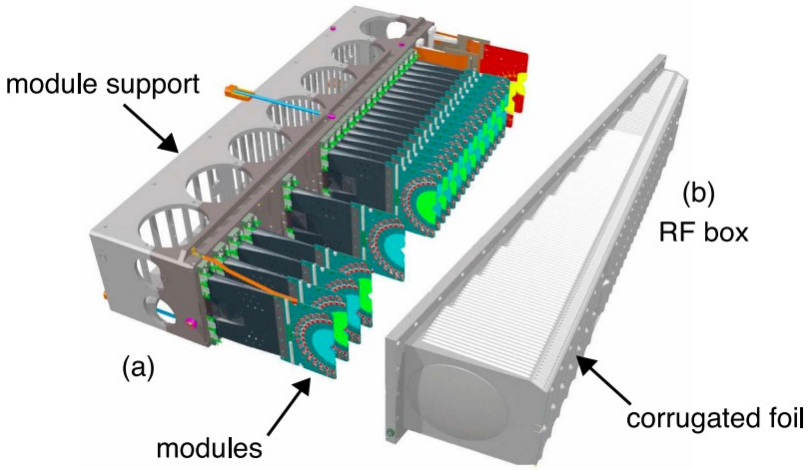


Figure 3.9: Exploded view of the VELO modules in their support structure and the RF box. Taken from Ref. [4].

is therefore known as the "RF-foil". This arrangement is shown in Figure 3.9. Altogether, the VELO's material comes to only 17.5% of a radiation length, which contributes to the excellent momentum resolution.

3.3.3 Tracker Turicensis

The Tracker Turicensis (TT) was a planar silicon-strip detector, formerly known as the "Trigger Tracker" due to the role it played in LHCb's track reconstruction, where it provided coarse information on the track momentum due to being in the fringe field of the dipole magnet. It also provided redundancy in the tracking system, which could be exploited in determining track reconstruction efficiencies, as will be seen in Chapter 7.

The layout of the TT sub-detector is shown in Figure 3.10. It was split into four layers spaced in z, each with approximately the same geometry, but arranged in a so-called (x, u, v, x) geometry, with the u (v) layers rotated by +(-) 5° with respect to the x layers. Each TT layer was split into 15 (first two layers) or 17 (last two layers) columnar modules, with the readout electronics at the top and bottom (outside the LHCb acceptance), and 14 sensors per module. Each sensor had silicon strips aligned vertically with a strip pitch of 183 µm. Vertical strip alignment was chosen because the magnet bends tracks in $x - z$ plane, so much finer granularity was needed in the x direction compared to y for a momentum measurement, particularly as the track direction before the magnetic field was well-constrained by the VELO. The (x, u, v, x) layer arrangement improved the y coordinate resolution. The sensors were read-out in sectors, with the sensors in each sector bonded together vertically. Since the occupancy is higher closer to the beam-pipe, more sensors were bonded together into a single sector as the distance from the beam-pipe increased. The entire TT sub-detector was encased by light-tight,

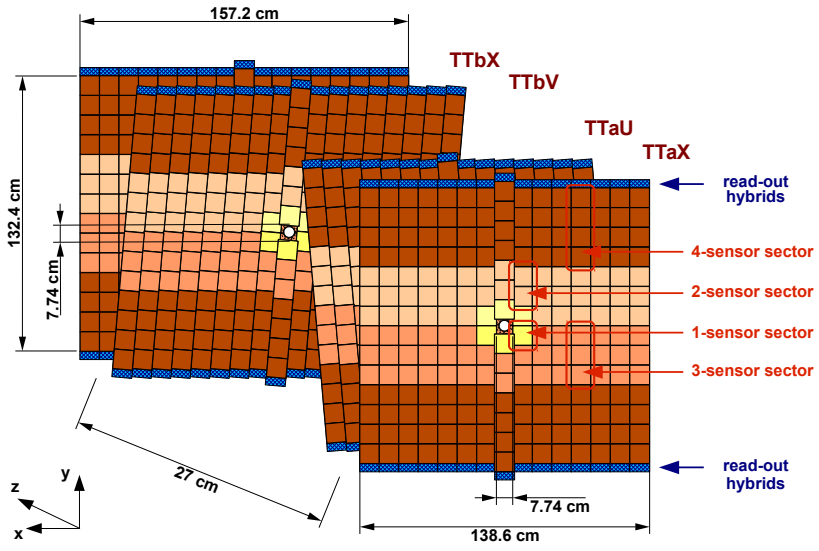


Figure 3.10: Layout (not to scale) of the Tracker Turicensis. The beam-pipe passes through the circular hole at the centre of each layer. Each approximately square box is a sensor, and adjacent sensors of the same colour in the same module (column) were bonded and read-out together. Taken from Ref. [68].

thermally and electrically-insulated box, within which C_6F_{14} gas was circulated to keep the detector below 5°C to reduce the build-up of radiation-induced leakage current. The hit resolution of the TT was measured to be 61 µm in 2012 [69].

3.3.4 Downstream tracking stations

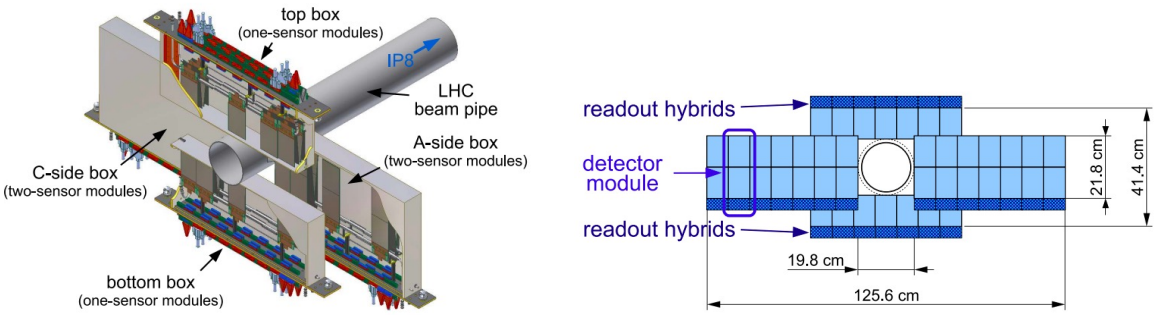


Figure 3.11: Layout schematic of the four detector boxes around the beam-pipe in one IT station (left), and the layout of the first layer of silicon sensors (right). In the latter, the sensors from each of the four boxes are shown without the ancillary structure of the boxes, thus just showing the active area of the detector layer. Taken from Ref. [4].

T1, T2 and T3 measured the track trajectory after the dipole magnet; three stations gives three independent measurements, which is crucial to achieve high momentum resolution and high track reconstruction efficiency. They also provided the possibility to reconstruct long-lived particles that have decayed after the VELO. Since the flux of particles is larger closer to the beam-pipe, the T stations were separated into the "Inner" and "Outer" trackers, or IT and OT. The IT, like the TT, was implemented in high-granularity silicon strip sensors, while the OT used lower-granularity drift tube technology.

Each of the three IT stations (one per T station) consisted of four independent detector boxes, arranged to the left, right, above and below the beam-pipe. Within each box were four layers of silicon strip sensors spaced in z and arranged in the same (x, u, v, x) geometry as the TT stations. The layout of the boxes, and the silicon sensors in the first layer of all four boxes, is shown in Figure 3.11. Each module consisted of one (in the top and bottom modules) or two (A- and C-side) silicon sensors (plus their readout electronics), and each sensor featured strips with a pitch of 198 µm. Within each detector box, C_6F_{14} was circulated to cool the sensors below 5°C. The hit resolution of the IT was measured to be 54 µm in 2012 [69].

Each of the three OT stations consisted of four layers - again in a (x, u, v, x) geometry - as is shown in Figure 3.12. Each layer here contained two staggered layers of vertically-aligned ($\pm 5°$ if in a u or v layer) drift tubes, separated into 14 long (~ 5

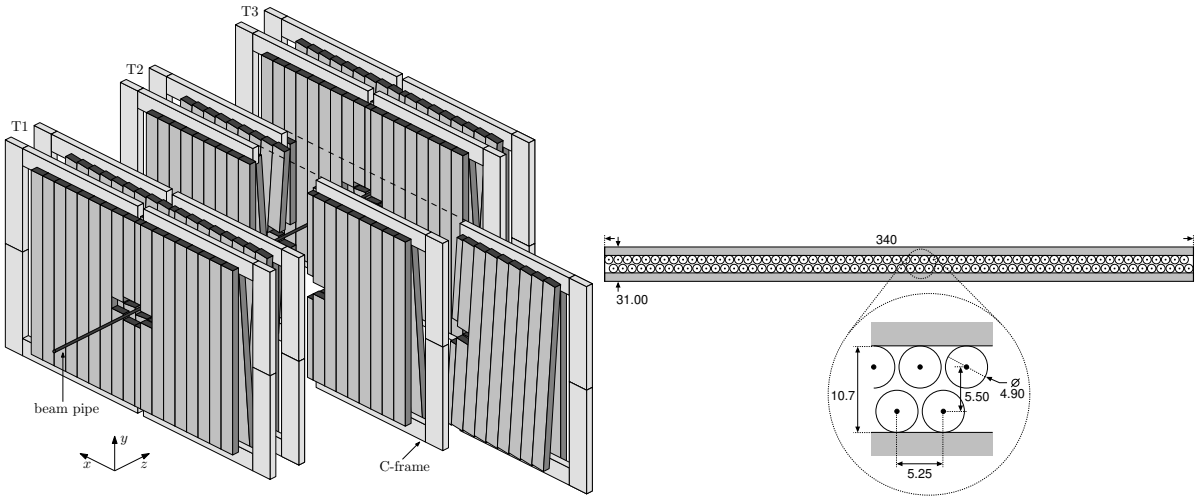


Figure 3.12: Layout of the modules in the three OT stations relative to the beam-pipe (left), and a view from above of the drift tubes in a single OT module (right). In the former the supporting C-frames are shown, and T2 is shown open (as would be for maintenance), while T1 and T3 are closed. A space around the beam-pipe can be seen, which would accommodate the IT stations. Distances in the right plot are in mm. Taken from Ref. [70].

metres) and 8 short ($\sim$ 2.5 metres) modules. In each module, the arrangement of the two staggered "monolayers" of the drift tubes is also shown in Figure 3.12. The drift gas (70% Argon and 30% CO_2) was ionized by the passage of high-energy charged particles, and the front-end electronics (for each tube) measures the time taken by ionization electrons to be collected at the central anode wire, giving a 200 µm hit-position resolution in x [70], which had been improved to 170 µm by Run 2, mainly due to introduction of the *real-time* alignment procedure discussed in Section 3.5 [63].

3.4 Particle identification in LHCb

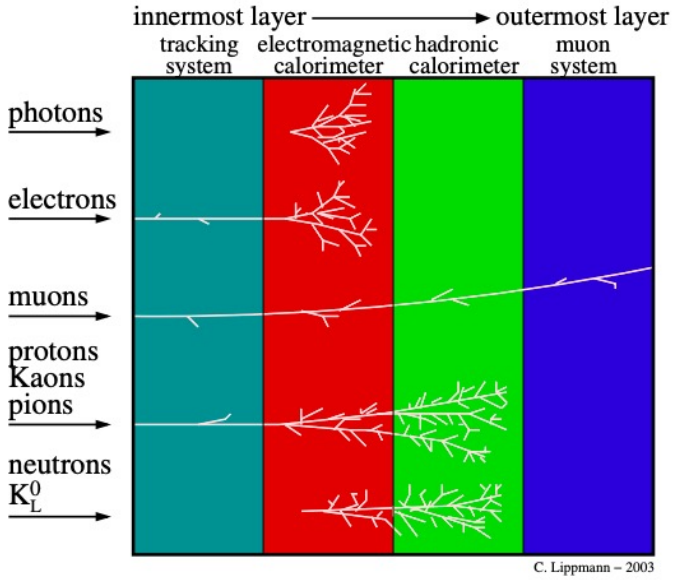


Figure 3.13: Schematic showing the interaction of different particle types with the segments of a "traditional" particle detector such as LHCb. Note that this figure assumes a magnetic field applied into/out of the page throughout the whole detector, whereas in LHCb all the corresponding sub-detectors except the tracking system operate in regions of negligible magnetic field. Taken from Ref. [71].

The need to differentiate between decay modes necessitates a system of sub-detectors to identify the species of each final-state decay product; to give particle identification (PID) information. Fortunately, different particle such as muons, electrons, and charged and neutral hadrons have distinct interactions with various sub-detector technologies, as shown schematically for a typical particle detector in Figure 3.13. Firstly, charged particles leave hits in tracking stations, whereas neutral particles do not. Muons are highly penetrating, and will typically traverse the entire detector, so typically calorimeters - which will stop everything except muons - will be placed before the muon system. In principle, signals in the muon system can therefore only signify muons. Particles stopped by the calorimeters deposit almost of their energy there, so calorimeters - as their name suggests - can also measure particle energy. Hadrons typically penetrate further into a calorimeter than electrons and photons, which can be exploited to differentiate them by segmenting into an electromagnetic and hadronic calorimeter, such that the energy from electrons and photons will be contained solely in the electromagnetic calorimeter.

LHCb contains sub-detectors fitting all of these functions: the tracking system has already been discussed; it has a four-component calorimeter system for differentiating between hadrons, photons and electrons and measuring their energy; and a muon system at

the far end of the detector. These will be described in the following subsections. However, differentiating between hadrons of different species (pions, kaons, protons etc.) is of particular importance to LHCb - without this capability different heavy-flavour decays could not be distinguished. Different hadronic species are hard to distinguish by their calorimeter signals, so in LHCb, hadron PID is achieved with the RICH detectors.

3.4.1 The RICH detectors

In principle, ring-imaging Cherenkov (RICH) detectors measure a particle's speed. Via the relativistic invariant equation $m^2c^4 = E^2 - |\vec{p}|^2c^2$[4], the speed can be combined with the momentum vector $\vec{p}$ (measured with a tracking system) to yield the particle's mass. This, along with the charge (known from the direction of track deflection in a magnetic field), reveals the unambiguous identity of the particle.

The method of determining the speed is as follows. Charged particles travelling through a refractive medium (with refractive index $n > 1$, also known as a "radiator") at a speed greater than the speed of light in that medium radiate so-called Cherenkov photons. In the lab frame, this light is emitted preferentially in the direction of motion of the particle, such that the trajectories of the photons emitted from a single point define a cone, with axis aligned with the particle's velocity vector $\vec{v}$ and opening angle θ defined by

$$\cos\theta = \frac{c}{|\vec{v}|n}\,, \tag{3.1}$$

where c is the speed of light in a vacuum. The cone of photons will form a ring of hits on a detector plane that is perpendicular to the direction of travel (hence ring-imaging), and from the ring radius and the known point of emission, θ can be determined [72]. With knowledge of θ and the radiator's refractive index, the speed can be derived via Eqn. 3.1. If the radiator is not a plane, but rather a 3D volume, then the emission will occur uniformly as the particle passes through the radiator, and the conical shell of Cherenkov photons will become a filled cone. The projection onto the detector plane therefore becomes (problematically) a filled circle. In this case, concave spherical mirrors can be used to focus the light back into a ring, by using their property that incident parallel rays in 2D - from any angle of incidence - will be focused to a single point on the focal plane of the mirror [73]. Generalizing to 3D and a filled circle will be focused to a ring, thus recovering the ability to measure the opening angle.

LHCb has two RICH detectors, shown schematically in Figure 3.14 in their Run 2 configuration. RICH1 was specialized for lower-momentum particles ($\sim$1-60 GeV) which could be bent out of LHCb by the magnetic field, so was placed upstream of the dipole

[4]In this subsection it is helpful to keep the speed of light in a vacuum (c) in the equations, so SI units are temporarily used.

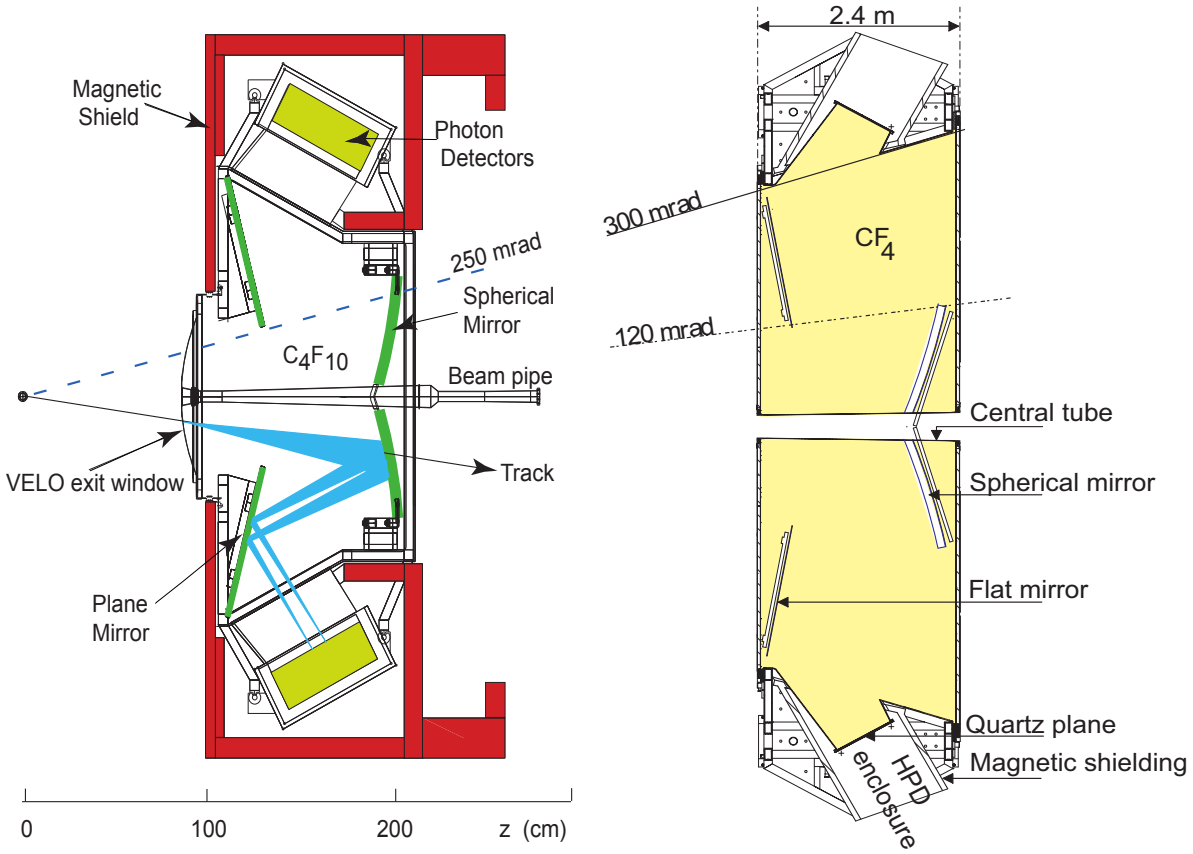


Figure 3.14: The optical layout of RICH1 (left) and RICH2 (right). Taken from Ref. [74].

magnet. RICH2 was downstream and effective from $p \gtrsim 15\,\text{GeV}$. The momentum spectra of particles produced in collisions is *softer* (peaked towards lower momentum) at larger polar angles, so RICH1 covered the whole LHCb acceptance, while RICH2 covered only $3 < \eta < 5$. RICH1 also featured a radiator of higher refractive index than RICH2 to give larger cone opening angles and hence more sensitivity to lower-momentum particles. Besides these differences, the operation and design of RICH1 and RICH2 is essentially the same. To fit the momentum spectra of LHCb, the RICH1 (2) radiators were volumes of C_4F_{10} (CF_4) gas in Run 2 [75], thus spherical mirrors were used to focus the Cherenkov photons from one particle into a ring, which were then further reflected by flat mirrors out of the LHCb acceptance onto hybrid photon detectors (HPDs). Having only the radiators and the spherical mirrors within the LHCb acceptance helped to keep the material budget down, and the flat mirrors ensured that the light was approximately normally incident on the HPD planes, which is advantageous for ray-tracing and the efficiency of photon detection.

The HPDs [76] featured a photocathode deposited onto the entry window of a vacuum tube, which produced photoelectrons (via the photoelectric effect) when hit by a Cherenkov photon of sufficient energy. The photoelectron was then guided by acceleration

optics to a silicon pixel sensor to give sensitivity to the incident ray's direction. The HPDs lose efficiency if operated in magnetic field above 3 mT, so were heavily shielded from the LHCb dipole magnet's field. The radiator gas was separated from the HPDs by a transparent quartz window.

3.4.2 Calorimeter system

Unlike tracking stations, which try to disturb the traversing particle as little as possible, calorimeters have large material budgets. Therefore, rather than ionizing or exciting the detector material, incident high-energy hadrons, electrons and photons dump all of their energy in them in the form of a *shower* of secondary particles (such as bremsstrahlung photons and pair-produced electrons). This cascade proceeds longitudinally until the secondary showering particles no longer have sufficient energy to propagate the shower further. Electromagnetic calorimeters are designed to contain the full shower of electrons and photons, while a hadronic shower will penetrate into the hadronic calorimeter as well. Calorimeters then measure the energy of the primary high-energy particle by summing up the energy of all parts of the shower it initiated.

The energy of the showered particles can be measured by collecting their *scintillation light*. The passage of energetic charged particles (created in the shower) through a scintillating material causes excitation of the crystal structure, and the subsequent de-excitation causes emission of visible light [72]. The deposited kinetic energy of the charged particle is ideally proportional to the light yield, although for heavy and/or very-high-momentum (> 100 GeV) particles there is significant non-linearity [77]. The scintillating material is then chosen to be transparent to its own scintillation light, and is typically collected and converted (proportionally) to an electronic signal by e.g. a photomultiplier tube. Scintillators will be interleaved with "absorbers", the latter being planes of high-budget material to propagate the shower.

The layout of the LHCb calorimeter system in Run 2 is shown in Figure 3.15. In order of ascending distance from the nominal interaction point, it consisted of the Scintillating Pad Detector (SPD), the PreShower (PS) detector, then the Electromagnetic and Hadronic Calorimeters (ECAL and HCAL respectively). There was also a 2.5-radiation-length lead foil, known as a "converter", between the SPD and PS. This system provided the discrimination described above and measured particle energy, as well as having a prominent role in first level of the LHCb trigger. The lead converter was responsible for initiating the electromagnetic shower, therefore signals in the SPD upstream of it signified a charged particle. This helped to discriminate between electrons and photons, which both produced showers downstream after the foil. Electrons can be misidentified as charged pions, although the former deposit more of their energy "earlier" in the

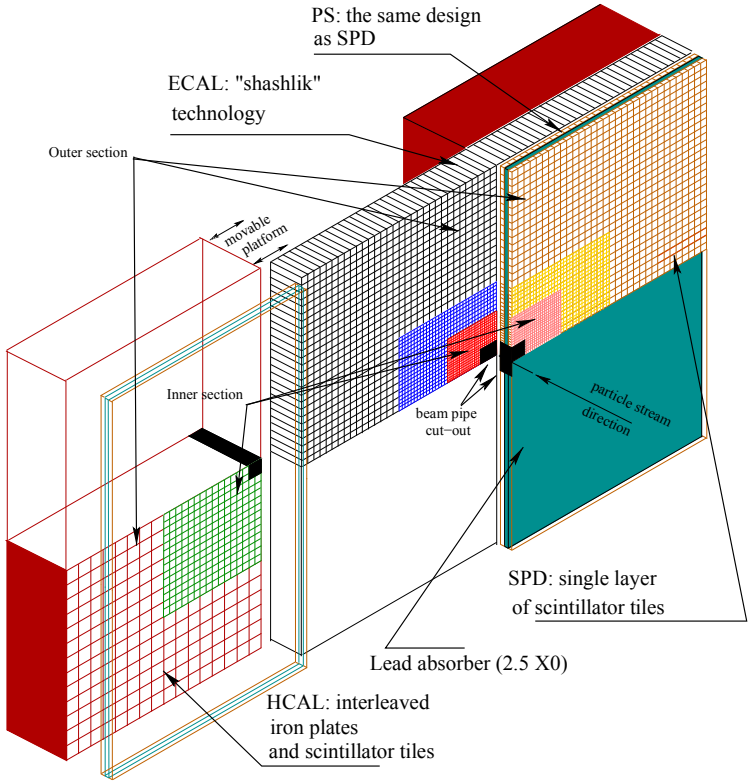


Figure 3.15: Layout of the calorimeter system. Taken from Ref. [78].

shower on average, so the presence of the PS as well as the ECAL probed the shower longitudinally. To accurately measure the energy, the PS and ECAL needed to fully contain the electromagnetic shower. Charged hadrons would have their energy (through momentum measurements by the tracking system and the RICH detectors) well-measured already, so full containment of hadronic showers was deemed less of a priority when considering cost and cavern space. The transverse energy of a hadronic candidate was reconstructed by summing the E_{T} of the HCAL cluster and the E_{T} of the ECAL cluster directly in front of it [78]. In the ECAL and HCAL, the gain of the photomultipliers scaled with the distance from the beam axis to give a constant E_{T} scale, since $E_{\mathrm{T}} = E \sin\theta$ where θ is the polar angle from the beam axis.

As with the tracking stations, the lateral segmentation was finer closer to the beam axis, as dictated by the larger particle density there. The SPD, PS and ECAL had three regions of lateral segmentation, while the HCAL had only two with larger cell sizes as hadronic showers tend to be broader in the perpendicular directions. The segments were sized such that the SPD-PS-ECAL group is projective with respect to the tracking systems. The SPD and PS were almost identical, and contained rectangular scintillating tiles made of doped polystyrene. Embedded in the tiles were optical fibres, which collected and

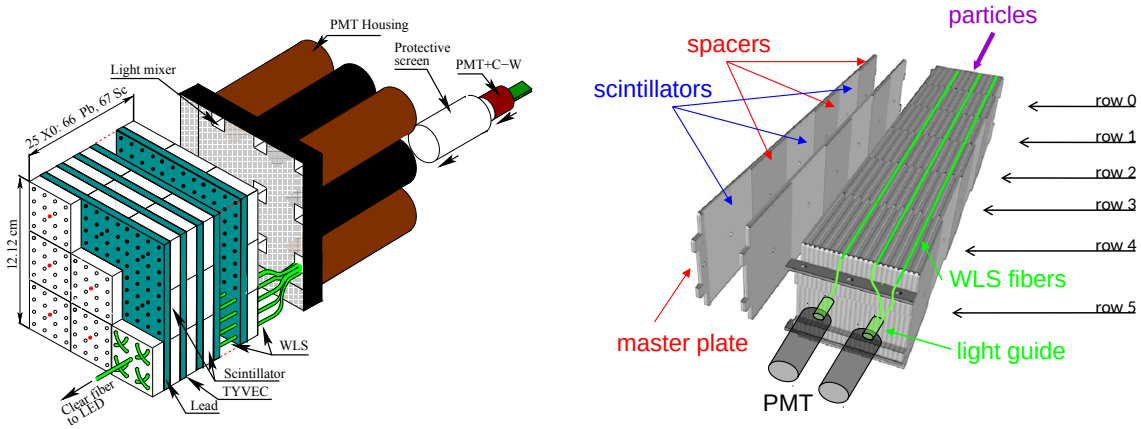


Figure 3.16: Schematics of an ECAL (left) and a HCAL cell (right). PMTs at the back of each cell were all read-out together as a single detector channel. Taken from Ref. [78].

transported the light to multi-anode photomultiplier tubes (MA-PMTs). The MA-PMTs were situated outside the detector acceptance, such that one MA-PMT channel received the light from one scintillator tile.

A schematic of an ECAL and a HCAL cell is shown in Figure 3.16. Longitudinally, these ECAL cells had a "Shashlik" structure, which consisted of alternating layers of absorbers and scintillator plates. In total, these layers added up to 25 radiation lengths - sufficient for containing an electromagnetic shower. Stacked at the end of the cell was a PMT and corresponding front-end electronics. Contrastingly, the HCAL cells had a "sampling" structure, with scintillating tiles and absorbers aligned parallel to the beam axis, and stacked laterally. Optical fibres ran along the edges of the tiles, delivering the scintillation light to PMTs stacked at the end of the module. Multiple fibres were grouped to create readout channels corresponding to different lateral dimension. The alignment of tiles parallel to the beam axis achieved better light collection that stacking of plates perpendicular to the beam axis [78].

3.4.3 Muon system

The LHCb muon stations were implemented in a combination of Multi-Wire Proportional Chambers (MWPCs) and Gas Electron Multipliers (GEMs), both of which are, in principle, proportional ionization chambers. Like the drift tubes in the OT described earlier, incident muons cause ionization of the drift gas, and the ionization electrons drift and are multiplied under applied electric field until collected at the anode. A multi-wire proportional chamber features many anode wires, giving position sensitivity, and the multiplication occurs as the electric field becomes larger closer to the wire [72]. GEMs, on the other hand, are simple conductive foils with machined holes which guide the electric field lines through them, geometrically providing the increase in electric field necessary for multiplication. Front-end

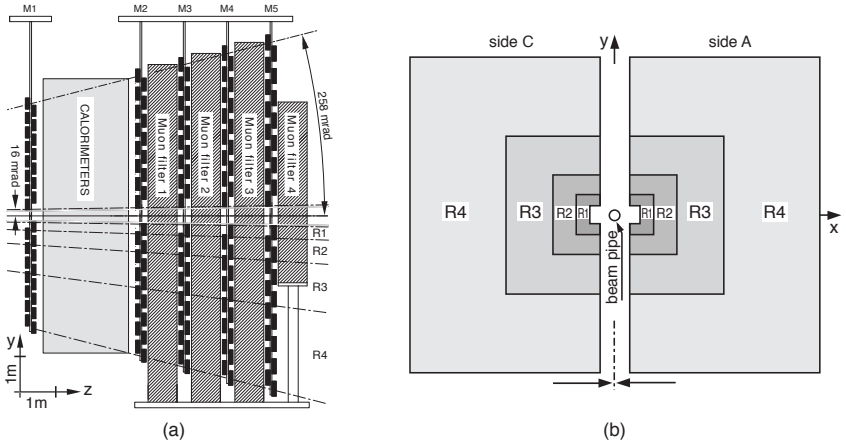


Figure 3.17: Layout of muon stations from the side (a) and looking down the beam-pipe (b). Taken from Ref. [65].

electronics read the electrical signals produced by the collected electrons, which signals the presence of a charged particle. As the muon stations are beyond the calorimeters, by design these particles should only be muons, giving muon PID information. The muon stations also had a pivotal role in the trigger system, as will be explained in Section 3.5.

The longitudinal layout of the muon stations, and their segmentation in the transverse plane, is shown in Figure 3.17. Station M1 was placed in front of the calorimeter system, with stations M2-M5 downstream. The inner region of M1 was instrumented with triple-GEMs (three layers of GEMs stacked in the incident particle direction) to handle the harsher environment closer to the beam-pipe [79], while the rest of M1-M5 were instrumented with MWPCs. M2-M5 were also interleaved with 80 cm thick iron absorbers, such that only muons with $p > 6\,\text{GeV}$ will be able to traverse the whole muon system; anything lower should be stopped. The chambers were constructed such that they form adjacent projective towers pointing towards the interaction point, as with the calorimeters. M1 was only used in the first stage of the LHCb trigger system, M2-M3 had the finest segmentation for the necessary position sensitivity, and finally the utility of M4-M5 was that only muons should get through to them. The MWPC wires were aligned in the y direction, again giving higher spatial resolution in the magnet's bending plane ($x - z$). These wires were grouped into channels that were read-out together, and further grouped into $x - y$ logical pads which returned binary signals if any one of the physical channels within returned a positive signal. The size of the logical pad therefore defined the position resolution of the muon system, which did not have to be particularly fine given that muons will have left hits in the main high-granularity tracking system upstream. The pad size increased with distance from the beam-pipe, and was grouped into regions, R1-R4, of constant pad size. M2-M3 (M4-M5) had twice (half) the number of logical pads as M1. As a consequence, the best x (y) hit resolution was 4 (10) mm for

region R1 of M1, increasing to 150 (180) mm for region R4 in M5 [65].

3.5 The LHCb trigger and data processing strategy

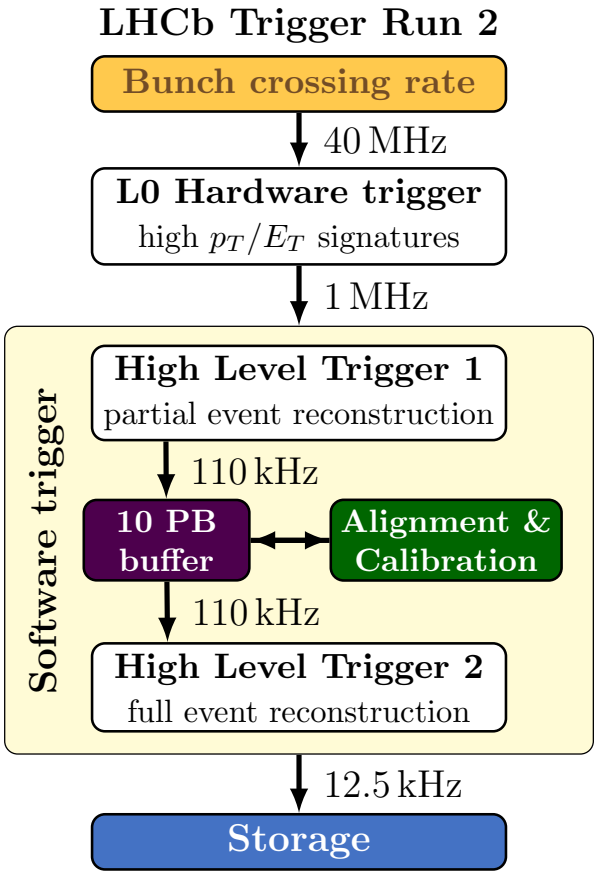


Figure 3.18: Schematic diagram of the Run 2 LHCb trigger system. Taken from Ref. [51].

The LHC minimum bunch spacing of 25 ns corresponds to a maximum collision frequency of 40 MHz. However, in Run 2 the maximum rate at which the front-end electronics of LHCb could be read-out was around 1 MHz. Data storage of the volumes in question is also prohibitively costly. It is unrealistic to record data from every pp collision, necessitating a *trigger system*, which decided which events were worth recording.

A schematic of LHCb's Run 2 trigger is shown in Figure 3.18. The trigger consisted of three stages, denoted L0 ("Level-0"), HLT1 and HLT2 ("High-Level Trigger 1, 2"). The first stage was a hardware trigger, which took information from the calorimeters and muon systems and made fast calculations of p_{T} and E_{T} of tracks on a system of field-programmable gate arrays (FPGAs). The two stages of the High-Level Trigger were software-based; a positive L0 decision meant the full detector was read-out, and the event's raw data was transferred to the Event Filter Farm (EFF) - a network of around

1700 computer nodes - which ran the HLT1 partial reconstruction and selection. Events passing HLT1 were then temporarily stored in a disk buffer with a capacity 10 PB, enough to absorb a maximum of two weeks of continuous data-taking. This buffer provided a safety mechanism against detector and trigger issues, and gave the flexibility to delay the execution of HLT2 to inter-fill periods, where more computing power was available. A subset of the events triggered by HLT1 were dedicated calibration samples, which were used to align and calibrate the detector run-by-run [80, 81]. HLT2 then performed a full event reconstruction followed by the final stage of data reduction, with the benefit of the up-to-date alignment and calibration information [82]. Events passing HLT2 were transferred to permanent storage.

Each stage of the trigger contains a set of classification algorithms: each algorithm checks the reconstructed information about the event against its requirements and decides if the event is interesting or not. This algorithm is typically called a trigger *line* or *selection*. The differing requirements reflect that an event may be interesting for different reasons to different analysts. An event needs a positive decision from at least one line to be passed on to the next trigger stage.

3.5.1 L0 hardware trigger

In Run 2, L0 had six trigger lines. The trigger line most relevant here is the so-called `L0MuonEW` line, part of a family of three L0 muon trigger lines. In every event, a simple muon-station-only reconstruction was run on FPGAs, which would first build straight-line tracks that pointed back to the interaction region from aligned hits in logical pads of all five muon stations. Once found, the p_{T} of these tracks could be estimated with around 25% uncertainty, using the track's direction, rough knowledge of the interaction point and under the assumption of a single kick at one point in the dipole magnet[5] [51]. Only stations M1 and M2 were used to determine the momentum at L0 [83] - this being M1's only use. The `L0MuonEW` trigger then fires if there is a muon in the event with $p_{\mathrm{T}} > 6\,\mathrm{GeV}$.

The remaining two muon lines - `L0Muon` and `L0DiMuon` - required a single muon or muon pair with lower p_{T} thresholds, facilitated by an additional "Global Event Cut" (GEC). The GEC was a requirement on the maximum number of hits in the SPD, which is a proxy for overall event complexity, and so the GEC thereby selects faster-to-reconstruct events. For most LHCb analyses the GEC makes little difference, but it is not tolerable for EW analyses (particularly cross section measurements), as it has a substantial and complicated effect on the trigger efficiencies of the particularly-high-p_{T} EW events. Finally, the `L0Hadron`, `L0Photon` and `L0Electron` triggers all used calorimeter information to measure E_{T} of clusters of energy and categorize them as from hadrons, electrons or

[5]This precision is important to remember when the L0Muon trigger efficiency is considered in Chapter 7.

photons as described in Section 3.4.2. At least one cluster had to have transverse energy above a certain threshold (typically around 3 GeV). These thresholds were again allowed to be low by the GEC.

3.5.2 High-level software trigger

HLT1 then performed a partial event reconstruction: charged tracks with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ that traversed the whole tracking system (known as "long tracks") were reconstructed, and positions of primary vertices (PVs) were fitted using VELO tracks [84]. Muon PID was carried out at HLT1 (possible because of its simplicity) by extrapolating the fitted long tracks to the muon stations and searching for the expected hits. Tracks with $p_{\mathrm{T}} < 3\,\mathrm{GeV}$ would be bent out of LHCb before reaching the muon stations, and as stated earlier, a minimum momentum of around 6 GeV is needed to penetrate through all the muon stations. The muon PID algorithm in HLT1 therefore requires hits in M2 and M3 for tracks with $p_{\mathrm{T}} < 6\,\mathrm{GeV}$, at least one additional hit in M4 or M5 for $6 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$, and hits in M2-M5 for $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. M1 was not used.

HLT1 then had around 20 trigger lines in Run 2. These lines mainly triggered on *inclusive* signals, i.e. a signature that doesn't correspond to any particular particle decay, such as a single track above a p_{T} threshold that is displaced from the PV, or a displaced two-track vertex. The most relevant to this thesis was the `Hlt1SingleMuonHighPt` line, which selected events with an identified muon that had $p_{\mathrm{T}} > 6\,\mathrm{GeV}$ and $p > 6\,\mathrm{GeV}$.

With more computer power available, the HLT2 could also perform a more complex calorimeter reconstruction and the RICH reconstruction, as well as repeating the HLT1 charged-particle tracking, but this time without a p_{T} requirement on the tracks. In Run 2, this reconstruction was "offline-quality", meaning that all the complexities of reconstruction that had to be performed *offline* (i.e. after the data had been saved to permanent storage) were now possible *online* (i.e. during the trigger reconstruction; before permanent storage). Muon identification in HLT2 was the same as HLT1 for the purposes of EW physics.

Around 500 HLT2 lines were present in Run 2, this time with both inclusive and many exclusive selections. The most salient one for this thesis was `Hlt2SingleMuonHighPt`, which selected events that had passed HLT1 and had identified muons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$.

3.5.3 Persistency and offline data processing

At the point at which an event passes HLT2, LHCb's knowledge of the event consisted of the raw readout of each sub-detector, the HLT2 reconstruction of the event, and a record of how the event fired the triggers. Motivated by the aim of data reduction, which parts of this information that were saved depended upon the persistency *stream*. Each HLT2 line

was associated to one of these streams, and if an event fired multiple HLT2 lines, it could be persisted through multiple streams. There were three such streams in Run 2 [85]:

- **Full stream**: raw detector readout and trigger summary reports are saved; trigger reconstruction is discarded,
- **Turbo stream**: the trigger summary reports and the reconstructed candidates for each trigger line are saved; the raw detector readout is discarded,
- **TurCAL stream**: both the above are saved.

The Full stream is the typical method of persistency for a HEP experiment, where it is not possible to fully and precisely reconstruct the event online, so the trigger's online reconstruction is superseded by a superior offline reconstruction. This was the case for LHCb in Run 1, but by Run 2 the offline and HLT2 reconstructions were of equal quality, facilitating the Turbo persistency model. Around a factor ten in data reduction can be achieved persisting to the Turbo stream rather than the Full stream. The TurCAL stream was used for deriving efficiencies and corrections for both the online and offline reconstructions [82].

The difference between the Turbo and Full persistency streams is of great importance to the design of trigger lines. In the case of Turbo (the default in Run 3), if the signal candidate is somehow misreconstructed or does not actually belong to a true signal decay, there is no opportunity later to rectify this. The likelihood of these false positives is key information, and informed the development of the trigger optimization tools detailed in Chapter 4. For the physics measurements in this thesis, all the data analysed came through the Full stream. This was necessary as there are quantities of interest for EW analyses (such as hadronic recoil and isolation) which rely on information from the full event, not just the signal candidate. Having the raw detector readout available means that tracks can also be re-fitted after applying more complex alignment and calibration procedures than what is run in real-time, as was necessary for the analyses here.

After persisting through the Full stream and an offline reconstruction, thousands of further selection lines were then run, each known as a *stripping line* [86]. These lines typically made exclusive selections, and the most important for this thesis were the `WMu` and `Z02MuMu` stripping lines: the former selecting a single muon with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$; and the latter selecting an oppositely-charged dimuon pair with invariant mass $m_{\mu\mu} > 40\,\mathrm{GeV}$ and each muon having $p_{\mathrm{T}} > 15\,\mathrm{GeV}$. Other stripping lines were used for control samples such as $\Upsilon(1S) \to \mu\mu$.

3.6 Reconstruction performance of LHCb for electroweak analyses

The performance of the Run 2 trigger is detailed fully in Ref. [51]. Since the HLT2 reconstruction in Run 2 was of identical quality to the offline reconstruction, the reconstruction performance given there is representative of LHCb's overall performance. In this subsection, the most important resolutions for EW studies are reproduced; their importance will become clear in later chapters. The trigger efficiencies on high-p_{T} muons are studied in Chapter 7.

As was stated at the beginning of Section 3.3, the muon's momentum in $W \to \mu\nu$ or $Z \to \mu\mu$ events needs to be precisely measured. The momentum resolution of LHCb's tracking system was measured in Run 1, and is shown in Figure 3.19. At higher momenta, the tracks have less curvature in the magnetic field (the high-p_{T} muons analysed in this thesis produce approximately straight tracks), and therefore the momentum resolution increases, and is around 1% at $p = 200\,\mathrm{GeV}$. At low momentum, LHCb has the best momentum resolution of the four main LHC experiments [7]. This filters through to an excellent invariant mass resolution, measured for the dimuon resonances and shown for Run 1 in Figure 3.20. The Z mass resolution was around 1.7 GeV, which was improved by around 35% when a Z-based detector alignment was applied, which was added to the real-time alignment procedure in 2018, but is applied offline as part of the m_W and $R_{\tau\mu}$ analyses [87].

The importance of reconstructing primary and secondary vertices, and the distance

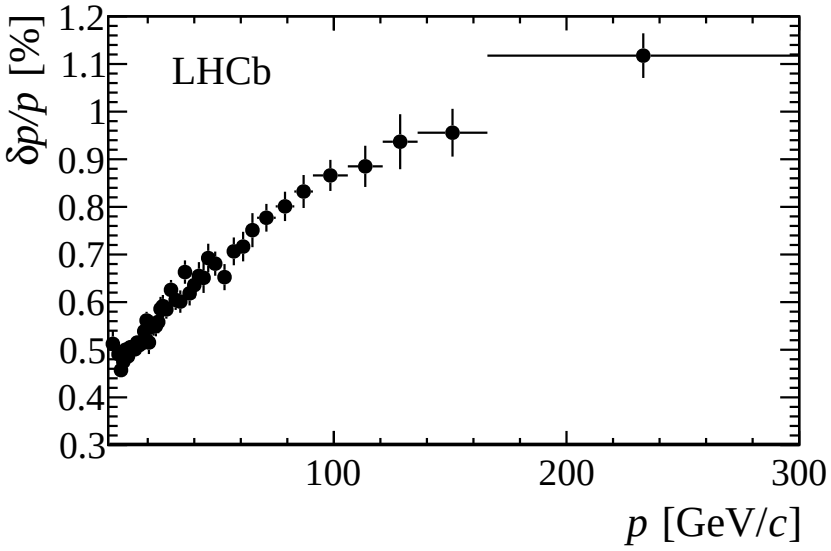


Figure 3.19: Momentum resolution of the LHCb tracking system in Run 1. Taken from Ref. [58].

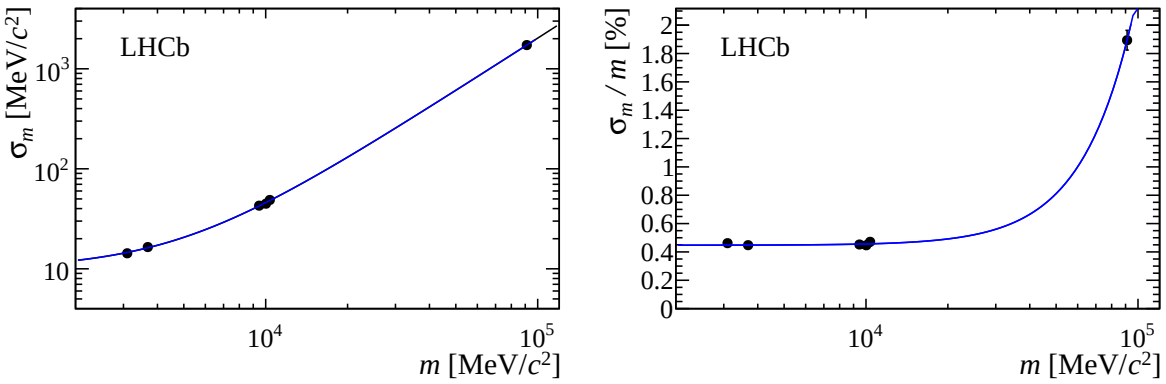


Figure 3.20: Mass resolution (σ_m) (left) and relative mass resolution (σ_m/m) (right) as a function of the invariant mass for dimuon resonances as measured during Run 1. The superimposed curve is from an empirical power law fit. Taken from Ref. [58].

between them, was discussed in Section 3.3. The performance of this reconstruction is quantified in terms of the impact parameter (IP[6]). The resolution of the x and y components of the IP are shown in Figure 3.21. The LHCb geometry means that the z component of the IP is negligible, and the IP is therefore the quadrature sum of x and y transverse components. Multiple scattering of the track with the detector material gives a $1/p$-dependent resolution on the track direction [88], which when combined with the VELO geometry and the fact that tracks in LHCb are produced at small polar angles, leads to an IP resolution with a linear dependence on $1/p_{\mathrm{T}}$. The best possible resolution is controlled by the hit position resolution in the VELO and the distance of extrapolation between the first hit and the interaction point, while the slope of the IP resolution is given by the amount of material between the measurement points and the track's origin [58]. Note here that the entire high-p_{T} region that is studied in this thesis exists solely in the first bin of these plots, with an IP component resolution of around 13 µm [51, 58].

The impact parameter resolution is naturally dependent on the resolution of the PV position itself, and the resolution of the PV x and z coordinates is shown in Figure 3.22. The x and y coordinates have equal resolution, while the LHCb geometry leads to a much worse resolution in z than the two transverse components. The more tracks pointing back to the PV, the more precisely its position can be determined, which can be seen in the figures. Although the z component of the IP is negligible, the PV's z resolution has a weak effect on the IP. Through the contribution of PV resolution to the IP resolution, it is to be expected that the IP resolution will also have a dependency on the number of tracks originating from the PV. These resolution effects are particularly important for the $R_{\tau\mu}$ measurement described in Chapter 9.

[6]Unfortunately, this is also the acronym for the LHC's interaction points as used at the beginning of this chapter. In the remainder of this thesis, IP will stand for "impact parameter".

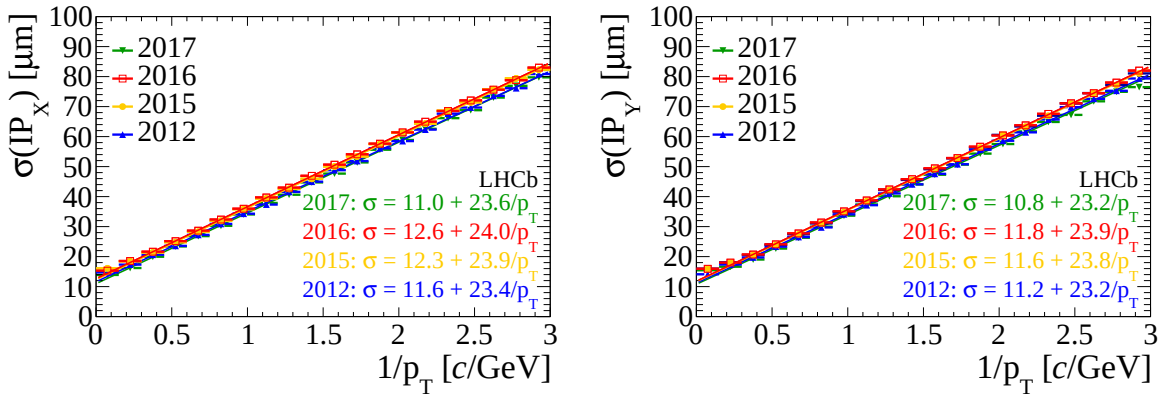


Figure 3.21: IP resolution in x (left) and y (right) as a function of the track's $1/p_{\mathrm{T}}$ for different data-taking periods. Taken from Ref. [51].

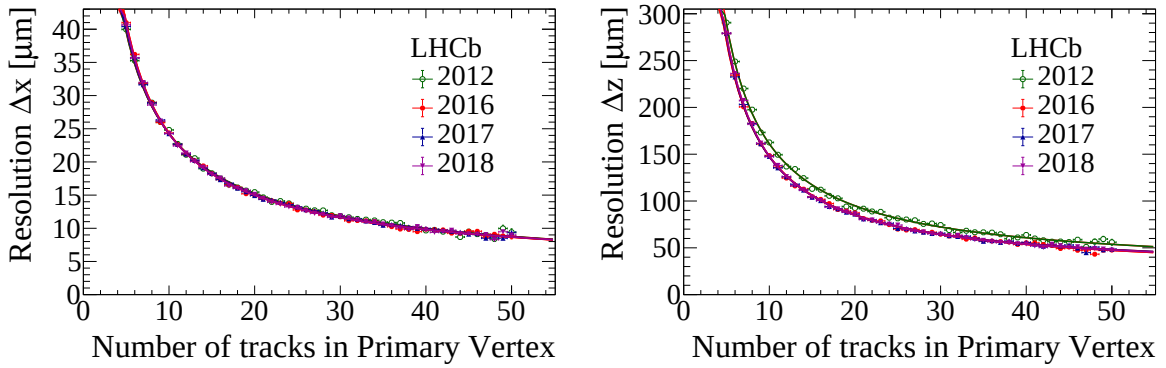


Figure 3.22: PV x (left) and z (right) resolution as a function of the number of tracks originating from the PV for each year in Run 2. A comparable set of points from Run 1 is also included. Taken from Ref. [51].

Many other quantities are of interest to EW analyses, and LHCb's performance in measuring them can be seen as they are analysed in the following chapters.

3.7 Overview of the LHCb simulation

Every LHCb analysis uses simulated events to some degree, but template-fitting analyses are particularly reliant on them as the templates are directly constructed from them. Ideally, any differences between simulated and real events come from the real physics phenomena under study, and not some mismatch between how the data and simulation are processed. LHCb tries to achieve this by treating the data and simulation in exactly the same way as early as is possible in their production chain, as is shown in Figure 3.23. The simulation workflow begins in a software application known as Gauss [89], which directs an event generation step with Pythia [46], followed by the decay of any unstable particles with EvtGen [90], and finally the simulation of interactions with the detector

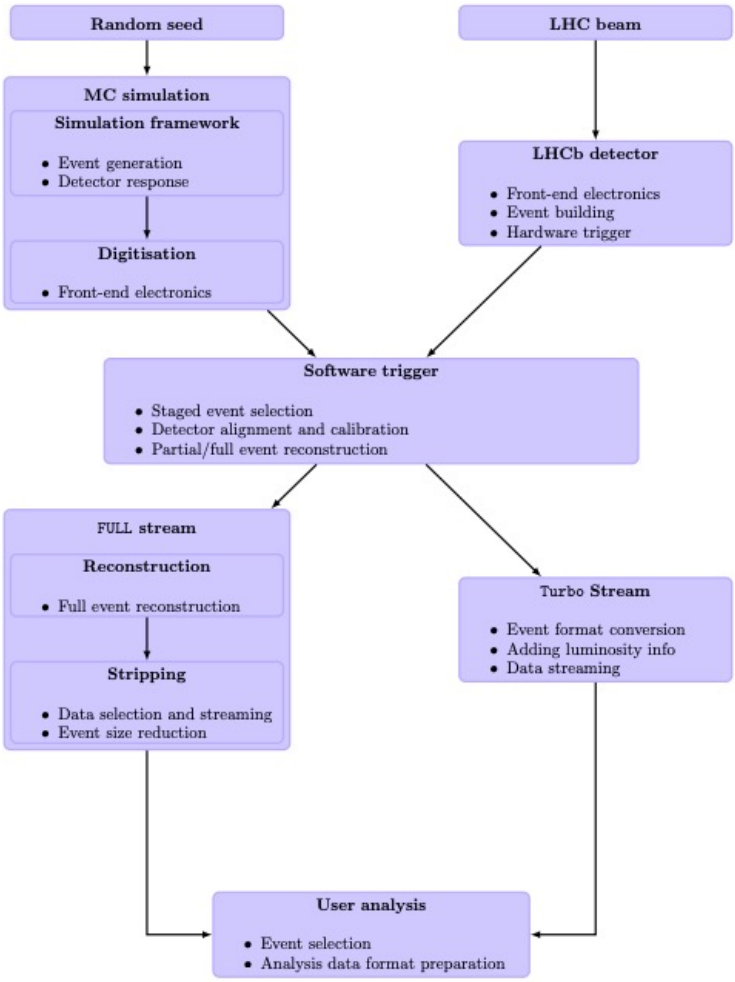


Figure 3.23: Schematic of the processing workflow for real and simulated data in Run 2. Taken from [86].

by GEANT4 [91, 92]. In the next stage, the software application BOOLE simulates the response of front-end electronics and digitization [93], after which the simulated data and real data can be treated equally by further processing steps. Like the real data would be, the next step for the simulated events is a pass through the trigger system, and so on down LHCb data processing chain described in the previous sections.

Whilst this production chain avoids many potential discrepancies between data and simulation, many differences remain, and they grow in importance as the target precision of a template-fitting analysis increases. For example, it is ultimately impossible to fully capture the real detector's geometry and behaviour in the GEANT4 simulation with perfect precision and accuracy. Furthermore, the generation step in PYTHIA will only predict the physics process to a certain order of perturbation theory. Both in the physics and detector modelling, there are assumptions, predictions and simplifications everywhere in the LHCb simulation. Studying the effects of these for EW analyses is the

subject of Chapters 5 and 6.

3.8 Outlook: the LHCb upgrade

The LHCb detector described in this chapter was shut down, along with the LHC, at the end of 2018 to begin the period known as Long Shutdown 2 (LS2). During LS2, LHCb underwent a major upgrade to prepare for a new cycle of data taking at higher instantaneous luminosity. To cope with the increased luminosity, more granularity and radiation tolerance was required, and over 90% of LHCb's active detector channels were to be replaced [94]. At the time of writing, commissioning of the upgraded detector and its trigger system are progressing well, although the installation of the UT has been delayed. The trigger system has also been completely redesigned: L0 has been removed completely; and a new HLT1 will process the full bunch crossing rate. Every trigger line at each level has been re-written, and all these lines must be optimized to fully exploit the new detector. This thesis turns next towards this effort of line-writing, which the author has contributed to by creating an automated tool for test-driven trigger line development, and by writing new lines for high-p_{T} muon physics in Run 3.

CHAPTER 4

Development of tools and selections for the Run 3 LHCb trigger

At the end of the last chapter the upgrade of the LHCb experiment for Run 3 was briefly introduced. In Run 3, to probe even deeper in the search for signs of physics beyond the Standard Model, LHCb will run at higher instantaneous luminosity and become the first hadron collider experiment to utilize a fully software-based trigger. However, the expected gains in signal yields manifest also as significant challenges in fitting within the confines of the available data storage capacity. This necessitates a set of highly optimized trigger selections. This Chapter will describe the author's contribution to this challenge, which begins with the creation of an automated software tool to facilitate the test-driven development of trigger lines. The motivation for this tool is detailed in Section 4.1, followed by its implementation and validation in Sections 4.2 and 4.3, and a showcase of its abilities in Section 4.4. These sections reproduce and expand on the published work (by the author) in Ref. [95]. Sections 4.5 and 4.6 then turn to the author's usage of the tool, firstly in comparing the physics performance of the CPU- and GPU-based proposals for HLT1 in Run 3, and in developing trigger lines for electroweak physics in Run 3. A summary and outlook on this work is given in Section 4.7.

4.1 Motivation for a trigger performance optimization tool

In Run 3 (2022-2025), the upgraded LHCb detector [96] will take data at an instantaneous luminosity of $\mathcal{L}_{\text{int}} = 2 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$ - five times that of Run 2, yielding on average five to six proton-proton (pp) interactions per bunch crossing [86]. To cope with this, the

trigger system has been completely redesigned. The Level-0 (L0) hardware trigger has been removed, and instead the upgraded detector will be readout after every pp bunch crossing, with the new, GPU-based, first stage of the High Level Trigger (HLT), HLT1, processing the full 30 MHz pp visible interaction rate[1]. Like in Run 2, HLT1 will perform a partial event reconstruction, whereas HLT2 will provide an offline-quality reconstruction with the best-available alignment and calibration. Via trigger lines at HLT1 and HLT2, the data rate to permanent storage will be reduced by three orders of magnitude. Coupled with the increase in interactions per event, removal of L0 is also projected to increase the trigger efficiency on typical hadronic decays of beauty hadrons by a factor of two [96, 97], leading to a typical 10× increase in signal yields per unit time with respect to Run 2. Further gains in physics reach are also being made at the time of writing due to the high technological ceiling of the GPU-based HLT1, named `Allen` [98].

The gains in signal yields lead also to challenges in terms of available data storage. The available capacity of permanent storage restricts the output bandwidth - defined as the rate of events output multiplied by the average event size - of HLT2 to 10 GB/s [86]. Section 3.5 outlined the considerable efforts made in the reduction of the output event size by the introduction of the Turbo stream in Run 2, but this alone is insufficient. Trigger lines must also be highly-optimized to find the right balance between a high signal efficiency and a low-enough trigger rate to not overload the available storage.

The task of optimizing a trigger line for Run 3 therefore had to be made easier than in previous years. Line development in LHCb has hitherto been conducted by independent line authors with no centralized or automated tools in place. This strategy doesn't use the available person-power efficiently, and is a recipe for inconsistent approaches, definitions, and standards of validation. The aim was to ameliorate these issues by providing a user-friendly, automated tool that all line authors can use, with transparent and well-motivated definitions of efficiencies and rates.

When trigger efficiencies were quoted in Run 2, they were typically "TOS" (*trigger-on-signal*) efficiencies [51, 66, 99]. A basic trigger efficiency would be calculated as the number of events in a sample that fire the trigger, divided by the total size of the sample. However, in a busy hadron collider environment, a positive trigger decision could be caused by something other than the presence of a signal decay. The real quantity of interest is the efficiency at which the trigger is firing on the signals it was designed to select. The TOS efficiency gives an estimate of this: its numerator is the number of events with positive trigger decisions, where also the trigger *candidate* (the software object representing e.g. the decay of a B meson, made by reconstructing sub-detector hits) matches to a signal candidate reconstructed with a high-quality offline reconstruction and selection. Since the

[1]The bunch crossing frequency is 40 MHz, but only roughly 30 MHz of these events produce a visible signal in LHCb.

latter was executed offline, it could be complex enough such that it would make a very pure selection, and hence a very pure denominator of signal candidates with which to evaluate the trigger efficiency against. However, calculating efficiencies with respect to the offline reconstruction adds a great deal of complication, bias (the offline reconstruction and selection may not give 100% purity) and inconsistency (the configuration of the offline reconstruction and selection will vary wildly from decay to decay). You also cannot perform an offline reconstruction unless the full raw event is saved - i.e. persisting via the Full stream - whereas the majority of lines in Run 3 will utilize the Turbo stream, discarding the raw event.

However, although this method may be broadly impractical for Run 3, a "TOS-like" efficiency is still a necessity. Regardless of its true origin, a candidate reconstructed and selected by a Turbo-persisting line will subsequently be treated as signal. The rate at which these candidates are "false-positives" - e.g. they were caused by a spurious or partially misreconstructed signal - is a crucial quantity for a line author. Even if part or all of the raw event is persisted, in reality only a TOS-like efficiency can be well-enough understood for use in a typical physics analysis, and thus this efficiency gives the realistic performance that can be expected. It was therefore a key aim to define and validate an efficiency like this that can be used in the absence of an offline reconstruction.

To achieve the aims stated here, the `HltEfficiencyChecker` software package was constructed.

4.2 Implementation of HltEfficiencyChecker

In the most basic sense, `HltEfficiencyChecker` gives the rates and efficiencies of trigger lines. These metrics are evaluated on simulated events, so the tool can be used now, before data-taking. The trigger efficiency is defined as

$$\varepsilon(\mathrm{DEC}) = \frac{\mathrm{N_{Triggered}}}{\mathrm{N_{Pre\text{-}selected}}}\,, \tag{4.1}$$

where $N_{\text{Pre-selected}}$ is the number of simulated signal candidates passing a pre-selection cut, or *denominator requirement*. This is sometimes referred to this as the "decision" (DEC) efficiency, to differentiate from the "matched" efficiency described below. To facilitate consistent efficiency definitions across many users/signals, `HltEfficiencyChecker` has a limited set of denominators from which the user can choose. Addition of new denominators is possible, but flexibility is deliberately not built in here - consistency was the aim. The default denominator requirement - denoted `CanRecoChildren`, abbreviated as C.R.C - requires that all final-state children are charged and are reconstructible as long tracks (hits in the VELO and downstream T stations)

in $2 < \eta < 5$, which is applicable to a large number of decays. It is also inclusive: it represents a simple but ambitious goal that the trigger line should catch any decays in LHCb's fiducial acceptance. All these requirements are made at the "truth-level" - i.e. on simulated kinematic quantities - rather than on reconstructed quantities. With LHCb's excellent reconstruction, the difference is expected to be negligible in most cases, but any difference shows up as an inefficiency of the trigger that the author would wish to know.

The natural extension to this to give a TOS-like efficiency is defined here (to avoid confusion) as the "TrueSim" efficiency:

$$\varepsilon(\text{TrueSim}) = \frac{\text{N}_{\text{Triggered \& Matched}}}{\text{N}_{\text{Pre-selected}}}, \tag{4.2}$$

where, to get into the numerator, an event must fire the trigger and at least one trigger candidate (many are possible, particularly for simpler and more inclusive lines) must match to the true simulated candidate. The matching algorithm, taking inspiration from the Run 2 TOS efficiencies, is as follows. Firstly, it breaks down each trigger candidate into its constituent reconstructed tracks, and collects the sub-detector hits from each track and stores them in one container, but retaining the knowledge of which hits come from which track, and which tracks come from each candidate. The simulated sub-detector hits of the simulated decay are similarly collected and stored. The trigger candidate's tracks are then iterated through, and if a minimum matching fraction $f = 0.7$ or more of the trigger candidate's hits are present in one of the simulated candidate's tracks, then the track is matched. Every track in the trigger candidate is required to match to a track in the true simulated candidate. If this is the case, then the algorithm declares that the trigger has fired on the true simulated signal in that event. This matching requirement is carefully worded. For example, an inclusive line looking for one high-p_{T} track may fire on one of the final-state muons of a decay like $B^0_s \to \text{J}/\psi\phi$ (with the J/ψ (ϕ) decaying to a pair of muons (kaons)), and with this matching requirement, a trigger candidate matching to one muon will be counted positively towards the TrueSim efficiency with respect to this muon, the intermediate J/ψ, and also the parent B^0_s. It will not count positively towards the TrueSim efficiency of the other three final-state particles, or the intermediate ϕ.

For code simplicity, the first iteration of this matching algorithm did not match on a track-by-track basis, but rather collected all hits from all tracks into a single, un-indexed container and required $f = 0.7$ or more of them to be present across the whole simulated signal. When showing the results from this algorithm in the next section, it will be referred to as the "prototype" matching algorithm; the track-by-track matching algorithm will be referred to as the "production" algorithm.

The other key metric of trigger line performance is the trigger rate, defined as

$$\text{Rate} = \frac{N_{\text{Triggered}}}{N_{\text{Events}}} \times \text{Input Rate}\,, \tag{4.3}$$

which is calculated by running the trigger over "minimum-bias" simulation, which is a proxy for real data-taking conditions. The input rate depends upon which stage of the trigger is under study: if HLT1, then the appropriate rate is the *pp* interaction rate of 30 MHz; if HLT2, an estimate of the HLT1 output rate (around 1 MHz) should be used.

Technically, the `HltEfficiencyChecker` package was built as an extension to the LHCb HLT software application `Moore`. Via a simple script, the user configures the trigger (HLT1, HLT2 or both), sets the simulated sample to run over and tells the package what information they would like to see. This script is then passed to `HltEfficiencyChecker` in a single shell command, which starts the running of the configured trigger job as a subprocess. The trigger decisions; as well as the simulated kinematic truth-level information required to perform the matching and plot efficiencies as a function as a function of decay kinematics; are saved to an output file. A second subprocess is then called, which processes that file to calculate and display the performance metrics of choice.

4.3 Validation of matching procedure

Before releasing the tool to line authors, it was crucial to validate the matching procedure, particularly the choice of minimum matching fraction $f = 0.7$. To accomplish this, a simple test setup of the high-level trigger was configured. A simulated sample of a standard decay of interest in LHCb was chosen: $B_s^0 \to \text{J}/\psi\phi$ (with $\text{J}/\psi \to \mu\mu$ and $\phi \to KK$). Two trigger lines were chosen: one building candidates with a single, displaced track; and the other building a displaced two-track vertex. These are similar to the one- and two-track lines that will be present in LHCb's HLT1 in Run 3. This choice enables study of the matching algorithm's behaviour against two differing topologies: the one-track line would be expected to trigger on one of the muons or kaons; while the two-track should fire on the J/ψ and ϕ combinations. In both cases, the selections are based on multivariate analysis (MVA) classifiers trained to find the products of heavy-flavour decays, but the specifics of the lines are unimportant here.

The absolute value of an efficiency is always subjective; it is highly dependent on the definition via the denominator requirement. In the case of the TrueSim efficiency, it is also expected to depend highly on the minimum matching fraction f, which is the only free parameter choice in the matching. Due to this subjectivity, the absolute value of the efficiency is not very important, provided that it is not biased to higher values by false positives (for example from spurious tracks that overlap with the signal track, and the pattern recognition has amalgamated the hits from both erroneously) or biased to lower

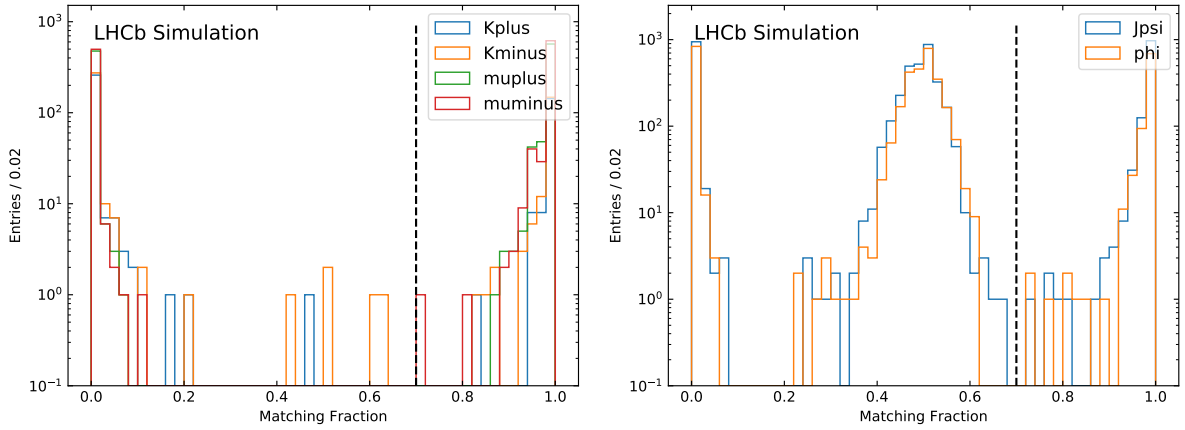


Figure 4.1: Histograms of the trigger candidate matching fractions to each of the true signal candidates, for the one-track line (left) and the two-track line (right) for triggered $B_s^0 \to J/\psi\phi$ events. The dashed line is at a matching fraction of 0.7.

values by rejecting perfectly good triggers that miss one or two simulated hits. The choice of minimum matching fraction should be optimized to avoid these two extremes. The Run-2 TOS efficiencies used a minimum matching fraction of around $f = 0.7$ (varying slightly for hits in different subdetectors), as does the algorithm commonly used to match reconstructed tracks to simulated tracks in calculating LHCb's track reconstruction efficiency [100]. However, no record of studies motivating this choice of $f = 0.7$ was found.

Figure 4.1 shows a histogram of the matching fractions of trigger candidates from the one- and two-track lines in response to $B_s^0 \to J/\psi\phi$. Here, the "prototype" matching algorithm has been used, which does not match on a track-by-track basis. Trigger candidates from the one-track line are strongly polarized towards matching fractions of 0 and 1, with very little in-between. However, the tail of candidates at low matching fractions illustrates that spurious tracks can partially overlap with the signal track at a non-negligible rate. It can be inferred that the tail falling away from 1 comes from well-matched tracks, but a handful of extra spurious hits have been added to the trigger candidate, or a small fraction of true hits have not been reconstructed. All of these should be true positives, so the minimum matching fraction shouldn't be set too close to 1. The large gap between 0 and 1 shows that it is extremely unlikely that a false positive can occur if a minimum matching fraction of around 0.7 is chosen.

In the two-track case, the same populations at 0 and 1 are observed, but also now a large population around 0.5 appears, which can be identified as belonging to two-track combinations where only one of the two tracks matches with the true signal candidate. These candidates should not be considered to be good matches. The broadness of the peak reflects the tail of overlap present near 0 and 1, and that the two tracks may have differing total numbers of hits. In both these simple cases, $f = 0.7$ appears to be sufficient

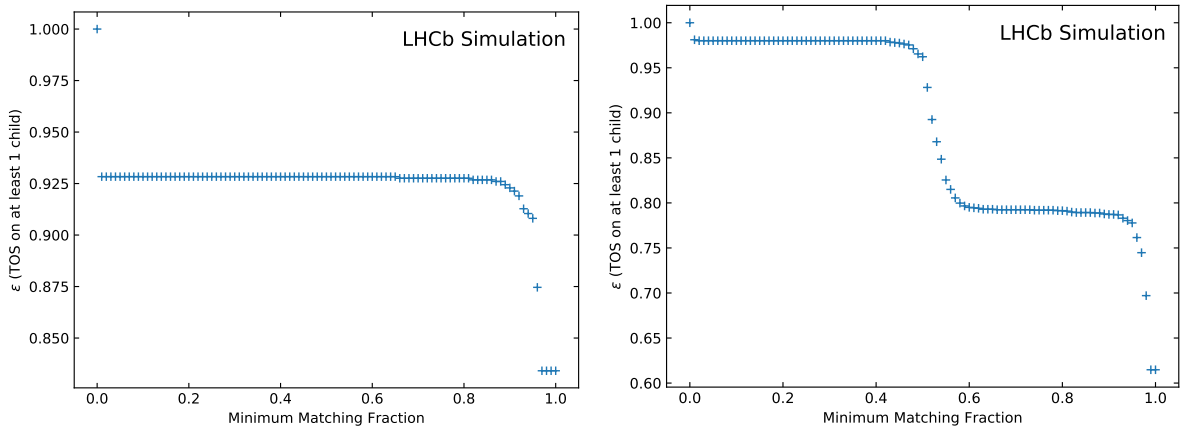


Figure 4.2: The efficiency that at least one one-track trigger candidate matched one of the kaons or muons in triggered events (left) and the efficiency that at least one two-track trigger candidate matched the J/ψ or ϕ in $B_s^0 \to$ J/ψϕ (right), as a function of the minimum matching fraction requirement.

to pick out the well-matched population, and the separation of the populations suggests that the TrueSim efficiency should not be a strong function of f near 0.7. This is indeed observed, as can be seen in Figure 4.2. Also observed here are large jumps near 0, 1, and 0.5 in the two-track case, which is asserted to be the inclusion of false positives or the rejection of good matches.

In conclusion, the prototype matching algorithm is performing as intended, and $f = 0.7$ is a good choice for the minimum matching fraction. The TrueSim efficiency here is stable around 0.7, which gives confidence that the absolute choice of f does not have a large impact. However, the two-track case begins to illustrate the problem with the prototype matching algorithm. Here, a population of trigger candidates appeared at 0.5, which was identified with one of the tracks being erroneous. If this was extended to a three- or four-track line, similar populations at 0.67 and 0.75 respectively would appear. These are unsettlingly close to the choice of $f = 0.7$, and might lead to a non-negligible amount of false positives in the four-track case.

Switching now to an early version of a line designed to trigger exclusively on $B_s^0 \to$ J/ψϕ at HLT2, these false positives begin to creep in, as can be seen in Figure 4.3. This histogram shows trigger candidates for this 4-track decay that are declared as "matched" by the prototype algorithm, as a function of the number of tracks that individually match to a true simulated track. There is a population, albeit small, of matched candidates where 1 of the 4 tracks does not match to a track from the simulated candidate - a population of false positives. By design, these false positives are removed by switching to the "production" version of the algorithm, which matches on a track-by-track basis. For simpler trigger candidates like the one- and two-track line above, this change is of no consequence. Changing the algorithm to the production version of track-by-track

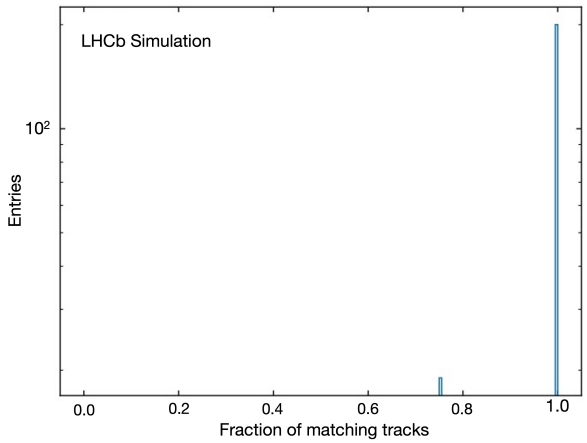


Figure 4.3: Histogram of $B_s^0 \to J/\psi\phi$ trigger candidates that were declared as matched by the prototype matching algorithm, plotted as a function of the fraction of tracks that individually match to a track in the true simulated decay.

matching made a per-cent level difference in the TrueSim efficiency for this HLT2 trigger line on $B_s^0 \to J/\psi\phi$.

4.4 Line tuning with HltEfficiencyChecker

Having outlined how `HltEfficiencyChecker` works, and validated its matching of trigger candidates to simulated decays, it is now possible to showcase some of the results that it gives to aid line tuning, as was the main aim of this work. There are a great deal more possibilities than what are shown here, and more of that capability will be seen in the following sections. The plots shown here should *not* be taken as indicative of LHCb's Run 3 trigger performance.

Sticking again with $B_s^0 \to J/\psi\phi$ and the simple one-track lines, Figure 4.4 shows the rate on minimum bias and the trigger efficiency (not the TrueSim efficiency for now) for six versions of the same line, with small arbitrary variations made in each case for the three MVA parameters. Plots like this facilitate the important task of balancing rate against efficiency. On the right-hand panel of Figure 4.4 is the trigger efficiency of three of the variations as a function of the true B_s^0 simulated p_{T}, which illustrates the typical trend that higher-p_{T} decays are easier to trigger on. Plots like this give substantial insight into the behaviour of a trigger line.

The DEC efficiency, and the TrueSim efficiency with respect to each child of the decay, can be plotted against a variety of true kinematic quantities. By default, the user will see both types of efficiency on the same plot to give them an idea of what fraction of trigger decisions are on the true signal. An example of this is given in Figure 4.5 for the one- and two-track lines as a function of $p_{\mathrm{T}}(B_s^0)$ in $B_s^0 \to J/\psi\phi$. The difference between

the two types of efficiency here is small, but non-negligible, perhaps reflecting the general ease of reconstructing this particular signal. Other signals, such as low-p_{T} multi-body charm-hadron decays, may not show quite the same agreement due to the abundance of low-p_{T} pions and kaons in a typical pp collision.

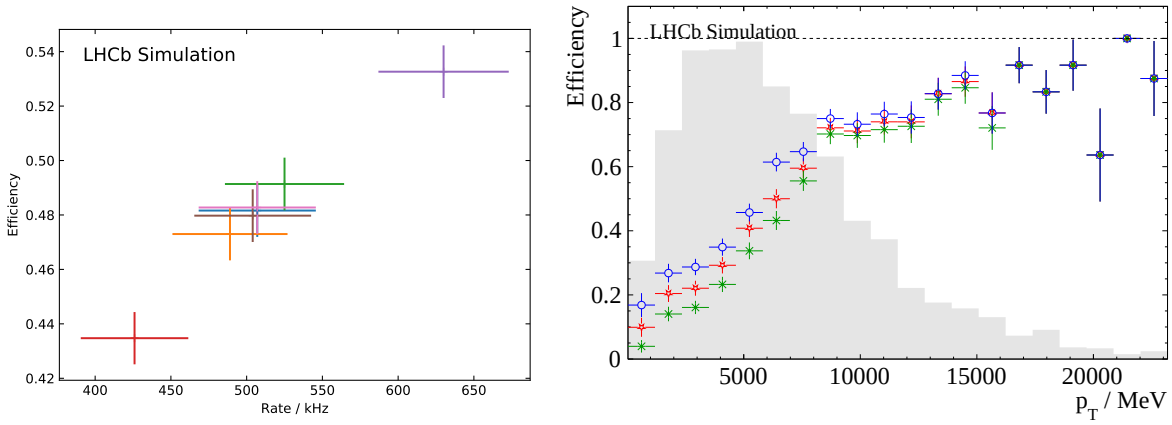


Figure 4.4: Scatter plot of the rate on minimum bias and the DEC efficiency on $B_s^0 \to \mathrm{J/\psi}\phi$ for the one-track line, plus six toy variations of the same line with different MVA parameters (left), and the DEC efficiencies on $B_s^0 \to \mathrm{J/\psi}\phi$ of the one-track line (blue), plus two toy variations of the line (red, green), plotted as function of B_s^0 transverse-momentum $p_{\mathrm{T}}(B_s^0)$ (right). The scaled $p_{\mathrm{T}}(B_s^0)$ distribution is under-laid in grey in the latter case.

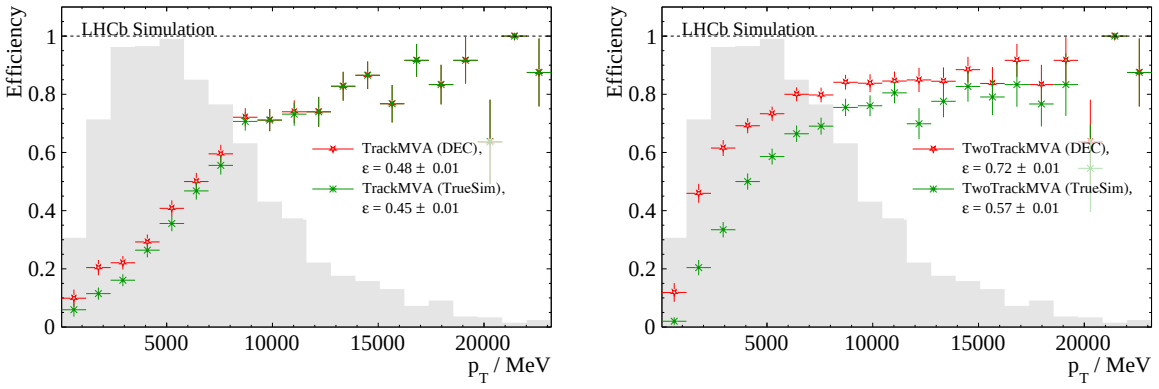


Figure 4.5: The DEC and TrueSim efficiencies of the one-track line (left) and two-track line (right) in response to $B_s^0 \to \mathrm{J/\psi}\phi$, as function of B_s^0 transverse-momentum $p_{\mathrm{T}}(B_s^0)$. The scaled $p_{\mathrm{T}}(B_s^0)$ distribution is underlaid in grey, and the efficiencies integrated over $p_{\mathrm{T}}(B_s^0)$ are listed in the legend.

4.5 Physics performance comparison of Run 3 HLT1 implementations

An important milestone in the preparation of the LHCb's new trigger system for Run 3 was the choice to implement HLT1 on GPUs [98, 101]. This choice was informed by a "cost-at-equal-performance" comparison study of CPU- and GPU-based HLT1 solutions that were available at the time, and has been published in Ref. [102]. As part of this study, the author used `HltEfficiencyChecker` to provide a comparison of the high-level performance of each HLT1 solution in terms of trigger rates and efficiencies, which is described in this section. Here only the DEC efficiency was used, as the TrueSim efficiency matching algorithm had not yet been implemented. The `CanRecoChildren` denominator is used throughout. As with the previous efficiencies shown, these should *not* be taken as indicative of the final performance of LHCb's HLT1 in Run 3.

Spanning LHCb's broad physics programme, simulated samples of six representative decays were chosen for this comparison: $B_s^0 \to \phi\phi$, J/$\psi \to \mu\mu$, $B \to K^{*0}ee$, $B \to K^{*0}\mu\mu$, $D_s^+ \to KK\pi$, and $Z \to \mu\mu$, each having 10,000 events. Four generic trigger lines were used from each HLT1 implementation:

- `TrackMVA` is similar to the one-track line used in the previous validation study: it searches for a displaced track above a p_{T} threshold, where "displaced" is defined by a minimum impact parameter significance, or χ^2_{IP}, with respect to any primary vertex.

- `TwoTrackMVA` is similar to the two-track line used previously: it searches for two-track combinations of a good combination quality (indicated by a low vertex fit χ^2) above a p_{T} threshold. The CPU- and GPU-based implementations differ here slightly: the former uses an MVA trained on heavy-flavour decays to make these selections, whilst the GPU-based implementation is based on rectangular selection cuts, and an MVA implementation was still pending at the time of the study. Per-cent level differences are therefore expected between the two implementations.

- `DiMuonLowMass` searches for two-track combinations of very low p and p_{T}, but with both having positive muon identification.

- `SingleHighPtMuon` selects single tracks of high p and p_{T} that are identified as muons.

Except for `TwoTrackMVA`, the requirements applied by each line were identical for the CPU- and GPU-based implementations.

4.5.1 HLT1 efficiencies

Figures 4.6 and 4.7 show the trigger efficiencies of the two HLT1 implementations as a function of parent p_{T} and parent decay time, τ (where applicable), for four of the signal samples. For figure clarity, only one trigger line (for CPU and GPU) is shown per plot. Tables 4.1 and 4.2 give the trigger efficiencies integrated over the decay kinematics for all four selections and all six signals. In each case, the trigger efficiencies are similar from the two HLT1 implementations.

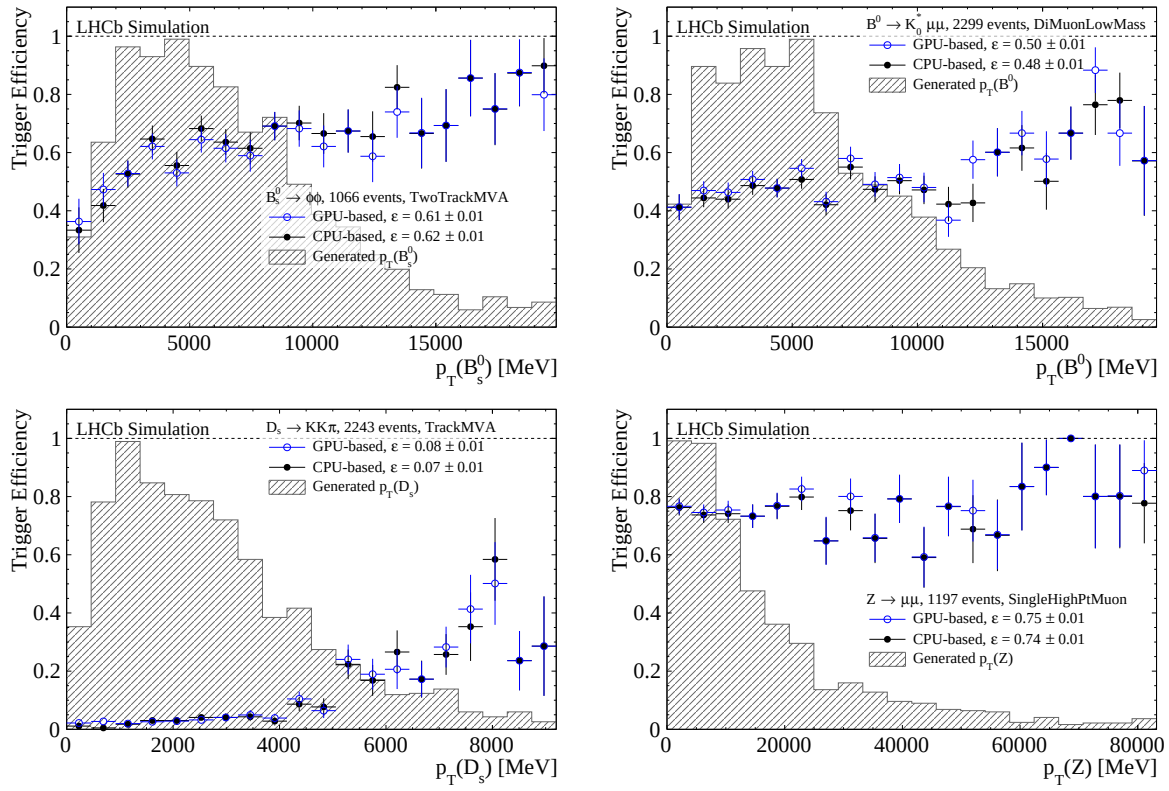


Figure 4.6: Trigger efficiencies for CPU-based and GPU-based HLT1 as a function of parent transverse momentum, with the `CanRecoChildren` denominator. Results are shown for the `TwoTrackMVA` (top left), `DiMuonLowMass` (top right), `TrackMVA` (bottom left) and `SingleHighPtMuon` (bottom right) selections firing on the $B_s^0 \to \phi\phi$, $B \to K^{*0}\mu\mu$, $D_s^+ \to KK\pi$ and $Z \to \mu\mu$ signal samples, respectively. The generated parent transverse momentum distribution is also shown for all events passing the denominator requirement.

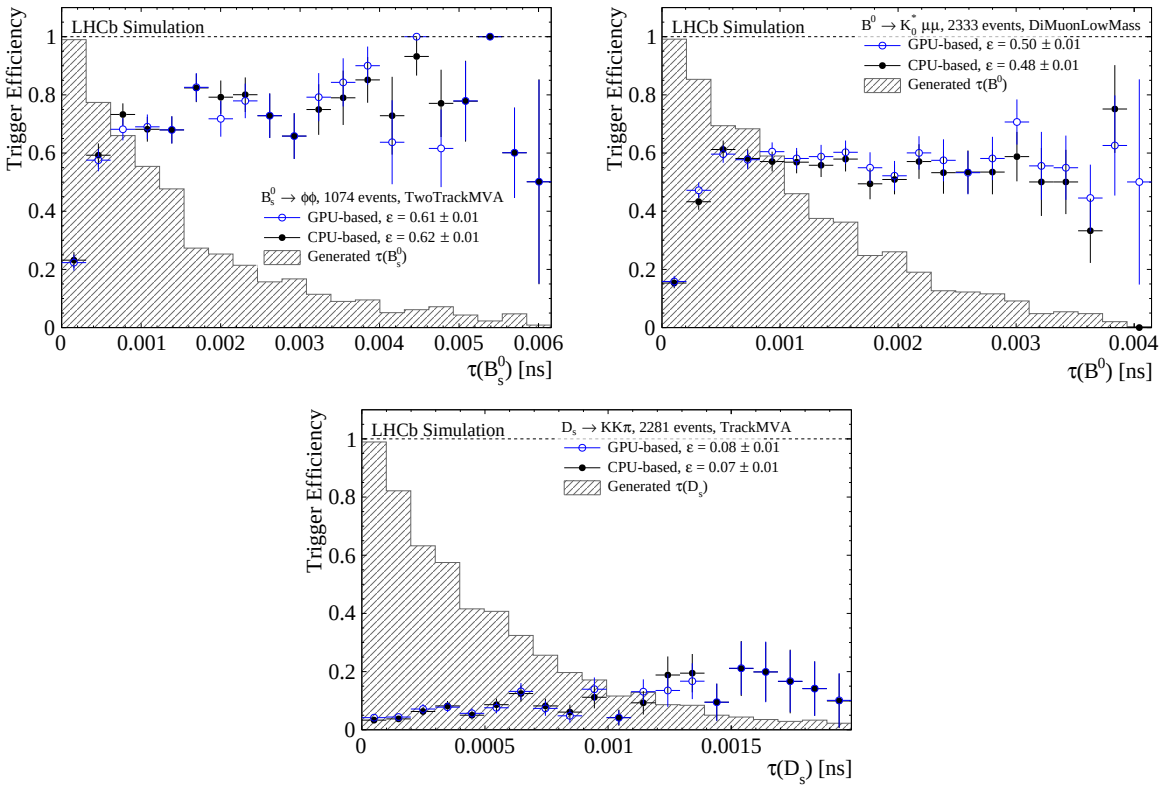


Figure 4.7: Trigger efficiencies for CPU-based and GPU-based HLT1 as a function of parent decay time, with the `CanRecoChildren` denominator. Results are shown for the `TwoTrackMVA` (top left), `DiMuonLowMass` (top right) and `TrackMVA` (bottom) selections firing on the $B_s^0 \to \phi\phi$, $B \to K^{*0}\mu\mu$ and $D_s^+ \to KK\pi$ signal samples, respectively. The generated parent decay time distribution is also shown for all events passing the denominator requirement.

Table 4.1: Comparison of trigger efficiencies integrated over the kinematic phase space of the candidates, for each of the six simulated signal samples for the `TrackMVA` and `TwoTrackMVA` selections. Statistical uncertainties are indicated in parentheses. The `CanRecoChildren` denominator has been used.

	`TrackMVA`		`TwoTrackMVA`	
Signal	GPU	CPU	GPU	CPU
$B_s^0 \to \phi\phi$	0.340(14)	0.332(14)	0.606(15)	0.621(15)
J/ψ $\to \mu\mu$	0.034(4)	0.031(3)	0.049(4)	0.042(4)
$B \to K^{*0}ee$	0.276(10)	0.278(10)	0.439(12)	0.473(12)
$B \to K^{*0}\mu\mu$	0.391(10)	0.385(10)	0.554(10)	0.582(10)
$D_s^+ \to KK\pi$	0.076(5)	0.073(5)	0.178(8)	0.193(8)
$Z \to \mu\mu$	0.051(6)	0.040(6)	0.024(4)	0.028(5)

Table 4.2: Comparison of trigger efficiencies integrated over the kinematic phase space of the candidates, for each of the six MC signal samples and the `DiMuonLowMass` and `SingleHighPtMuon` selections. Statistical uncertainties are indicated in parentheses. The `CanRecoChildren` denominator has been used.

	`DiMuonLowMass`		`SingleHighPtMuon`	
Signal	GPU	CPU	GPU	CPU
$B_s^0 \to \phi\phi$	0.025(5)	0.024(5)	0.005(2)	0.004(2)
$J/\psi \to \mu\mu$	0.078(5)	0.067(5)	0.048(4)	0.045(4)
$B \to K^{*0}ee$	0.024(4)	0.027(4)	0.0011(8)	0.0011(8)
$B \to K^{*0}\mu\mu$	0.502(10)	0.482(10)	0.091(6)	0.088(6)
$D_s^+ \to KK\pi$	0.018(3)	0.019(3)	0.0013(7)	0.0013(7)
$Z \to \mu\mu$	0.033(5)	0.036(5)	0.749(12)	0.740(13)

4.5.2 HLT1 rates

Using 10,000 minimum bias events, the HLT1 rate - as defined by Equation 4.3 - was measured for each of the four representative trigger lines on each architecture, as shown in Figure 4.8. The rates of `TrackMVA`, `DiMuonLowMass` and `SingleHighPtMuon` are similar in each case, but the differing implementation of the `TwoTrackMVA` line on CPU and GPU (the GPU-based implementation was still based on rectangular selection cuts - not an MVA - at the time of the study) leads to a discrepancy of around 30%. The inclusive rate of the four selections was measured to be 912 ± 52 (798 ± 48) kHz for the GPU-based (CPU-based) implementation, which is largely due to the `TwoTrackMVA` lines.

4.5.3 Global Event Cut efficiency

As was described in Section 3.5, most of the L0 trigger lines in Run 2 had a global event cut (GEC) on the number of hits in the SPD sub-detector, rejecting those events that were most time-consuming to reconstruct. In Run 3, the GEC will instead be based on the number of clusters in the new tracking stations before and after the magnet. The same requirement was applied in both the CPU- and GPU-based HLT1 options, and consequently, the GEC efficiencies for each sample were found to be same for both architectures. This efficiency was 0.75 ± 0.01 on $Z \to \mu\mu$, whilst the other B and D decay samples under study had GEC efficiencies of about 85%, with statistical uncertainties of $\sim 1\%$. The efficiency on minimum bias events was 0.931 ± 0.003.

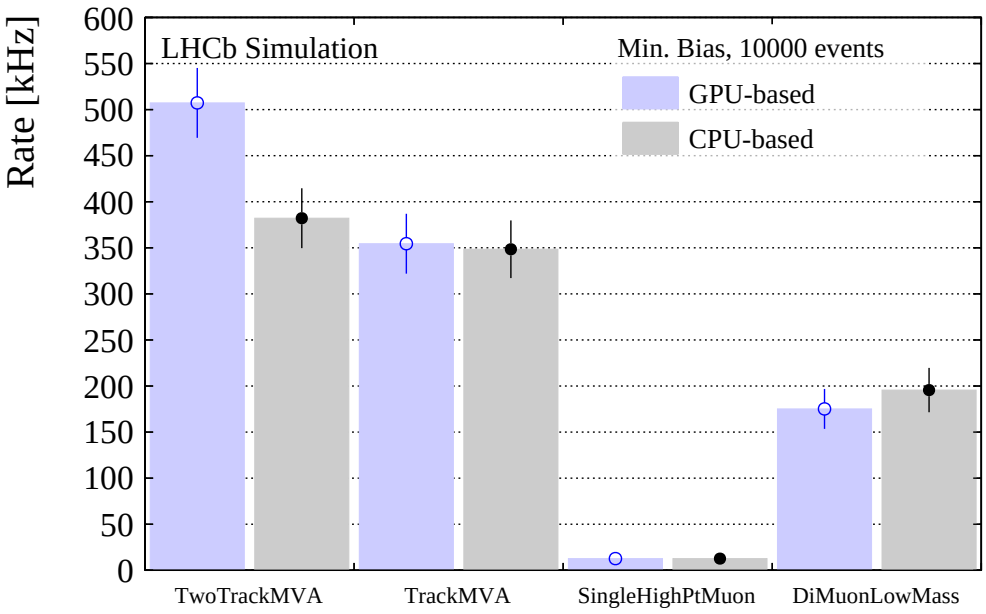


Figure 4.8: Trigger rates of the CPU- and GPU-based implementation for the four trigger selections of interest. The difference in the rate of the two `TwoTrackMVA` selections can be explained by the differing implementation.

4.5.4 Conclusions of the comparison study

The previous sections showed a near-identical high-level physics performance of the two HLT1 options in almost all aspects - the only exception being the `TwoTrackMVA` rates, which were understood to be due to the differing CPU- and GPU-based implementations of the line at the time of the study. Work by others presented in Ref. [102] came to similar conclusions with regard to lower-level quantities, such as the track reconstruction efficiency, momentum resolution and PV reconstruction efficiency. This was somewhat by design of the "cost-at-equal-performance" strategy that the comparison took, but nonetheless these conclusions were not trivial and not to be assumed. The conclusions enabled the decision-making process to come down to monetary costs, costs in person-power and projected future performance gains. Whilst operating a "hybrid" HLT - with GPU-based HLT1 and CPU-based HLT2 - adds to the overhead of software maintenance and developer training, the immediate cost savings, expectation of rapid advancement in GPU technology, and capitalization on emerging trends of software development led to the collaboration deciding to proceed with a GPU-based HLT1.

4.6 Trigger selections for high-p_{T} muon physics in Run 3

Sections 4.1 - 4.4 illustrated the development of `HltEfficiencyChecker` as a tool that line authors - primarily physics analysts - could use to help them write optimized trigger lines for Run 3. This section will now describe how the author used the tool in the writing of trigger lines at HLT1 and HLT2 for single, high-p_{T} muons in Run 3. The trigger strategy here is simple: achieving a high efficiency on EW signals should be straightforward, so push the p_{T} threshold as low as can be allowed by rate constraints. Descending in p_{T}, larger and larger amounts of background start to appear - pushing up the rate - but this therefore gives an important control region to help model the backgrounds in the higher-p_{T} region that will be selected, or allow for bigger signal samples in e.g. low-mass Drell-Yan ($\gamma^* \to \mu\mu$) or $W \to \tau\nu$. Background-dominated samples should also be selected by making copies of the lines without MuonID requirements. All rates and efficiencies shown in the following are calculated and plotted with `HltEfficiencyChecker`.

4.6.1 High-p_{T} single muon lines in HLT1

The CPU-based `Hlt1SingleHighPtMuon` line used in the physics performance comparison of the previous section was developed by the author, and simply requires a track to be positively identified as a muon, $p > 6\,\mathrm{GeV}$, and $p_{\mathrm{T}} > 6\,\mathrm{GeV}$. Subsequently, the GPU-based version of the line was written by collaborators and aligned to have these requirements, alongside a loose cut on the track fit quality $\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 100$. These requirements are taken from the `Hlt1SingleMuonHighPt` line in Run 2, described in Section 3.5. The performance of the new line was shown in the previous section, where it is important to notice that the integrated efficiency of around 75% is conspicuously close to the GEC efficiency on $Z \to \mu\mu$. For clean, high-p_{T} muons coming from Z-boson decays, the efficiency of such a simple selection should be close to 100% - the only significant rejection is coming from the GEC. At the time of writing, work has begun to remove the GEC from `Hlt1SingleHighPtMuon`, however this is complicated by the tight constraints on throughput in HLT1. In Run 2, the vast majority of hard-to-reconstruct events failing the GEC would be thrown away by L0; only the tiny proportion of them that passed `L0MuonEW` would need to be reconstructed at HLT1. With no L0 in Run 3, removing the GEC for `Hlt1SingleHighPtMuon` means that a reconstruction must run on *every* event to work out whether there is a high-p_{T} muon or not. The associated drop in throughput (around 15-20%) cannot be absorbed. The future solution will be to write a simpler, faster reconstruction sequence feeding only `Hlt1SingleHighPtMuon` that takes a negligible amount of GPU time. This reconstruction can then be run *after* the default, with-GEC reconstruction sequence feeding all other

lines.

In addition to the signal-selecting `Hlt1SingleHighPtMuon` line, the `Hlt1SingleHighPtNoMuIDMuon` line has been written by the author for selecting control samples for the studies of backgrounds in the high-p_{T} muon datasets. In principle, the line is a copy of `Hlt1SingleHighPtMuon`, but the candidates have no requirement of hits in the muon chambers - they are just reconstructed long tracks with an assigned muon mass hypothesis. Ostensibly, the rate of such a selection is unsatisfactorily large, but has been controlled by increasing the requirement on p_{T} to 8 GeV, and by adding a *prescale* (randomly throwing away all but this fraction of events before attempting to reconstruction and selection) of 0.05. In April 2022 when this line was added, `Hlt1SingleHighPtMuon` (`Hlt1SingleHighPtNoMuIDMuon`) had a rate of approximately 15(40) kHz, which is very small compared to the main selections for B and D physics. `Hlt1SingleHighPtNoMuIDMuon` can be included in either the GEC or without-GEC sequences, as the absolute trigger efficiency of this selection is not important - typically only the kinematic shapes of these control samples are used in physics analysis.

4.6.2 High-p_{T} single muon lines in HLT2

The preliminary HLT2 high-p_{T} muon selections are described in Table 4.3, as well as their rate and TrueSim trigger efficiencies on 1,000 simulated $Z \to \mu\mu$ events. Isolation requirements and prescales facilitate lowering the p_{T} threshold, but this adds complexity (also in potential mismodelling effects), and reduces the trigger efficiency respectively. However, it is expected that `Hlt2SingleHighPtMuon` will be the main line for selecting signals for physics analysis; the other two will provide control regions. The `Hlt2SingleHighPtNoMuIDMuon` line picks up the muon-ID-unbiased control sample selected at HLT1.

Figure 4.9 also shows the TrueSim efficiency of `Hlt2SingleHighPtMuon` and `Hlt2SingleHighPtMuonIso`, as a function of p_{T}^{Z} and p_{T}^{μ}, with the `CanRecoChildren` denominator. The two other control lines are omitted as their efficiency is low by consequence of their design. The trigger efficiency is reassuringly flat; slopes as a function of p_{T} are problematic for shape measurements of the p_{T} spectrum. The rates at this stage are satisfactorily low, but as commissioning of the detector and trigger system on the real data progresses, all of these thresholds are subject to change.

Table 4.3: HLT2 trigger lines for high-p_{T} muons in Run 3. The common `Hlt2SingleHighPt` prefix in the line names are omitted for brevity. Rates were calculated on 50,000 minimum bias events which had been filtered by a semi-representative HLT1, and TrueSim trigger efficiencies are calculated by triggering and matching to one of the final state muons in 1,000 simulated $Z \to \mu\mu$ decays. The efficiency denominator requirement is `CanRecoChildren`, with the additional requirement of true muon $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. Statistical uncertainties on the last digit(s) are indicated in parentheses after the value where appropriate.

Name	Muon-ID?	Kinematic Selection	Prescale	Rate [kHz]	Efficiency
`Muon`	✓	$p_{\mathrm{T}}^{\mu} > 15\,\mathrm{GeV}$	-	0.04(2)	1(0)
`MuonIso`	✓	$p_{\mathrm{T}}^{\mu} > 12.5\,\mathrm{GeV}, p_{\mathrm{T}}^{\mathrm{cone}} < 10\,\mathrm{GeV}$	-	0.12(4)	0.972(10)
`MuonPrescale`	✓	$p_{\mathrm{T}}^{\mu} > 10\,\mathrm{GeV}$	5%	0.02(1)	0.056(14)
`MuonNoMuID`	✗	$p_{\mathrm{T}}^{\mu} > 15\,\mathrm{GeV}$	10%	0.36(8)	0.083(17)

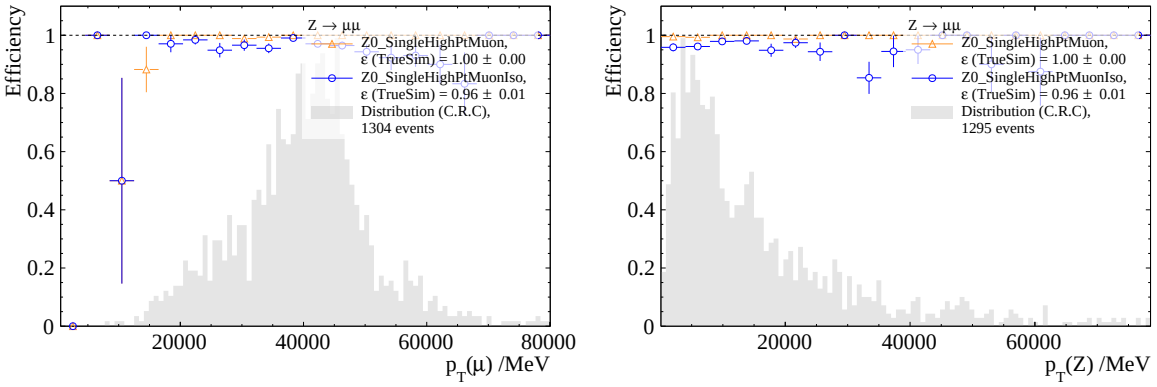


Figure 4.9: HLT2 TrueSim trigger efficiencies for `Hlt2SingleHighPtMuon` and `Hlt2SingleHighPtMuonIso` matched to the Z in $Z \to \mu\mu$ decays, as a function of the muon p_{T} (left) and Z-boson p_{T} (right). The `CanRecoChildren` denominator has been used in the efficiency calculation, and the generation kinematic distributions are underlaid in grey. Note that the integrated efficiencies that show as "1.00 ± 0.00" are in fact below 1 at the $\mathcal{O}(10^{-3})$ level, so round up to 1.

4.7 Summary and outlook

Run 3 places a substantial challenge on the shoulders of the new LHCb trigger system. The physics benefits of higher luminosity, removing L0, and embracing GPUs in HLT1 will be large, but only if they can be made to align with the constraints on HLT output rates and available permanent data storage capacity. These constraints necessitate a set of well-tuned trigger lines, that find the right balance between signal efficiency and rate. This chapter presented the `HltEfficiencyChecker` software package, which allows this important work to progress before data-taking in a consistent, test-driven and automated way. Since its release, it has been used widely across the LHCb collaboration.

As well as provided the trigger rate and simple DEC efficiency, the implementation and validation of an algorithm that matches of the trigger's reconstructed candidate to

the true simulated signal was also detailed. This allows computation of the so-called TrueSim efficiency, which is a measure of how often the trigger fires on the true signal in the event. In combination with the DEC efficiency, this tells the author the likelihood at which their line fires on spurious or partially misreconstructed signals, which is key in the upcoming era where Turbo lines - that persist only the trigger candidate to file - will be the new normal.

The remainder of the chapter showed how `HltEfficiencyChecker` can be useful in the line tuning procedure: first with a toy test case; then to the author's own uses of the tool in comparing the CPU- and GPU-based implementations of HLT1; and in helping write and tune the trigger lines that will select high-p_{T} muons in Run 3. In the near future, these lines will be put to use on real pp collision data.

CHAPTER 5

Precise modelling of W-boson physics

After spending the last chapter discussing preparations for LHCb's Run 3, this chapter and the remainder of this thesis now returns to Run 2 and the precision electroweak (EW) measurements that the author has contributed to. Both of the measurements are template fits to data of the Drell-Yan process, as introduced and motivated in Chapter 2. A key aspect of such measurements is the modelling of the underlying Drell-Yan physics: the vector boson production and decay. Predictions for this process are available from many different programmes, and it must be understood how dependent a measurement of an observable is on which prediction was used. This chapter describes the strategies developed to control and understand these physics modelling aspects at LHCb. These approaches were first developed by collaborators for the m_W measurement, where they are far more pertinent, and later incorporated into the $R_{\tau\mu}$ analysis by the author. As such, description of this work by others is necessary to fully understand the methods and results of later chapters. Chapters 8 and 9 will describe the application of (and show the results of) these strategies to the measurements of the W boson mass, and a test of its lepton flavour universality.

Chapter 2 ended by introducing the idea of reweighting a parton shower simulation to mimic a higher-order event generator in the hard process. The implementation of this Quantum Chromodynamics (QCD) reweighting strategy for these analyses is explained in Section 5.1. The evaluation of different candidate models for this reweighting is the subject of Section 5.2. Similar considerations are made for the Quantum Electrodynamics (QED) final-state radiation (FSR) model in Section 5.3. A final choice to be made in the physics modelling is the set of parton distribution functions (PDFs) to be used, which is described in Section 5.4.

5.1 Reweighting to higher-order QCD models

The first step in constructing templates is the event generation, which begins with simulation of the hard process. In the LHCb simulation, as introduced in Section 3.7, this step is executed with a LHCb-specific tune of Pythia 8 [103], whose leading-order (LO) predictions do not model the physics of vector boson production and decay with sufficient accuracy. However, it has been extensively validated for more than 10 years in LHCb, and gives a good description of the complete event, so serves as a useful base model. The full LHCb simulation chain does not easily allow the switching of the event generator, so the presence of a higher-order generator must instead be mimicked by reweighting the template events to their predictions.

The weights to be applied are calculated at the generator level (i.e. including no detector effects), as they factorize from all other parts of the simulation. This calculation begins by running the exact version of Pythia as is used by the LHCb simulation to make generator-level predictions, which form the denominator of the weights. Similar predictions are then made using a higher-order tool of choice, which become the numerator. These generator-level predictions are much faster than the full LHCb simulation, in which the main bottleneck is the detector simulation.

The perturbative QCD (pQCD) weights are in fact further factorized into two parts. There is also a third physics weight for the QED final-state radiation (FSR) treatment, which will be covered in Section 5.3. Eqn. 2.23 showed how the Drell-Yan differential cross section could be factorized into an unpolarized cross section σ^{unpol} and a series of angular terms. The QCD-related weight therefore factorizes into an unpolarized cross section weight and an angular weight.

The unpolarized cross section is a function of the vector boson's mass M, rapidity y and transverse momentum p_{T}^V, and so different models give different predictions for the unpolarized cross section in this 3D space. The numerical Monte-Carlo event generators that are being used do not give smooth continuous predictions but rather 3D histograms of $\mathrm{d}\sigma/\mathrm{d}M\mathrm{d}y\mathrm{d}p_{\text{T}}$, which can be made smoother or more finely-binned by increasing the number of events generated. The weight (denoted w_{unpol}) is then the ratio of the target generator's $\mathrm{d}\sigma/\mathrm{d}M\mathrm{d}y\mathrm{d}p_{\text{T}}$ to that of LO Pythia:

$$w_{\text{unpol}} = \frac{\frac{1}{\sigma_{\text{target}}}\frac{d^3\sigma}{dp_T dy dM}(\alpha_s, k_{\text{T}}^{\text{intr}})\big|_{\text{target}}}{\frac{1}{\sigma_{\text{Pythia}}}\frac{d^3\sigma}{dp_T dy dM}\big|_{\text{Pythia}}} \,. \tag{5.1}$$

Achieving sufficient granularity in (M, y, p_{T}) is computationally impractical; instead the points of a coarse histogram are linearly interpolated between to get the weight at the proper coordinates for each boson's kinematics. The 3D histograms are therefore referred

to as *interpolation histograms*. Each of the candidate programs has a number of tuning parameters that can also be varied to improve the model's agreement with the data. For example, in POWHEG (described shortly) it is possible to change the value of the coupling α_s and the amount of intrinsic transverse momentum in the proton $k_{\mathrm{T}}^{\mathrm{intr}}$. As was stated in Section 2.3, although these parameters represent real physics parameters, their values are highly dependent on the order and scales of the calculation, and so tuning them to fit the data is helping to account for missing higher orders. There is therefore an array of 3D interpolation histograms for each program, with each array member being produced with different values of the tuning parameters. A further (*bicubic spline*) interpolation is used to give continuous variation of the tuning parameters between the different histograms in each array.

The angular weight w_{ang} is similarly formed by the ratio of the angular terms of Eqn. 2.23, with the PYTHIA predictions for each angular coefficient in the denominator, and the target model in the numerator:

$$w_{\mathrm{ang}} = \frac{1 + \cos^2\vartheta + \frac{1}{2}A_0^{\mathrm{target}}(1 - 3\cos^2\vartheta) + ...}{1 + \cos^2\vartheta + \frac{1}{2}A_0^{\mathrm{pythia}}(1 - 3\cos^2\vartheta) + ...}\,. \tag{5.2}$$

The event generators provide predictions of differential cross sections, and so predictions for the angular coefficients must be extracted from the differential cross section using a projection method, as detailed in e.g. Ref. [42]. Referring back to Eqn. 2.23 it can be seen how this is done in principle: firstly, take the differential cross section; average it over ϑ and φ; use an appropriate spherical-harmonic projection function (the spherical harmonics are an orthogonal basis, so projecting out one term is possible); and finally divide away the unpolarized cross section. The angular coefficients are also a function of (M, y, p_{T}), and so there are eight such 3D interpolation histograms - one for each angular coefficient.

5.2 Evaluation of candidate pQCD models

In the context of the W mass measurement, several programs were studied for their ability to model the 5D differential cross section of Drell-Yan events in LHCb:

- PYTHIA-LHCb: Hard process generated at LO (α_s), plus parton shower, all with the same version of Pythia8 [46] as used in the LHCb simulation, including the same values of the tuning parameters α_s and $k_{\mathrm{T}}^{\mathrm{intr}}$.
- PYTHIA: PYTHIA 8 [46] as above, but with several values of α_s and $k_{\mathrm{T}}^{\mathrm{intr}}$.
- POWHEGPYTHIA: NLO hard process generated with POWHEGBOXV2 [104], interfaced to PYTHIA 8 [46] for parton shower.
- POWHEGHERWIG: NLO hard process generated with POWHEGBOXV2 [104],

showered with HERWIG.

- HERWIG: NLO hard process and parton shower fully implemented in HERWIG [47, 105].
- DYTURBO (NLO): Analytic resummation at NLL accuracy with matching to fixed order predictions at $\mathcal{O}(\alpha_s)$ accuracy [48].
- DYTURBO (NNLO): Analytic resummation at NNLL accuracy with matching to fixed order at $\mathcal{O}(\alpha_s^2)$ accuracy [48].

In the early development of the m_W analysis, it was concluded that DYTURBO is the only available program that predicts the angular coefficients with sufficient reliability. In a recent measurement of the angular coefficients in $pp \to Z \to \mu\mu$ at $\sqrt{s} = 8\,\text{TeV}$ by the ATLAS collaboration [106], the DYNNLO program - the program on which DYTURBO is based - gave a good description of the data. An exception to this was the quantity $A_0 - A_2$, although this has little effect on the muon p_T distribution at this level of precision. This publication also showed a poor description of several of the coefficients by POWHEGPYTHIA, including unphysical negative values of A_0 at low boson p_T. These conclusions are roughly reiterated by the very recent LHCb measurement of the angular coefficients [107], where it was also shown that the LHCb PYTHIA tune gives a very poor prediction.

All of the candidate models are evaluated for their use in predicting the unpolarized cross section. This evaluation proceeded by comparison to the fully-reconstructible p_T^Z distribution, as shown in Figure 5.1. On the top left of this figure, the normalized p_T^Z distributions from each of the candidate programs (the NNLO version of DYTURBO has been used) is shown with default settings, overlaid on the 2016 LHCb p_T^Z data. On the top right, the best fits to the data by tuning the available parameters is shown. For a fair comparison, in each case the angular coefficients come from DYTURBO at $\mathcal{O}(\alpha_s^2)$, and a modern PDF set is used. Since the PDFs used in the LHCb tune of PYTHIA is the older CT09MCS set, this is also shown. The pre-fit rapidity distributions from each of the models is also shown in the bottom panel of Figure 5.1.

Although DYTURBO gives the best pre-fit description, it predicts a p_T distribution that is too hard after floating the available tuning parameter g, which relates to the non-perturbative part of the prediction. LHCb's tune of PYTHIA gives a good description of the p_T distribution, although it disagrees with the Z rapidity distribution by up to 20%. The choice of PDF set for PYTHIA makes little difference to the p_T prediction, but is important for the rapidity distribution. This is to be expected as the rapidity distribution is mainly controlled by the PDFs (see e.g. Eqn. 1.3). POWHEGPYTHIA, POWHEGHERWIG and HERWIG - which are all fixed-order NLO predictions interfaced to parton showers - start off with a description that is far too soft, but improve greatly

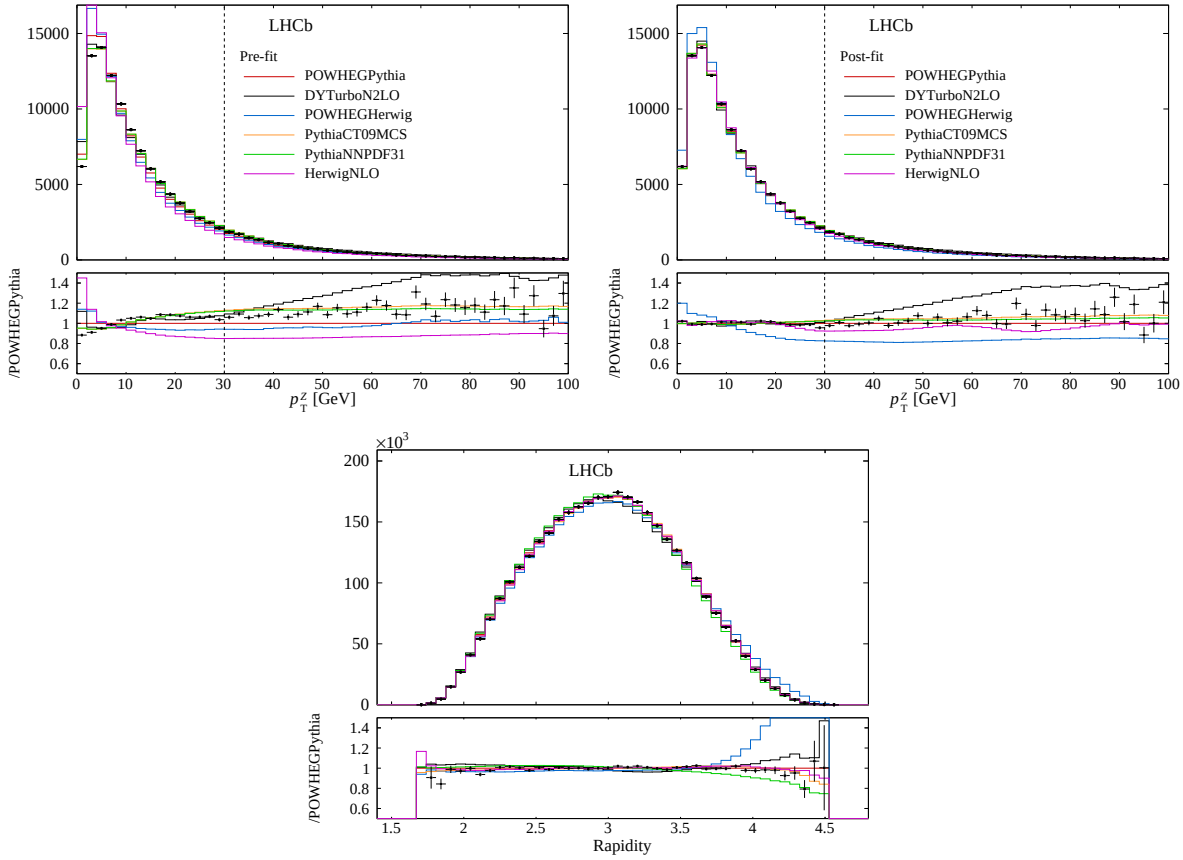


Figure 5.1: Predictions from the candidate models of the p_{T}^{Z} distribution before (top left) and after (top right) floating of their respective tuning parameters, and predictions of Z rapidity distribution using the default settings. In each case, the 2016 LHCb Z data is overlaid (black points).

with tuning of their α_s and $k_{\mathrm{T}}^{\mathrm{intr}}$ parameters. Of these, POWHEGPYTHIA gives the best description of the data, particularly at the lower p_{T}^{Z} values which dominate the dataset.

Based upon this, POWHEGPYTHIA was selected as the base model to reweight the templates to for both the m_W and $R_{\tau\mu}$ analyses. Since the rapidity distribution is largely PDF-dependent, and the mass distribution is a function of the vector boson mass and its width, this choice of generator mainly impacts the boson p_{T} distribution. The other models were retained for use in assessing the systematic uncertainty in this choice of p_{T} model, which is particularly important for the m_W measurement. Technically, the uncertainty in the p_{T} model is assessed by producing interpolation histograms using the base POWHEGPYTHIA 3D cross section, but in each bin of rapidity and mass allowing the *shape* of the p_{T} distribution to be predicted by another model. This hybrid approach leaves the rapidity and mass distributions unchanged from POWHEGPYTHIA.

5.2.1 Parametric correction at high boson transverse momentum

Despite POWHEGPYTHIA providing a good description of p_{T}^{Z} across the bulk of the distribution, Figure 5.1 shows an underestimation of the cross section that grows with p_{T}, reaching around 10% at 100 GeV. This can be corrected by multiplying the predicted cross section with a p_{T}-dependent error function of the form

$$f(p_{\mathrm{T}}) = 1 + A(1 + \mathrm{Erf}(B(p_{\mathrm{T}} - C))) \times (1 + Dp_{\mathrm{T}})\,. \tag{5.3}$$

After a rough tuning of the simulation in α_s and $k_{\mathrm{T}}^{\mathrm{intr}}$, the parameters A, B, C and D can be fitted to the ratio of the p_{T}^{Z} distribution in data and simulation. Since this underestimation only becomes well above the Jacobian peak, such a correction is not expected to lead to any noticeable bias on observables such as m_W.

5.3 QED FSR models

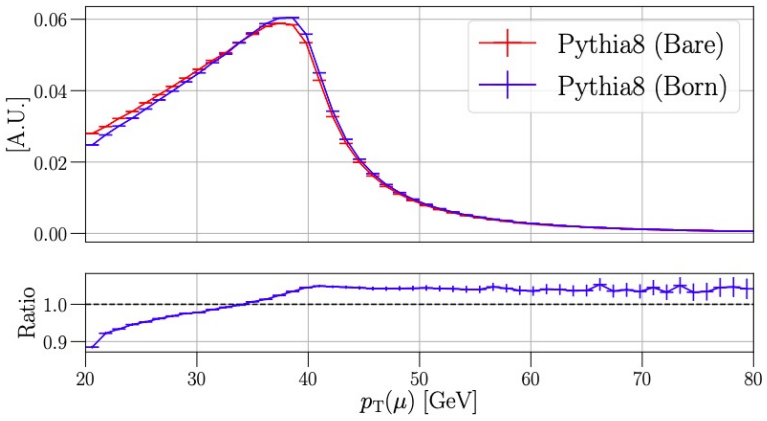


Figure 5.2: Comparison of the p_{T} spectrum of muons in W-boson decays before (Born) and after (Bare) final-state radiation. The ratio of Born to Bare is also shown.

The reweighting of the simulation due to differing angular coefficient and unpolarized cross section predictions is all carried out at the Born level, which was defined at the end of Section 2.3. Final-state QED radiation causes a loss of energy in the boson system, which shifts down the Bare-level muon p_{T} spectrum in $W \to \mu\nu$ with respect to the Born-level muon p_{T}, as shown in Figure 5.2. PYTHIA provides a model of FSR as part of its simulation, however members of the m_W analysis team evaluated the use of HERWIG, PHOTOS [108], DIRE [109] and VINCIA [110] for their usage in precision EW measurements. For Z production and decay, it was found that all five generators give a similar description

of FSR, both in terms of the relative energy loss of final-state leptons due to FSR, defined as

$$\frac{\Delta E}{E}(\ell\ell) = \frac{E^{\text{Born}}(\ell\ell) - E^{\text{Bare}}(\ell\ell)}{E^{\text{Born}}(\ell\ell)}\,, \tag{5.4}$$

and also in the subsequent effect on the dimuon mass distribution. However, for W decays Vincia and Dire gave drastically different descriptions of the relative energy loss spectrum compared to the other three, with differences at the percent level being propagated to the muon p_{T} distribution. Owing to this disagreement, and the fact that (in contrast to Pythia, Herwig and Photos) neither Vincia nor Dire have been extensively used in precision EW analyses, they were concluded to be unsuitable. The study's second conclusion was that no reason was found to treat any of Pythia, Herwig or Photos above the others: all three are treated equally in the following analyses. This is achieved by using a default QED FSR weight of $(1 + w_{\text{Photos}} + w_{\text{Herwig}})/3$, where the weights are taken from predictions of the relative energy loss distribution by Photos and Herwig, and the denominator of the weight is the Pythia prediction that is used by default in the LHCb simulation. The individual weights for each of Pythia, Photos and Herwig are retained to help assess the systematic uncertainty due to the model choice propagated to the parameter of interest.

5.4 Choice of PDFs

As mentioned in Section 2.4, there are a number of different PDF sets available from different PDF fitting groups. It was chosen here to use three modern PDF sets: NNPDF3.1 [33], CT18 [111] and MSHT20 [112]. These three are all "global" PDF sets, in that they are fitted to a wide variety of data including deep inelastic scattering, Drell-Yan and boson+jet processes; and were the contemporary updates of three of the PDF sets recommended for use by PDF4LHC in 2015 [41]. The data used to extract the three PDF set overlaps almost entirely, but fortunately none include the 2016 LHCb data used in this thesis. Given this, and that there is deemed to be no other reason to prefer one over the other, the three PDF sets are treated equally, except that in this case one PDF set must be chosen for the default model. The choice made for the analyses here is, arbitrarily, NNPDF3.1.

The PDF set used in the event generation step of the LHCb simulation is CT09MCS [113], which analogously to the event generator itself cannot be easily changed. However, in running the different pQCD models at generator level as described in Sections 5.1 and 5.2, a PDF choice can be made. All the pQCD predictions made above used the NNPDF3.1 PDF set, so reweighting to this more modern PDF has already been achieved. The effect of using a different PDF set (or a variation within one PDF set)

in the template preparation - whilst keeping the same predictions for the unpolarized cross section and angular coefficients - can be similarly mimicked via an additional weight. Interpolation histograms are made with the central and target PDF sets, with all other choices the same, from which the PDF weights

$$w_{\mathrm{PDF}} = \frac{x_1 f_1(x_1, Q^2) \cdot x_2 f_2(x_2, Q^2)|_{\mathrm{target}}}{x_1 f_1(x_1, Q^2) \cdot x_2 f_2(x_2, Q^2)|_{\mathrm{nominal}}}\,, \tag{5.5}$$

can be extracted. Here, $x_{1,2}$ and $f_{1,2}$ are the Bjorken-x and PDFs for partons 1 and 2 involved in the interaction.

CHAPTER 6

Detector modelling strategies for precision EW studies

As well as having a good understanding of the physics modelling aspects of Drell-Yan, there are also a plethora of potential detector effects that could cause intolerable biases on the physics observables. For example, since measuring m_W essentially entails precisely measuring the muon p_{T} distribution, any potential biases in LHCb's momentum calculation at high p_{T} are key. The topic of this chapter is the many detector modelling strategies that have been developed to control these possible biases.

Since the m_W and $R_{\tau\mu}$ measurements share essentially the same dataset of high-p_{T} muons collected with LHCb in 2016, most of the modelling procedures are applicable to both analyses. As with the preceding chapter, since the m_W measurement began first and had a greater number of collaborators, most of the strategies outlined in this chapter were developed first for that analysis by others, and were later incorporated into the workflow of the $R_{\tau\mu}$ analysis by the author of this thesis. It is therefore necessary to (relatively briefly) explain this work by others to fully understand the methods and results of later chapters. The treatment of the muon reconstruction efficiencies, which was wholly the author's work (under direction from the supervisor) is omitted here and described separately - in greater detail - in Chapter 7.

The layout of this chapter follows the execution workflow of the m_W and $R_{\tau\mu}$ analyses. The first section details the control samples used, which are identical for both measurements. Sections 6.2, 6.3 and 6.4 describe the large effort towards accurately modelling LHCb's momentum measurement. Section 6.5 then describes the correction of the simulation's impact parameter (IP) and track fit-quality resolutions. Sections 6.6 and 6.7 describe the method to correct for any mismodelling in the efficiencies of the key selection requirements. Section 6.8 outlines the dedicated simulation of the decay-in-flight

background from light hadrons. Important results of these methods when applied in the m_W and $R_{\tau\mu}$ measurements will be shown in Chapters 8 and 9.

6.1 Selection of control samples

The six control samples of data and simulation are as follows. First, there are three dimuon samples with positively ID-ed muons: $Z \to \mu\mu$, $\Upsilon(1S) \to \mu\mu$ and J/$\psi \to \mu\mu$, all of which have positively-identified muons (as defined in Section 3.5.2), referred to as the `Z`, `U1S` and `Jpsi` samples respectively. There are two additional $Z \to \mu\mu$ samples where only one positively ID-ed muon is required: `ZTrkEff` and `ZMuID` are used to study muon tracking efficiencies and muon trigger and identification efficiencies respectively. Finally, in order to train a parametric model for the decay-in-flight background (described in Section 6.8) the m_W analysis requires a sample of unidentified muons that otherwise mimic $W \to \mu\nu$ signal decays. This is known as the `NoMuID` sample. The single-muon signal selections differ slightly between the m_W and $R_{\tau\mu}$ analyses, and are described in the later chapters.

The three Z samples, the `U1S` sample and `Jpsi` sample in data were all collected in 2016. The `NoMuID` sample is from 2017-2018, since the trigger lines necessary to accumulate this sample were not present in 2016. The simulated control samples to compare to were simulated with identical conditions as their real-data counterparts. In all cases, the selections made by the trigger and stripping lines were superseded by selection cuts made at the analysis stage. Tables 6.1, 6.2, 6.3, 6.4 and 6.5 show the selection cuts imposed on each control sample at the analysis stage. A few of the selections require further explanation:

- The muon isolation variable $\mathcal{I}^{\mathrm{PF},0.4}$ is defined as the scalar sum of the transverse momenta of all charged and neutral particles - as selected by a particle-flow algorithm described in Ref. [114] - within a cone defined by $\Delta R = 0.4$ around the muon,
- The muon χ^2_{IP} (sometimes called impact parameter significance) is defined by the difference in the vertex fit quality of the primary vertex (PV) including and excluding the muon. A large IP correlates strongly with large χ^2_{IP},
- The muon χ^2_{trk} is the fit quality of the muon's track,
- "TOS" stands for "trigger-on-signal", meaning that the signal muon was responsible for firing the trigger lines. Unless stated otherwise, the trigger lines are as described in Section 3.5: `L0MuonEW`, `Hlt1SingleMuonHighPt` and `Hlt2SingleMuonHighPt`,
- `PIDmu` is a secondary muon identification variable, which takes likelihood ratios from the calorimeters and muon systems and computes the difference in log-likelihoods between the muon and pion hypotheses as $\log \mathcal{L}_\mu - \log \mathcal{L}_\pi$ [115].

Table 6.1: The `Z` and `ZMuID` selection requirements. These samples differ only in that the latter's stripping line did not require that both muons were positively-identified.

Description	Requirement
Dimuon mass	$77\,\text{GeV} < m_{\ell\ell} < 105\,\text{GeV}$
Minimum p_T	$p_\text{T} > 20\,\text{GeV}$ (both muons)
Pile-up	Number of candidates in event = 1
Momentum error	$\Delta p/p < 0.06$ (both muons)
Sanity cut	$p < 2\,\text{TeV}$ (both muons)
Trigger	μ^+ or μ^- is TOS at L0, HLT1 and HLT2
η acceptance	$1.7 < \eta < 5.0$ (both muons)
Isolation	$\mathcal{I}^{\text{PF},0.4} < 10\,\text{GeV}$ (both muons)
Track quality	$\chi^2_\text{trk} < 1.8$ (both muons)
IP significance	$\chi^2_\text{IP} < 100$ (both muons)

Table 6.2: The `ZTrkEff` selection requirements.

Description	Requirement
Pile-up	Number of candidates in event = 1
Sanity cut	$p < 2\,\text{TeV}$ (both muons)
$Z \to \mu\mu$ combination quality	$\chi^2/\text{ndf} < 10$

Table 6.3: The `Jpsi` selection cuts. The J/ψ mass used is 3.097 GeV [88].

Description	Requirement
Dimuon mass	$\lvert m_{\ell\ell} - m_{\text{J}/\psi}\rvert < 20\,\text{MeV}$
Minimum p_T	$p_\text{T} > 3\,\text{GeV}$ (both muons)
Pile-up	Number of candidates in event = 1
Momentum error	$\Delta p/p < 0.06$ (both muons)
Sanity cut	$p < 2\,\text{TeV}$ (both muons)
Hits in the TT	At least one TT cluster (both muons)
Trigger	μ^+ or μ^- is TOS on the `L0Muon` line
Track quality	$\chi^2_\text{trk} < 1.8$ (both muons)
Muon hypothesis	`PIDmu` > 2 (both muons)

Table 6.4: The `U1S` selection cuts. The $\Upsilon(1S)$ mass used is 9.46 GeV [88].

Description	Requirement
Dimuon mass	$\lvert m_{\ell\ell} - m_{\Upsilon(1S)}\rvert < 50\,\text{MeV}$
Minimum p_T	$p_\text{T} > 3\,\text{GeV}$ (both muons)
Pile-up	Number of candidates in event = 1
Momentum error	$\Delta p/p < 0.06$ (both muons)
Sanity cut	$p < 2\,\text{TeV}$ (both muons)
Trigger	μ^+ or μ^- is TOS on the `L0Muon` line
Track quality	$\chi^2_\text{trk} < 1.8$ (both muons)
Muon hypothesis	`PIDmu` > 2 (both muons)

Table 6.5: The `NoMuID` selection cuts.

Description	Requirement
Pile-up	Number of candidates in event = 1
Sanity cut	$p < 2\,\text{TeV}$
Hits in the TT	At least one TT cluster
Track quality	$\chi^2_\text{trk} < 1.8$
Muon hypothesis	Muon is *not* positively-identified

During the processing of the datasets for the m_W and $R_{\tau\mu}$ analyses, each event is allowed to fail up to one of the relevant selection requirements. This means that at any point, plots of the cut variables can be made and the region outside of the cut requirement can still be seen, which is very useful for assessing that the cuts and any corrections made to the cut-variable distributions are well-understood. However, before any computation or fitting etc. is done using the datasets, the cuts in these tables are re-applied in full - the "all-but-one-cut" strategy is purely used for monitoring purposes.

6.2 Fill-dependent momentum scale corrections

To understand the *scale* of the LHCb momentum measurement, the masses of "standard candle" resonances such as J/ψ, $\Upsilon(1S)$ and Z are often measured and compared to their known values. In 2016, it was observed that the measurement of these standard-candle masses by LHCb drifted with time, and so the first analysis step in calibrating the momentum scale was to remove this time-dependent bias. The selected $\Upsilon(1S)$ data was split by fill number (a proxy for time) into 20 equally-populated subsets, and for each subset the dilepton mass spectrum was fitted at the $\Upsilon(1S)$ peak with a sum of two Crystal-Ball [116] functions, both sharing the α_{CB}, n_{CB} and mean (m_{CB}) parameters, but with independent σ parameters. Combinatorial background was described by an exponential function with floating slope and normalization. From these fits a fill-dependent correction to the momentum scale can be derived as $\alpha = M_{\text{PDG}}/m_{\text{CB}}$, where $M_{\text{PDG}} = 9460.30 \pm 0.26\,\text{MeV}$ is the 2020 PDG value of the $\Upsilon(1S)$ mass [88]. The dependence of the momentum scale on this PDG value is later removed by fitting this peak again after offline detector re-alignment, as described in Section 6.4. The fill-dependent momentum scale corrections derived in the m_W analysis can be seen in Section 8.5, and were generally of $\mathcal{O}(10^{-4})$, although they had little effect when integrated in time over the whole dataset.

6.3 Detector alignment

Misalignments in the tracking system are particularly important for the momentum measurement of high-p_{T} muons, because such high-p_{T} tracks are not largely deflected by LHCb's magnetic field. The deflection of a track in LHCb in metres is approximately $3\ \text{GeV}/p$, meaning that a typical muon track of several hundred GeV will only be deflected by a few tens of millimetres. This argument can be reverse-engineered: in Ref. [117] it was found that a bias of $\mathcal{O}(50)\,\text{MeV}$ would be caused by just a $5\,\mu\text{m}$ shift in of the first T station along the bending (x) axis. Such translations thus contribute *curvature biases* to the momentum measurement, where a curvature bias δ is defined by

$$\frac{q}{p} \to \frac{q}{p} + \delta\,. \tag{6.1}$$

As briefly introduced in Section 3.5, LHCb has an online alignment procedure that aligns all measurement stations from constraints given by standard-candle masses. This alignment is designed with the broader LHCb physics programme in mind, which generally features tracks of lower p_{T} and standard candles at lower mass than is sufficient here. Therefore, at the offline data processing stage, an "offline" alignment was performed using a Z mass constraint as detailed in Ref. [118], and the tracks re-fitted with this improved alignment. Although this does greatly improve the Z mass resolution [87], significant curvature biases were seen to remain, as evidenced by substantial disagreements between m_W extraction from the data taken with the two different magnet polarities (in which curvature biases have the opposite effect).

The curvature bias problem was solved by the development of the *pseudomass* method [117]. The difficulty of correcting for curvature biases using the standard-candle masses is that the parent mass is typically calculated by reconstructing two child tracks of opposite charge - e.g. the Z mass would be found by reconstructing and combining two oppositely-charged muon tracks. Any curvature bias due to e.g. a station misalignment would have opposite effects on the two tracks (to first order) and largely cancel out in the mass measurement. The pseudomass gets around this problem by using the full momentum information from one track, but only the direction of the second track (as determined before the magnet kick). It is defined as

$$M^{\pm} = \sqrt{2p^{\pm}p_T^{\pm}\frac{p^{\mp}}{p_T^{\mp}}(1-\cos\theta)}, \tag{6.2}$$

where $p^{\pm}$ is the momentum of the $\mu^{\pm}$ and θ is the opening angle of the dimuon system. Since this does not require information of the second muon's momentum (it cancels in the $p^{\mp}/p_{\mathrm{T}}^{\mp}$ ratio), it is a direct probe of curvature bias affects on only the $\mu^{\pm}$. The pseudomass would be equal to the true mass if the Z candidate has zero p_{T} with respect to the axis bisecting the two lepton decays, and would therefore give a peak at m_Z that can be used to constrain away curvature biases. Fortunately, most Z events at LHCb do have $p_{\mathrm{T}}^Z < m_Z$, and the smallest p_{T}^Z values can be selected independently of the muon momenta by cutting on the ϕ^* variable [119], defined as:

$$\phi^* \equiv \tan\phi_{\mathrm{acop}}\sin\theta_\eta^* \approx \frac{p_T^Z}{m_Z}, \tag{6.3}$$

where $\phi_{\mathrm{acop}} = (\pi - \Delta\phi)/2$ is the "acoplanarity" angle, $\Delta\phi$ is the azimuthal opening angle between the two leptons, $\theta_\eta^* \equiv \arccos\left(\tanh\left(\frac{\eta^- - \eta^+}{2}\right)\right)$, and η^- and η^+ are the pseudorapidities of the negatively and positively charged leptons respectively. For the

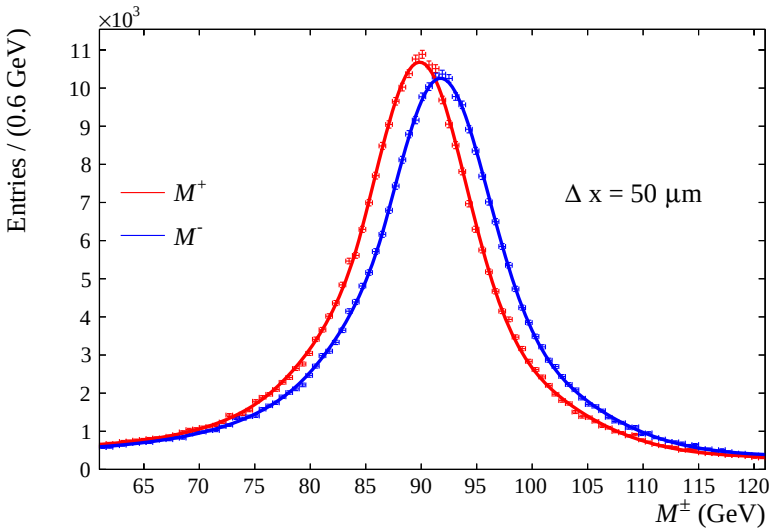


Figure 6.1: M^+ and M^- pseudomass distributions in simulated events, assuming the LHCb detector resolution, overlaid with a parametric fit function fitted to each. An artificial shift of 50 μm has been applied at the measurement plane. Taken from Ref. [117].

purpose of deriving the pseudomass corrections, Z events were selected with $\phi^* < 0.05$.

If curvature biases are present via translation of the measurement stations, it causes the two pseudomass peaks from muons of opposite charge to separate. This is illustrated in Figure 6.1, where an artificial 50 μm shift has been introduced at the measurement plane, yielding a $\sim 2\,\text{GeV}$ separation between the peaks. By fitting the pseudomass peaks in real data and simulation, this process can be reverse-engineered: observed shifts between pseudomass peaks can be translated back to shifts in the measurement plane, which can then be corrected for, and the observed biases are thus removed. The fitting function used is a sum of a Crystal-Ball [116] function describing the peak, plus an exponential component. The means of the $M^\pm$ Crystal-Ball functions are written in terms of an average mass $\overline{M}$ and a pseudomass asymmetry A as $\overline{M}(1 \pm A)$, to which the charge-dependent curvature bias can be related as

$$\delta \approx A \frac{< \frac{1}{p^+} > + < \frac{1}{p^-} >}{2} . \tag{6.4}$$

Since this equation is only approximately true, the procedure must be applied iteratively: the peaks are fitted to yield an asymmetry; this asymmetry is corrected for by "realigning" the measurement plane; the tracks are re-fitted; the pseudomass peaks are fitted etc. Curvature bias corrections are applied additively from iteration-to-iteration, and the iterations stop when the residual corrections are zero within their statistical uncertainties for most bins.

This procedure decouples information from the two muons, giving freedom to bin

the fits (and hence the corrections) in terms of muon η, ϕ etc. The fits were also separated into regions where the muon in question falls in the OT or IT station, as these could deform/shift independently of one another.

The results of this method applied in the m_W analysis are shown in Section 8.5, where curvature bias corrections of $\mathcal{O}(10^{-4}\,\mathrm{GeV}^{-1})$ were found.

6.4 Momentum scale calibration

The previous section dealt with improving the momentum resolution in data by removing curvature bias effects. The next step is to make the simulation's p_{T} measurement more representative of data by applying a momentum smearing. This transforms the momentum of simulated tracks to account for four effects that may differ between data and simulation:

1. **Residual curvature bias**: $\frac{q}{p} \to \frac{q}{p} + \delta$ that could be leftover from the pseudomass corrections,
2. **Momentum scale bias**: $p \to (1+\alpha)p$,
3. **Mismodelling of multiple scattering**: multiple scattering leads to a smearing of the momentum measurement of the form $p \to p \cdot \mathcal{N}(1, \sigma_{\mathrm{MS}})$, where $\mathcal{N}(1, \sigma_{\mathrm{MS}})$ is a Gaussian-distributed random number of unit mean and width σ_{MS}. σ_{MS} is scaled up by a factor 1.5 for muons in $\eta > 3.3$ to account for the extra material seen towards the outer edges of the detector up to the magnet [120].
4. **Momentum-dependent curvature resolution**: $p \to p{\cdot}\mathcal{N}(1, p{\cdot}\sigma_\delta/\cosh\eta)$, where the $1/\cosh\eta$ scaling improves the agreement in the fit.

In total, this gives a smearing function of the form:

$$\frac{q}{p} \to \frac{q}{p \cdot \mathcal{N}(1+\alpha, \sigma_{\mathrm{MS}})} + \mathcal{N}\left(\delta, \frac{\sigma_\delta}{\cosh\eta}\right), \tag{6.5}$$

the parameters of which can be fitted to the Z, $\Upsilon(1S)$ and J/ψ invariant mass peaks. These fits are binned in magnet polarity and categorized by η of the two muons (the categories are delineated as shown in Table 6.6). Categories 1, 2 and 3 are not used for the quarkonia samples ($\Upsilon(1S)$ and J/ψ); only Z is used there since this η region does not overlap with the `W` signal selection, but a wider selection is used for the Z portion of the m_W fit. In categories 4, 5 and 6 the Z sample is further split into three bins of muon momentum asymmetry, defined as

$$A_P = \frac{|\mathbf{p}^+| - |\mathbf{p}^-|}{|\mathbf{p}^+| + |\mathbf{p}^-|}, \tag{6.6}$$

Table 6.6: Categorization of the dimuon mass distributions in bins of η used in the fit.

	$1.8 < \eta^+ < 2.2$	$2.2 < \eta^+ < 3.3$	$\eta^+ > 3.3$
$1.8 < \eta^- < 2.2$	1	2	3
$2.2 < \eta^- < 3.3$	2	4	5
$\eta^- > 3.3$	3	5	6

where $\mathbf{p}^+$ ($\mathbf{p}^-$) is the 3-momentum of the μ^+ (μ^-). A curvature bias would cancel to first-order in an event with zero momentum asymmetry, but have a much more noticeable effect at higher momentum asymmetries, so binning in this variable helps to determine σ_δ. The fit is conducted simultaneously, with 36 total fit categories across the three control samples. There are six fit parameters in total: two momentum scale (α) parameters covering $2.2 < \eta < 4.4$ and $\eta < 2.2$; two resolution σ_δ parameters, also for $2.2 < \eta < 4.4$ and $\eta < 2.2$; a single curvature bias δ for $2.2 < \eta < 4.4$, which is fixed to zero for $\eta < 2.2$; and a single multiple-scattering width σ_{MS} was found to be sufficient to cover all η values (subject to the fixed scaling between the two η regions described above).

Final-state radiation (FSR) makes a significant effect on the shape of these resonance peaks used to tune the momentum scale. The Z events in the momentum smearing fitter are weighted according to the average of FSR predictions made by Pythia, Herwig and Photos, as motivated in Section 5.3. For the quarkonia samples, where effect is expected to be less pronounced, the default Pythia FSR modelling is used (i.e. no FSR reweighting).

The results of the application of the momentum scale calibration in the m_W and $R_{\tau\mu}$ analyses can be seen in Sections 8.5 and 9.6 respectively.

6.5 Smearing of impact parameter and track fit quality

The IP variable is used (via the χ^2_{IP}) to discriminate between muons from semileptonic heavy-flavour decays, $W \to \tau\nu$ decays and "prompt" muons from $W \to \mu\nu$ and $Z \to \mu\mu$ decays. Tau leptons and heavy-flavour hadrons have a non-zero IP on average, while the true muon IP for $W \to \mu\nu$ and $Z \to \mu\mu$ should be zero, but the IP resolution of LHCb broadens this into a distribution. This resolution must be sufficiently well-modelled in the simulation to give accurate efficiencies if cutting on IP-related variables (in the case of m_W), or to be able to fit the IP distribution ($R_{\tau\mu}$). Out-of-the-box, the simulation of the $Z \to \mu\mu$ muon IP resolution requires further smearing to match the resolution observed in data. A Gaussian smearing model was applied to the x and y components

of the IP, IP_x and IP_y, the parameters of which were fitted to the $Z \to \mu\mu$ data in bins of muon η and ϕ. The components are used because they are roughly Gaussian quantities, whereas the IP - being the length of a 2D $(\mathrm{IP}_x, \mathrm{IP}_y)$ vector - is not, and the z-component of the IP is negligible due to LHCb's forward geometry. This smearing broadened the distributions according to the Gaussian's width σ and shifted their mean values by the Gaussian's mean μ. Fitting was achieved (in each bin independently) by minimization of an unbinned Anderson-Darling statistic [121] with respect to σ and μ. The Anderson-Darling statistic is proportional to the cumulative "distance" between two samples, calculated as the integral (over some variable x - here IP_x and IP_y) of the squared difference in the cumulative distribution functions of the two samples. The integral is also weighted by a factor that grows towards the integral limits, therefore giving more weight to the tails in the statistic and hence more emphasis in fitting them correctly in a minimization. No charge dependence is expected, so muons of both charges are aggregated together in the fits. The best-fitting values of the smearing parameters are then used to smear IP_x and IP_y, from which the IP and χ^2_{IP} are re-calculated. The procedure is very similar for the χ^2_{trk} distribution: here the χ^2_{trk} distribution itself is smeared via the Gaussian model, which on average broadens and shifts the simulation's χ^2_{trk} down to lower values, better approximating the χ^2_{trk} distribution as seen in $Z \to \mu\mu$ data.

6.6 Isolation efficiency

The m_W analysis signal selection makes a cut on the muon isolation to reject heavy flavour and decay-in-flight backgrounds. This cut has the largest rejection of any applied to the $W \to \mu\nu$ signal data sample, so particular care is needed in the modelling of that cut's efficiency in the simulation. The efficiency is evaluated in data and simulation with the $Z \to \mu\mu$ control sample, which passes all of the `Z` selection cuts except the isolation requirement, exploiting the "all-but-one" cut strategy. This gives a sample of W-like muons with negligible residual background that can be used to determine the isolation efficiency. This efficiency is then evaluated by simply counting the number of muons that pass the isolation cut (the numerator of the efficiency) and dividing by the total number of muons. It is expected that the isolation efficiency would vary across the detector (particularly near the edges, where the isolation cone won't be fully instrumented), and that it would also depend on the overlap between the isolation cone and the system that the Z recoils against. Naively, the isolation cut is expected to become less effective at vetoing backgrounds if this overlap is significant. This motivated binning the sample in muon pseudorapidity η and the recoil projection variable u, defined as

$$u = \frac{\mathbf{p}_{\mathrm{T}}^{V} \cdot \mathbf{p}_{\mathrm{T}}^{\mu}}{p_{\mathrm{T}}^{\mu}}\,, \tag{6.7}$$

where $\mathbf{p}_{\mathrm{T}}^{\mu}$ and $\mathbf{p}_{\mathrm{T}}^{V}$ are the transverse momentum vectors of the muon and the parent vector boson. By momentum conservation, the recoiling hadronic system will have equal-and-opposite p_{T} to the parent vector boson, so the expectation is that the isolation cut has a lower efficiency at large negative values of u. This variable can also account for how differences between the vector boson production between W and Z affect the isolation efficiency, as it is related to the boson p_{T}.

The 2D efficiency results (binned in reconstructed u and η) for data and simulation are divided bin-by-bin to produce binned efficiency scale factors $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$, which is the relevant quantity for correcting the simulation via weights. The binned scale factors were then smoothed to reduce any noise due to limited control sample size. The application of this method to the m_W analysis can be seen in Section 8.5, giving scale factors generally at the per-cent level, rising to $\mathcal{O}(10\%)$ at very large-and-negative u.

6.7 Impact parameter significance and track quality cut efficiencies

The simple Gaussian smearing model detailed in Section 6.5 does not perfectly reproduce the χ^2_{IP} and χ^2_{trk} distributions in the data, which means the efficiency of cutting on the simulated distribution may be slightly biased. This is corrected using the same tool developed for the isolation efficiencies described in the previous section. Efficiency scale factors $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ for requirements on χ^2_{IP} and χ^2_{trk} were evaluated to be used as weights, binned in muon η and ϕ. In the end, these weights made a negligible impact to the m_W analysis (quantified in Section 8.5) and are not currently deemed necessary to be used in the $R_{\tau\mu}$ analysis.

6.8 Parametric modelling of the decay-in-flight hadron background

A troublesome background present in the single, high-p_{T} muon dataset is that coming from light hadrons produced in the primary interaction and decaying in-flight to muons before the muon stations (denoted "DIF"). At LHCb, since these tracks are boosted into the forward direction, there are many events where any kink in the track from the decay-in-flight cannot be resolved - giving a signature very similar to a muon coming from an EW process. Requiring the muon is isolated does reduce the size of this background,

but not to the level where it can be considered negligible. Accurate templates of this significant background must therefore be prepared.

Early attempts to create these templates focused on studying an inclusive sample of pp collisions in LHCb with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, simulated with PYTHIA at the generator level. This sample is dominated by charged hadrons in the final state, but requiring that they passed the loose, signal-like `NoMuID` selection gave only a 0.1% retention, and further selections necessary to mimic signal on quantities such as isolation or IP significance had efficiencies of $\mathcal{O}(10\%)$. To use these events for a template, the full LHCb detector simulation would be required, and generating enough simulated events with such low retention rates was undesirable and potentially unnecessary for m_W. The generator-level simulation is not so prohibitively resource-intensive, and so a second attempt was made to smear the generator-level simulated events with the known p_{T} and IP resolutions of LHCb, as given in Refs. [58, 122]. Unfortunately, the resulting candidates showed a markedly different p_{T} distribution to data and background candidates selected with the `NoMuID` selection, which was difficult to motivate and understand sufficiently.

A third-and-final data-driven approach was adopted based upon muon candidates in data selected with the `NoMuID` selection cuts. Activity in the hadronic calorimeter, and a negative `PIDmu` requirement both ensure that there is no signal left in this sample, and an IP significance cut reduces heavy-flavour contamination to negligible levels. What remains is a pure sample of light hadrons, as verified by the population of the aforementioned generator-level inclusive sample. The probability for an unstable hadron to decay in-flight within a detector of length d is

$$1 - \exp\left(-\frac{md}{\tau p}\right) \approx \frac{md}{\tau p}\,, \tag{6.8}$$

where m, τ and p are the particle's mass, lifetime and momentum respectively. Since the abundances of different particle species in the sample is unknown, and in the end the yield of this component will float freely, only the $1/p$ dependence of these weights is important. Therefore, weights $w_{\mathrm{DIF}} = 1/p$ are applied to the `NoMuID` control sample, which morphs the momentum dependence to mimic decays in-flight. To remove any statistical uncertainty that may be propagated due to limited control sample size, a parametric model was then trained on the weighted `NoMuID` sample, from which any number of template events can be generated. The parametrization used is the so-called "Hagedorn" function:

$$f(p_{\mathrm{T}}) \propto \frac{1}{(1 + \frac{p_{\mathrm{T}}}{p_{\mathrm{T},0}})^n}\,, \tag{6.9}$$

which has been shown to give a good description of inclusive p_{T} spectra in hadronic collisions from [123]. It is also observed to give a good fit to the `NoMuID` control sample -

as can be seen when applied in the m_W analysis at the end of Section 8.5 - for hadrons of both charges that satisfy requirements on χ^2_{IP} and isolation cuts such as those applied to a $W \to \mu\nu$ signal sample. The Hagedorn function is therefore fitted to the weighted `NoMuID` data in bins of isolation and χ^2_{IP}, and the parametrization from the lowest isolation and lowest χ^2_{IP} bin is taken as the model from which template events are generated, as this is closest to the background present in the signal selection of the data. The fitting-and-generating is done separately for the two hadron charges, as they are observed to have different p, p_{T} and η shapes, as well as an overall 10% asymmetry towards positive charges. The resulting templates were used directly in a fit to the muon p_{T} distribution, although the modelling of any other variables was based upon educated guesses and no dedicated training.

CHAPTER 7

Studies of high-p_{T} muon reconstruction efficiencies

The foremost selection requirement for high-p_{T} muons for precision EW studies is that they fired the relevant trigger lines at L0, HLT1 and HLT2. Implicitly, this also means that the high-p_{T} track was reconstructed by the LHCb tracking system, and that the track was positively identified as a muon, i.e. that it could be matched to hits in all four (at these momenta) of the muon stations downstream of the LHCb calorimeters. All of these requirements come with a corresponding efficiency, which can in principle vary greatly with factors such as the detector region or the momentum of the track. It is crucial that these efficiencies, and how they vary, is accurately modelled in the LHCb simulation, particularly those variations that correlate with the physics observables. If this is not the case, there is a danger of biasing the extraction.

This chapter details the treatment of the muon reconstruction efficiencies used in the m_W and $R_{\tau\mu}$ measurements with the 2016 LHCb data. The strategy was developed for the m_W analysis, and is in use with minimal modification for the $R_{\tau\mu}$ analysis. The first section introduces the commonly-used tag-and-probe method of measuring efficiencies, which is the basis of the treatment. From there, the chapter focuses in on the application to precision EW analyses, which is wholly the author's work under direction of the supervisor. This begins with definition of the relevant efficiencies and the tag-and-probe sample preparation in Sections 7.2 and 7.3, followed by the extraction (and parametrization) of binned efficiencies in Section 7.4. Section 7.5 then details the construction of event weights to correct for the observed data/simulation mismodelling in the efficiencies. Section 7.6 describes variations to the efficiency treatment from which systematic uncertainties can be evaluated. Finally, a closure test of the method and a conclusion is given in Sections 7.7 and 7.8.

The key results of the treatment when applied to both the m_W and $R_{\tau\mu}$ measurements are shown in their respective chapters. However, figures are necessary in this

chapter to properly explain the method and elucidate the conclusions. Figures here are therefore taken from the finalized m_W measurement (as published in Ref [124]) where necessary, but do not feature again in later chapters or in the published paper.

7.1 The tag-and-probe method

In these analyses, muon reconstruction efficiencies are calculated in data and simulation using the tag-and-probe method. This method requires a control sample of decays that have two particles in the final state which can be treated as identical with respect to the efficiencies under study. LHCb's muon reconstruction efficiencies are expected to be independent of muon charge, so dimuon resonance decays are typically used. For each event of the control sample, a strict set of requirements is placed on one muon (the *tag*), which "tags" the decay and suppresses background contributions. The other muon in the decay - the *probe* - is then free to pass or fail an identification requirement, which allows the calculation of the efficiency of said identification requirement as

$$\varepsilon = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}\,, \tag{7.1}$$

where N_{pass} and N_{fail} are the number of probes that pass or fail the identification requirement respectively. Typically, all muons of one charge are assigned to be the tags and the other to be the probes, N_{pass} and N_{fail} are calculated for this assignment, and then the process is repeated with the other assignment of charges. The charge-summed N_{pass} and N_{fail} would then be used to work out the efficiencies. The identification requirement is generally some requirement on a correctly-reconstructed muon. An example of such a requirement is that the muon is positively identified as a muon (hereafter "has MuonID"). In this case, a control sample of dimuon decays is needed where at least one muon has MuonID, but the other could just be a long track[1] with a muon mass hypothesis. The tag is required to have MuonID, and then how many probes do and do not have MuonID can be counted. This is illustrated in Figure 7.1.

For this method to work, the pass and fail samples must be free of background contamination (which is typically present at higher concentration in the fail category, biasing the efficiency low). Taking the efficiencies from this method at face value would mean an assumption that the tag requirements are completely unbiasing, which is a strong assumption, but fortunately the primary quantity of interest here is the ratio of efficiency in data to simulation, $\varepsilon_{\text{D}}/\varepsilon_{\text{MC}}$, where any potential bias should largely cancel.

[1] A "long" track was defined earlier to be a track with hits in all of LHCb's tracking stations.

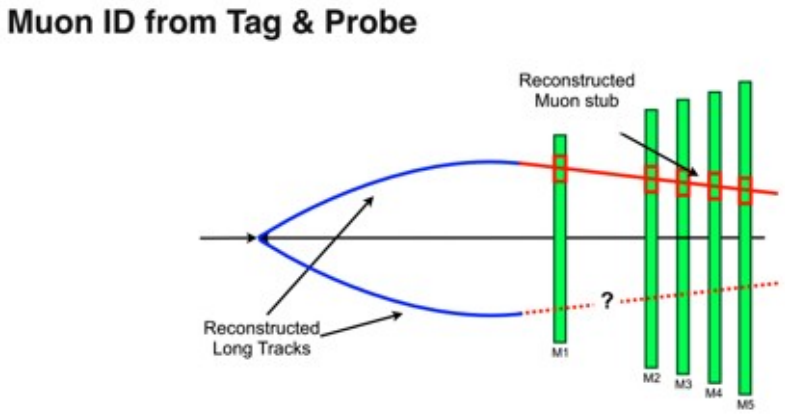


Figure 7.1: Schematic of the tag-and-probe method for measuring the MuonID efficiency in LHCb.

7.2 Efficiency definitions

Ultimately, the efficiency of interest here corresponds to a total requirement that signal muons were reconstructed, have MuonID and triggered the relevant lines. Unfortunately, there is no single control sample available to calculate this total efficiency, and so it must be factorized. Firstly, the tracking efficiency can be separated out, as it can be calculated from the dedicated `ZTrkEff` sample (defined in Table 6.2), which was the only dimuon sample available that doesn't require two properly-reconstructed long tracks. The other specialist Z control sample, `ZMuID` (see Table 6.1) requires that at least one muon has MuonID and has fired the triggers, so can be used for the remaining efficiency. The L0 trigger had a p_{T} threshold of 6 GeV, but as described in Section 3.5 the L0 muon momentum resolution is poor, and so it is desirable to also use lower-p_{T} muons from $\Upsilon(1S) \to \mu\mu$ to study the L0 efficiency turn-on. However, the $\Upsilon(1S) \to \mu\mu$ control sample - `U1S`, see Table 6.4 - is formed of muons that all have MuonID, so cannot be used to calculate a MuonID efficiency. Therefore, the remaining efficiency must be split up into trigger ($\varepsilon_{\mathrm{Trig}}$) and MuonID ($\varepsilon_{\mathrm{ID}}$) efficiencies. This can be factorized as either $\varepsilon_{\mathrm{Trig|ID}} \times \varepsilon_{\mathrm{ID}}$, or $\varepsilon_{\mathrm{ID|Trig}} \times \varepsilon_{\mathrm{Trig}}$. Preliminary studies showed that the trigger efficiency had the largest inefficiency and the worst mismodelling in simulation of all the requirements here. The former factorization was therefore chosen, because the absolute $\varepsilon_{\mathrm{Trig|ID}}$ is larger than $\varepsilon_{\mathrm{Trig}}$, meaning the former will suffer less from statistical uncertainties in the efficiency numerator. In summary, the total efficiency is factorized as

$$\varepsilon = \varepsilon_{\mathrm{Trig|ID}} \times \varepsilon_{\mathrm{ID}} \times \varepsilon_{\mathrm{trk}} \,. \tag{7.2}$$

The pass requirements for each of the three efficiencies are listed in Table 7.1.

The trigger lines are as described in Section 3.5: `L0MuonEW`, `Hlt1SingleMuonHighPt` and `Hlt2SingleMuonHighPt`. A HLT2 requirement is omitted because the inefficiency of HLT2 with respect to HLT1 on the signal (W) muons was found to be $< 10^{-6}$, and so can be ignored. The tracking efficiency is calculated by asking how often "MuonTT stubs" (a track reconstructed from only hits in the muon and TT stations - no initial requirement made on the presence of VELO clusters) can be matched to fully-reconstructed long tracks. This method assumes that the long-track reconstruction efficiency is mostly driven by the VELO [125], but any bias in this should cancel in the ratio of efficiencies $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$. The matching procedure of a MuonTT probe to a long track is as follows:

- The fraction of hits shared between the probe and the long track in the muon stations is $> 40\%$,
- The fraction of hits shared between the probe and the long track in the TT station is $> 60\%$,
- The combined mass of the long track and the tag $M > 40$ GeV.

Table 7.1: Pass requirements for the three muon reconstruction efficiencies under study. "TOS" stands for "trigger-on-signal", meaning that the probe was the part of the event responsible for firing the trigger lines.

Efficiency	Pass Requirement
$\varepsilon_{\mathrm{Trig\|ID}}$	Probe is TOS at L0 & HLT1
$\varepsilon_{\mathrm{ID}}$	Muon probe track has MuonID
$\varepsilon_{\mathrm{trk}}$	MuonTT stub matches to a long track

7.3 Efficiency sample preparation

In addition to the cuts already applied to the `ZTrkEff`, `ZMuID` and `U1S` samples, further requirements are applied on the tag, probe and the whole event to increase the sample purities, and to ensure the factorization as expressed by 7.2. These requirements are listed in Tables 7.2, 7.3, and 7.4 (7.5) for the tracking, MuonID, and trigger $Z \to \mu\mu$ ($\Upsilon(1S) \to \mu\mu$) efficiency samples respectively. The isolation variable $\mathcal{I}^{\mathrm{PF},0.4}$ is defined in Section 6.1. This cannot be used for the tracking probes, as it may count the muon's long track towards the sum. A "hollow" isolation variable is therefore constructed, $\mathcal{I}^{\mathrm{hollow}}$, which is the difference of isolation variables with cones of $\Delta R < 0.5$ and $\Delta R < 0.05$. The HLT2 denominator requirement is not applied to the $\Upsilon(1S)$ sample as these decays are not high-enough momentum for this line, but since this requirement is negligible on signal, this inconsistency should also be negligible.

Table 7.2: Selection requirements on the tag-and-probe sample used to determine tracking efficiencies.

Tag	Probe	Dimuon Candidate
Long track	MuonTT track	
$p_{\mathrm{T}} > 20$ GeV	$p_{\mathrm{T}} > 15$ GeV	$\lvert m_{\mu\mu} - m_Z \rvert < 25$ GeV
$\mathcal{I}^{\mathrm{PF},0.4} < 4$ GeV	$\mathcal{I}^{\mathrm{hollow}} < 4$ GeV	
Has MuonID		
TOS at L0, HLT1 & HLT2		

Table 7.3: Selection requirements on the tag-and-probe sample used to determine ID efficiencies.

Tag	Probe	Dimuon Candidate
Long track	Long track	
$p_{\mathrm{T}} > 30$ GeV	$p_{\mathrm{T}} > 15$ GeV	$\lvert m_{\mu\mu} - m_Z \rvert < 15$ GeV
$\mathcal{I}^{\mathrm{PF},0.4} < 6$ GeV	$\mathcal{I}^{\mathrm{PF},0.4} < 6$ GeV	
Has MuonID		
TOS at L0, HLT1 & HLT2		

Table 7.4: Selection requirements on the $Z \to \mu\mu$ tag-and-probe sample used to determine trigger efficiencies.

Tag	Probe	Dimuon Candidate
Long track	Long track	
$p_{\mathrm{T}} > 20$ GeV	$p_{\mathrm{T}} > 15$ GeV	$\lvert m_{\mu\mu} - m_Z \rvert < 15$ GeV
Has MuonID	Has MuonID	
TOS at L0, HLT1 & HLT2		

Table 7.5: Selection requirements on the $\Upsilon(1S) \to \mu\mu$ tag-and-probe sample used to determine trigger efficiencies. The `Hlt1Phys` line is used as a proxy for the `Hlt1SingleMuonHighPT` line, and `L0Muon` is used over `L0MuonEW` for the tag for its lower p_{T} threshold and therefore a larger sample size.

Tag	Probe	Dimuon Candidate
Long track	Long track	
		$9.0 \leq m_{\mu\mu} \leq 9.8$ GeV
Has MuonID	Has MuonID	
TOS on `L0Muon` & `Hlt1Phys`		

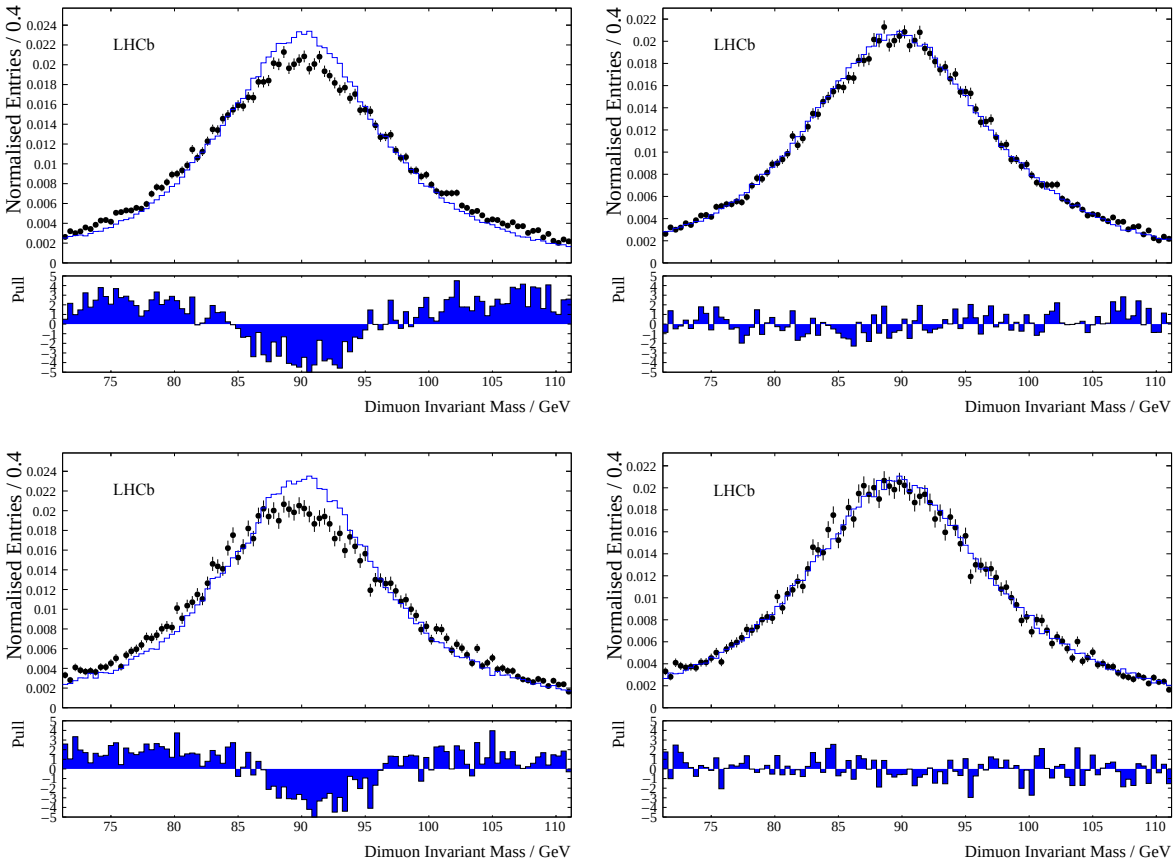


Figure 7.2: Effect of the additional momentum smearing on MuonTT probes in simulation. The left panels show the dimuon invariant mass (combination of the tag's long track and probe's MuonTT stub) before smearing, and the right panel after. The data are the black points, and the blue line is the simulation. The upper (lower) panels are for μ^+ (μ^-) probes.

A final set of cuts is also made to require that the probe muons are "W-like": i.e. the same as those that would be applied to signal muons in the m_W or $R_{\tau\mu}$ fits (see Tables 8.2 and 9.1), apart from any that would clearly bias the efficiency (e.g. trigger requirements, cuts veto-ing $Z \to \mu\mu$ events etc.). This helps to eliminate any bias from extrapolating from the control to signal samples. Each set of tag-and-probe requirements applies a cut on the dimuon mass to pick out the signal-rich regions. The mass window is wider for the tracking efficiencies because - with no VELO hits - the momentum, and therefore mass, resolution of the MuonTT probes is much poorer than equivalent long tracks. This loss of resolution is also observed to be badly mismodelled in the simulation. To correct this, a simple Gaussian curvature smearing is applied (in addition to the momentum smearing applied to all tracks detailed in Section 6.4). The results of this smearing can be seen in Figure 7.2.

These requirements are sufficient to assume residual background in the $Z \to \mu\mu$ samples is negligible, which is checked by looking for non-peaking components via a

signal+background fit to the m_Z peaks, the result of which can be seen in Figures 7.3, 7.4 and 7.5 for the trigger, MuonID and tracking efficiency samples respectively. In each case, the size of the exponential background component is seen to be very small. Residual background levels are also assessed as a potential systematic uncertainty by varying the tag requirements (see Section 7.6). Some residual background is expected for the $\Upsilon(1S)$ control sample, the handling of which is described in Section 7.4.

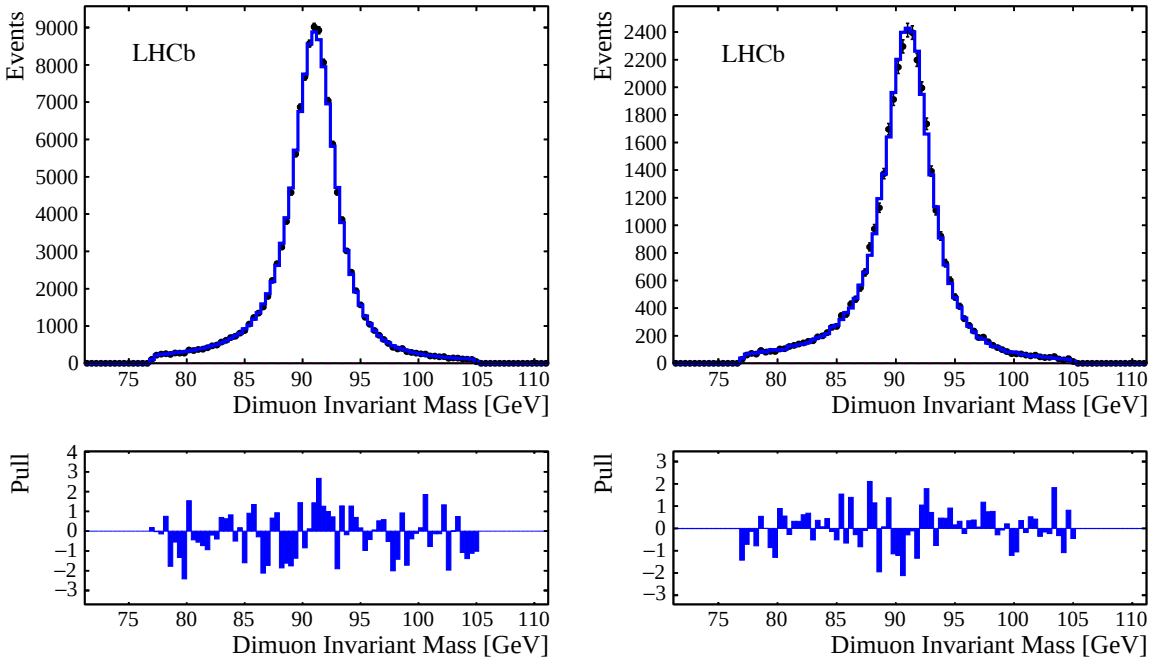


Figure 7.3: Template fits of the trigger efficiency tag-and-probe simulation sample (blue) to the corresponding tag-and-probe data sample (black) near the Z mass peak, with probes passing (left) and failing (right) the trigger efficiency requirement. μ^+ is arbitrarily used as probe. The fitted background shape is too small to be seen.

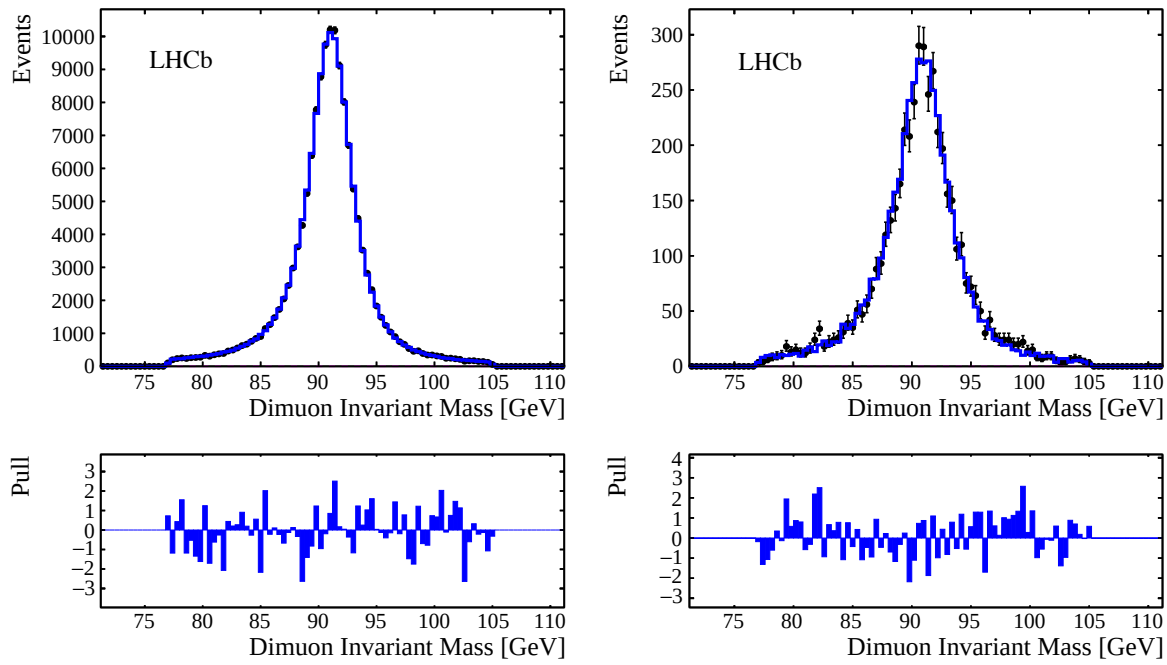


Figure 7.4: Template fits of the MuonID efficiency tag-and-probe simulation sample (blue) to the corresponding tag-and-probe data sample (black) near the Z mass peak, with probes passing (left) and failing (right) the MuonID efficiency requirement. μ^- is arbitrarily used as probe. The fitted background shape is too small to be seen.

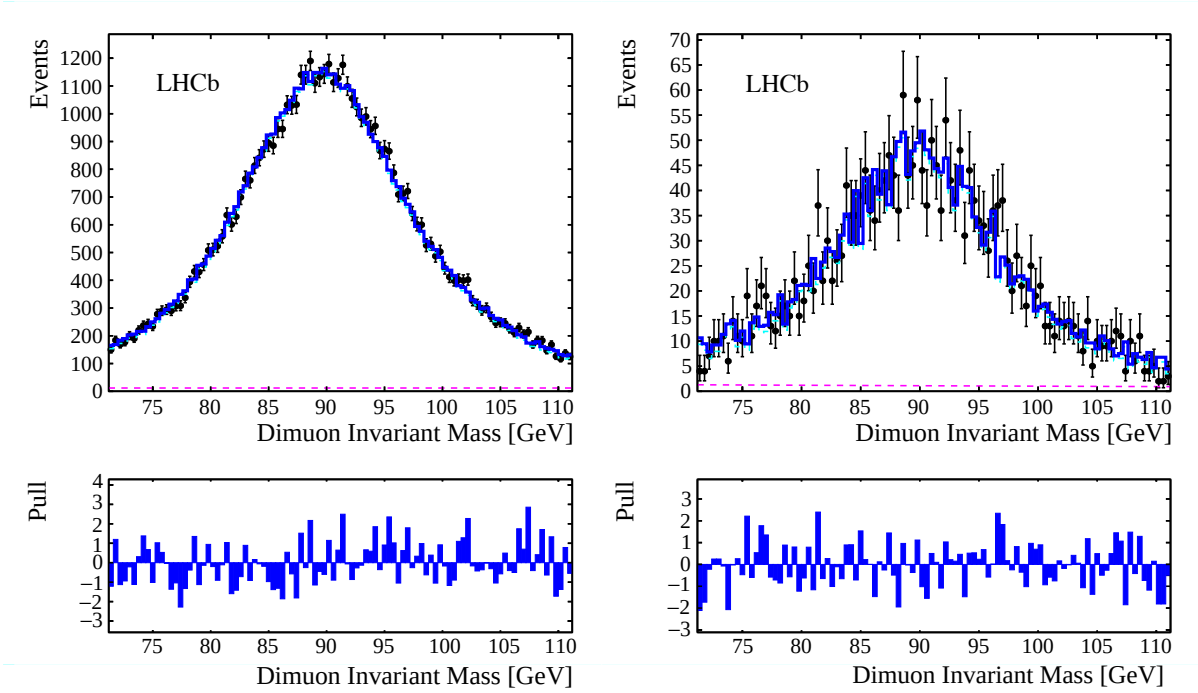


Figure 7.5: Template fits of the tracking efficiency tag-and-probe simulation sample (blue) to the corresponding tag-and-probe data sample (black) near the Z mass peak, with probes passing (left) and failing (right) the tracking efficiency requirement. μ^+ is arbitrarily used as probe. The fitted background shape can be seen here in magenta, which is a negligibly small $1-2\%$ of the total model yield.

7.4 Extraction of efficiency parametrizations

The samples are now ready for the calculation of efficiencies. To properly model any variations in these efficiencies across the detector and the muon's kinematics, the samples are binned up in probe η, ϕ and $1/p_{\mathrm{T}}$. The granularity of this binning is constrained by the size of the control samples. Four uniformly-spaced bins are used for ϕ, and eleven in η. Eight (six) of these cover the η range required of W muons for the m_W ($R_{\tau\mu}$) fitter, and are uniformly-spaced in this interval. For the $\Upsilon(1S)$ sample, three bins of similar population (not uniformly-spaced) at low p_{T} are used. However, due to the limited size of the control sample compared to the signal samples (owing to the greater W production cross section), the Z samples cannot be binned sufficiently finely so as not to introduce unwanted statistical noise into the m_W or $R_{\tau\mu}$ fits. This is partially, but not sufficiently, solved by applying an adaptive binning algorithm to give similarly-populated bins in $1/p_{\mathrm{T}}$. The full solution was to assume that the true efficiency varies smoothly with $1/p_{\mathrm{T}}$, such that it could be described by a simple functional form. Such a form could therefore be fitted to the coarsely-binned points, increasing the statistical power of the control sample. The adaptive binning algorithm begins with very fine bins, then aggregates them together into coarser bins, but it keeps track of the centre-of-gravity of each bin as it is constructed. The fit therefore uses the centres-of-gravity rather than the bin centres, which gives a more accurate parametrization of the efficiency. It is not necessary to apply this binning algorithm to the $\Upsilon(1S) \to \mu\mu$ points.

In each $(\eta, \phi, 1/p_{\mathrm{T}})$ bin, the number of candidates (after the cuts described above) is counted and gives the efficiency denominator. The efficiency numerator is then the number of probes which pass, in addition to the denominator requirements, the relevant pass requirement given in Table 7.1. The efficiency in simulation $\varepsilon_{\mathrm{MC}}$, and the ratio of data to simulation $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ is therefore determined - the latter as it is the quantity relevant for correcting mismodelling in the simulation and is expected to have a shape that is easy to parametrize. Whilst it is sufficient to simply count events near m_Z for the Z samples, there is still a non-negligible amount of background in the $\Upsilon(1S)$ data, which demands that a fit is performed to the invariant mass peak. In each $(\eta, \phi, 1/p_{\mathrm{T}})$ bin, this is a simultaneous fit to the four categories - (data, MC) $\times$ (pass, fail) - to determine $\varepsilon_{\mathrm{MC}}$ and $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$. The signal is modelled with a Crystal-Ball (CB) function [116], and the residual background is flat in invariant mass. The four CB shapes share a common mean (constrained to be within 0.1 MeV of the known $\Upsilon(1S)$ mass), and α parameter controlling the position at which the power-law tail turns on. The CB n parameter is fixed to $n = 2$ throughout to ensure fit stability, and the Gaussian core width σ floats freely in each of the four categories. A representative example of the simultaneous fit is given in Figure 7.6.

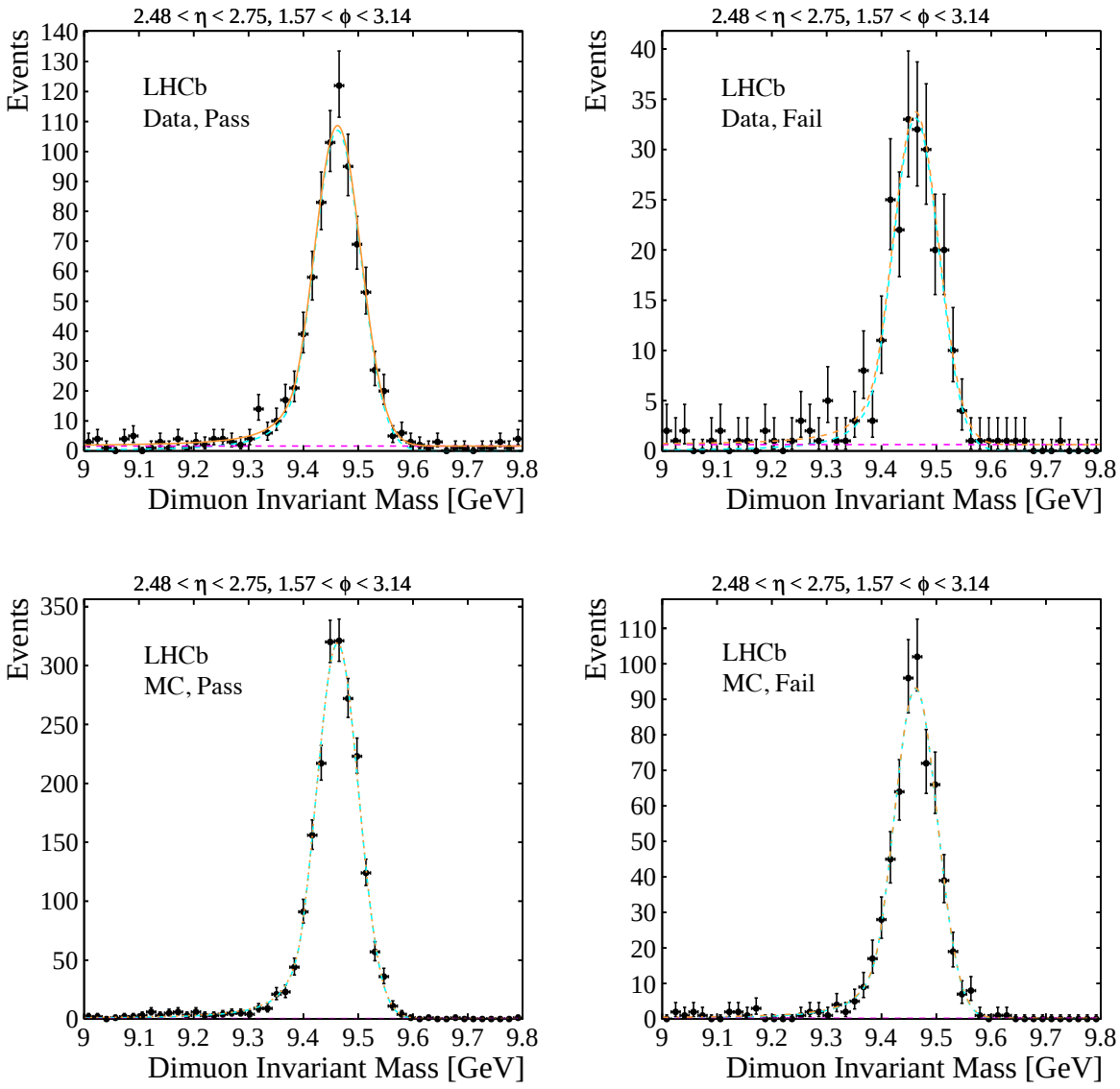


Figure 7.6: Representative binned fit to the $\Upsilon(1S)$ peak, in four simultaneous categories, to extract the trigger efficiency. The η and ϕ ranges are indicated, and for this bin $10 < p_{\mathrm{T}} < 12.5$.

With the binned $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ and $\varepsilon_{\mathrm{MC}}$ values now available, the fitting of a functional form to the values as a function of $1/p_{\mathrm{T}}$ can proceed in each (η, ϕ) bin. For the trigger efficiency, the turn-on of L0 at $p_{\mathrm{T}} = 6\,\mathrm{GeV}$ (and the poor momentum resolution of L0) motivates the usage of an error function for $\varepsilon_{\mathrm{MC}}$ in the trigger efficiencies, with the turn-on controlled by the $\Upsilon(1S)$ points. A linear function is found to be sufficient to fit the $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ points. No such turn-on is expected with p_{T} for the MuonID and tracking efficiencies, and in any case coverage is lacking at low p_{T} in the control samples. Fitting with a flat (linear) function for the tracking (MuonID) efficiencies describes the $1/p_{\mathrm{T}}$ shape well, although this is varied as a systematic uncertainty in both the m_W and $R_{\tau\mu}$ analyses. These fits are the key result of this treatment of the muon reconstruction efficiencies, and as such

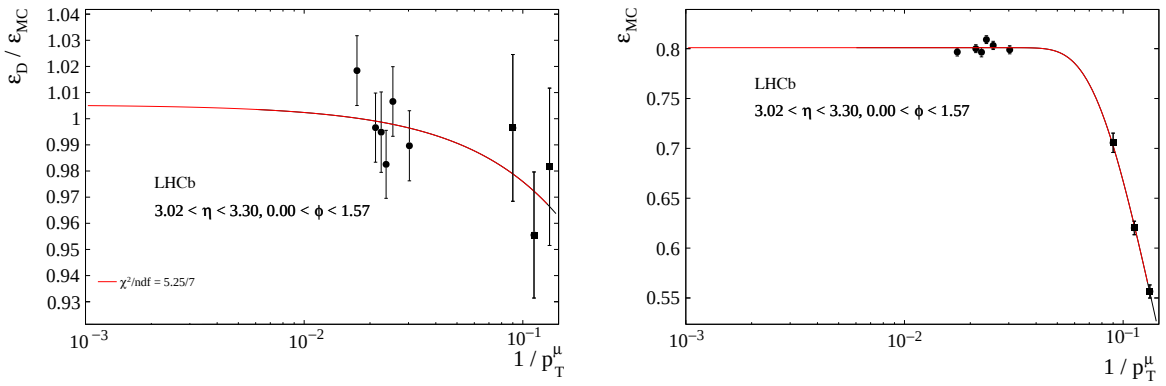


Figure 7.7: Representative binned fit to $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ $(1/p_{\mathrm{T}})$ (left) and $\varepsilon_{\mathrm{MC}}$ $(1/p_{\mathrm{T}})$ (right). The η and ϕ ranges are indicated. The rightmost three points (lowest p_{T}) in each plot belong to the $\Upsilon(1S) \to \mu\mu$ probes - the rest are from $Z \to \mu\mu$.

are included at the relevant point in each of Chapters 8 and 9. A representative example of the binned trigger efficiency fits is shown in Figure 7.7 for a single (η, ϕ) bin.

7.5 Event weights

In both the m_W and $R_{\tau\mu}$ analyses, mismodelling of the muon reconstruction efficiencies is corrected via the application of weights to each event in simulation[2] depending on the muon kinematics. For $W \to \mu\nu$ signal muons (and any backgrounds with a signal muon in the final state), the appropriate weight is simply the product of the three $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ factors - one for each efficiency, evaluated from the relevant parametrized form at the $(\eta, \phi, 1/p_{\mathrm{T}})$ of the signal muon. Dimuon signal samples require that either muon satisfies the identification requirements (see e.g. Table 6.1), so in this case the appropriate weight depends on how many of the muons passed the requirements. For example, if the positive muon was to pass, but the negative muon failed, then the appropriate weight factor for each efficiency w_i is

$$w_i = \frac{\varepsilon_{\mathrm{D},i}(\mu^+)}{\varepsilon_{\mathrm{MC},i}(\mu^+)} \times \frac{1 - \varepsilon_{\mathrm{D},i}(\mu^-)}{1 - \varepsilon_{\mathrm{MC},i}(\mu^-)}\,, \tag{7.3}$$

thereby correcting the mismodelled efficiency for the positive muon, and the mismodelled *inefficiency* for the negative muon. The final case is of dimuon backgrounds where one muon was not reconstructed, which therefore becomes a background for $W \to \mu\nu$. Here, both the efficiency of correctly reconstructing one muon, and the inefficiency of not reconstructing the second, need to be corrected. If the second muon simply does not fall

[2]Note that the DIF simulation in the m_W analysis (see Section 6.8) does not receive these weights, as the model is trained on data.

in the LHCb acceptance then no weight is required to correct for the inefficiency, but if it has failed to be reconstructed as a long track or to have MuonID, then a factor is required. This is applied based upon the true simulated kinematics of the second muon, as no reconstruction-level information is available.

7.6 Systematic uncertainties

There were a number of semi-arbitrary choices in the muon reconstruction efficiency treatment. All of these should be assessed as potential systematic uncertainties to the parameter of interest (m_W or $R_{\tau\mu}$). There is also statistical uncertainty propagated due to the limited size of the control samples. To assess these sources of uncertainty, in each case, a single variation is made in the treatment, and the appropriate parts of the analysis are re-run to extract the value of m_W or $R_{\tau\mu}$ after making the variation. A set of such values is then used to calculate the uncertainty, as will be described for each analysis in Sections 8.7 and 9.8. The variations to the treatment for both analyses are the same, and are as follows.

Statistical uncertainty is assessed by varying the $\varepsilon_{\mathrm{MC}}$ and $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ points within their error bars according to Gaussian random numbers, and then re-performing the efficiency parametrization fits. Systematic uncertainties in the adaptive binning algorithm are assessed by increasing or decreasing the number of coarse $1/p_{\mathrm{T}}$ bins the algorithm aims to produce, as well as changing how it makes its decisions of how to aggregate bins. The tag requirements shown in Tables 7.2, 7.3, 7.4 and 7.5 were loosened and tightened, and the number of η and ϕ bins was increased and decreased. For the tracking efficiencies alone, the amount of additional momentum smearing applied was increased and decreased, and a linear function was tried in the $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ fits. A flat function was used for the MuonID $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ and $\varepsilon_{\mathrm{MC}}$ fits, and an error function for the trigger $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ fits, as well as both of these together. The largest variations are found to come from these variations in the choice of functional form, which is not surprising given it is the efficiency mismodelling as a function of $1/p_{\mathrm{T}}$ that is most important to the two analyses.

7.7 Closure test

A reassuring sanity check of the method is to perform a *closure test*; to measure the efficiency in simulation again after the application of corrections via weights, and see that the efficiency matches the efficiency in data. The largest kinematic dependency and largest mismodelling was seen as a function of η for the trigger efficiencies. It was therefore chosen to re-measure the trigger efficiency in simulation in coarse bins of η, as is shown in

Figure 7.8. These efficiency calculations are made by a simple script that is independent of all the code used in the rest of the efficiency treatment, which gives greater confidence in the method. As the figure shows, the data efficiency is reproduced after the application of the standard corrections.

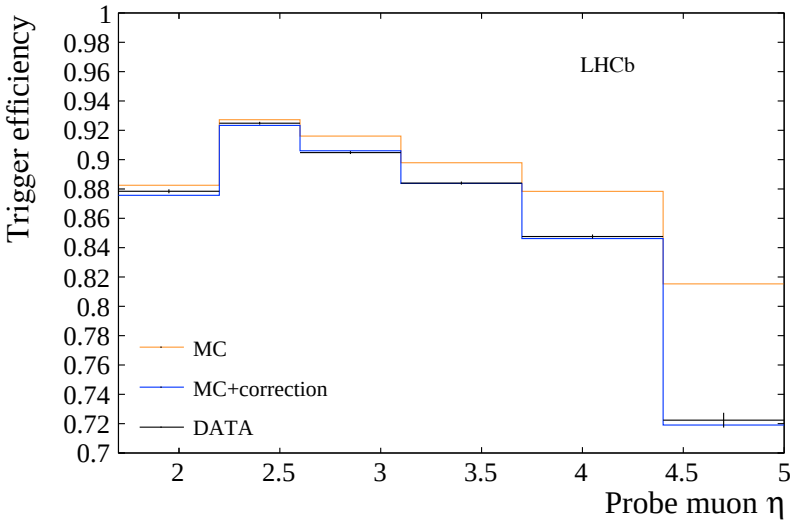


Figure 7.8: Trigger efficiency as a function of η for data, simulation and simulation after correction, testing the closure of the method.

7.8 Conclusion

In this chapter, the author's work on devising and implementing a treatment of high-p_{T} muon reconstruction efficiencies for precision electroweak analyses has been presented. The development of these studies was driven by the advancing m_W measurement, and the high precision of that measurement required a treatment that was complex and comprehensive. In particular, the problem of the (comparatively) small Z control samples drove the development of an adaptive binning algorithm, and then the decision to parametrize the efficiencies as a function of $1/p_{\mathrm{T}}$. Without these steps in the method, the statistical uncertainty propagated to the m_W fit via the $1/p_{\mathrm{T}}$ shape was intolerable. The additional information from the $\Upsilon(1S)$ points came relatively late-on, as it was realized that points from just the Z give little coverage across $1/p_{\mathrm{T}}$, particularly when it is known that L0 has a broad turn-on at low p_{T}. In the end, the treatment was sufficient to ensure that the muon reconstruction efficiencies were not a large systematic uncertainty for the 2016 m_W measurement, as will be shown at the end of Chapter 8. For the $R_{\tau\mu}$ measurement, where the exact $1/p_{\mathrm{T}}$ shapes are less important than their normalization, the muon reconstruction efficiencies take on less importance, but are nonetheless comprehensively

controlled by the treatment outlined here.

CHAPTER 8

Measurement of the W boson mass

In the preceding chapters the physics and detector modelling strategies developed for high-precision measurements of electroweak (EW) parameters have been introduced. This chapter now details the first such measurement that the author was involved in: the measurement of the W boson mass m_W with the 2016 LHCb dataset, published in Ref [124]. This measurement was accomplished by a team of analysts, and the author's main contribution was detailed in Chapter 7. The work in this chapter is therefore largely the work of others.

The chapter proceeds by first giving an introduction to the context and motivation of the measurement, before moving on to detail the fitting strategy in Section 8.2, which is validated in Section 8.3. The selection requirements applied to the data to extract $W \to \mu\nu$ signal events are detailed in Section 8.4, followed by the preparation of signal and background templates in Section 8.5, which reviews and puts into context the methods described in Chapters 6 and 7. Important results of the application of those methods for the m_W measurement are shown in this section. The central m_W fit result is then showcased in Section 8.6. Sections 8.7 and 8.8 describe the evaluation of systematic uncertainties and cross-checks on the measurement respectively. The conclusions and outlook are stated in Section 8.9.

8.1 Introduction

Section 2.2 reviewed how the Higgs mechanism of spontaneous symmetry breaking provides a way to generate the masses of the W and Z bosons, as well as how they relate to one another and other Standard Model (SM) parameters. This allows the derivation of equations such as

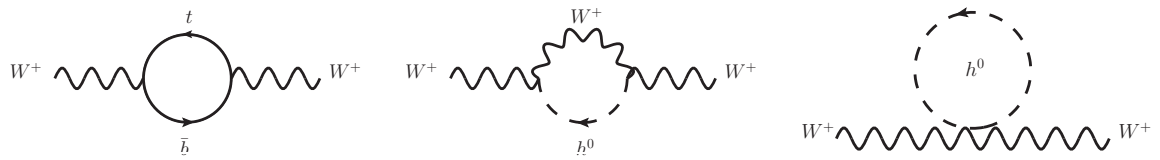


Figure 8.1: Feynman diagrams representing quantum corrections to the $W^{\pm}$ mass. Produced with the JaxoDraw tool [26].

$$m_W^2 \left(1 - \frac{m_W^2}{m_Z^2}\right) = \frac{\pi\alpha}{\sqrt{2}G_F}(1 + \Delta)\,, \tag{8.1}$$

where α is the fine structure constant of Quantum Electrodynamics (QED) and G_F is the Fermi constant of weak decays. The Δ term encodes the effect of higher-order loop corrections to m_W, for instance due to the top quark and Higgs masses [126, 127]. Feynman diagrams representing such corrections are shown in Figure 8.1. Global EW fits can therefore be performed, where the EW model is tested for consistency with all direct measurements of the relevant SM parameters. Additionally, if direct measurements of m_W are removed from this fit, it can be used to give an indirect SM prediction. Comparing this prediction with direct measurements is a stringent test of the SM, and any significant discrepancy would be a strong hint of new physics. An example of such phenomena would be new heavy particles coupling to the W boson via diagrams such as those in Figure 8.1, which would modify masses, couplings and/or relations between them. At the time of publication, the latest prediction from a global EW fit for m_W was [128]

$$m_W^{\text{pred}} = 80354 \pm 7\,\text{MeV}\,, \tag{8.2}$$

the uncertainty of which has since been improved to 6 MeV [129]. The single highest-precision measurement of m_W at that time (and also the sole measurement from the LHC despite over 10 years of running) was by the ATLAS collaboration, achieving a 19 MeV uncertainty [130]. m_W had also previously been measured to a precision of 33 MeV by the combination of the LEP experiments [131] and 16 MeV by the Tevatron experiments [132]. An official ATLAS-Tevatron combination was (and still is) in progress. The 2020 PDG average of direct measurements was [88]:

$$m_W^{\text{PDG}} = 80379 \pm 12\ \text{MeV}\,. \tag{8.3}$$

This gap - of almost a factor of two in precision - between direct and indirect measurements of m_W provides strong motivation for new, high-precision direct measurements of m_W. The ATLAS measurement, despite using only a small subset of their data collected to date, was already limited by the modelling of W boson production, in

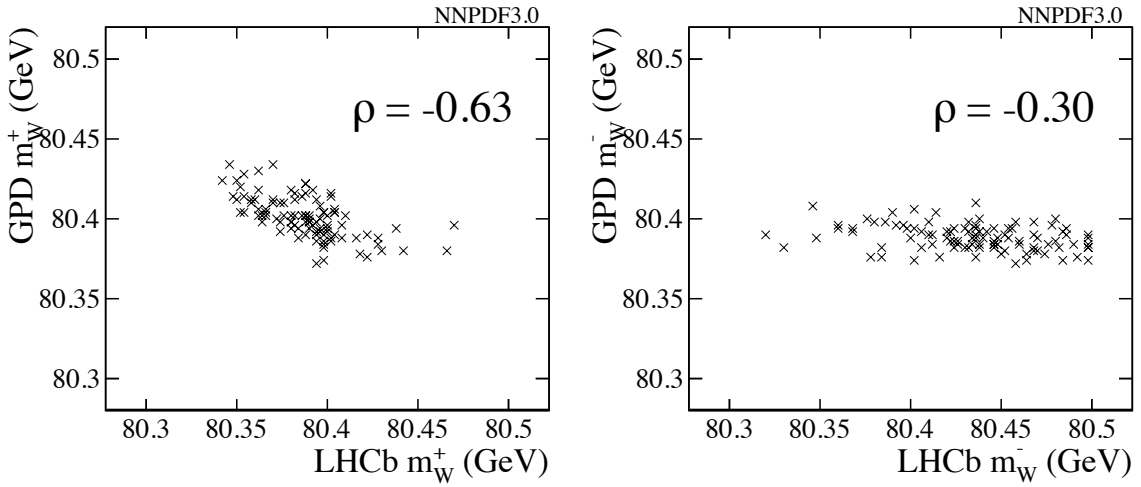


Figure 8.2: Results of template fits for m_W using simulated toy datasets with a GPD and with LHCb. Although the spread in LHCb measurements is larger owing the larger PDF uncertainty, a linear combination with lower overall variance can be found. Taken from Ref. [134].

particular by the PDF uncertainty. This uncertainty originates from the extraction of the PDFs themselves from experimental data. At the Tevatron, W production in $p\overline{p}$ collisions was dominated by valence quark annihilation, whereas processes involving gluons and sea quarks - with less precisely known PDFs - are far more important in pp collisions at the LHC [133]. This situation is worse for LHCb than the "general-purpose detectors" (GPDs) ATLAS and CMS, due to the larger uncertainty in sea quark PDFs in the high partonic x region that LHCb probes. However, Ref. [134] showed that, if a m_W measurement from LHCb were combined with one from ATLAS or CMS, the PDF-related uncertainty may partially cancel. This derives from their largely orthogonal pseudorapidity (η) acceptances, which leads to an anti-correlation in m_W measurements because of PDF momentum sum rules. This anti-correlation is illustrated in Figure 8.2. An appropriately-weighted average can therefore yield a substantially reduced total PDF uncertainty, even if the LHCb result alone has a larger uncertainty than a GPD measurement. In Ref. [134] it was estimated that with the full Run 2 (2015-2018) LHCb data, a statistical precision of 10 MeV would be achievable. If systematic uncertainties could be controlled at a similar level, this would have a substantial impact in reducing the precision gap between direct and indirect m_W measurements. Unfortunately the PDF uncertainty is not the only important modelling challenge, and it was clear that substantial collaboration is needed with the theory community to control all the relevant uncertainties at the 10-20 MeV level. This measurement therefore uses only 2016 LHCb data, and serves as a proof-of-principle of LHCb's capability and a first step towards the full Run 2 measurement.

8.1.1 Strategies for measuring the W mass

In hadron colliders, W boson properties are studied via W production in the Drell-Yan process $pp \to W + X \to \ell\nu + X$ as introduced in Section 2.5. The neutrino is not reconstructible here, so direct reconstruction of the dilepton invariant mass - as might be used to measure the mass of the Z boson - is not possible. Instead, three observables are typically used to infer m_W indirectly: the charged lepton p_{T}; the neutrino p_{T}; and the transverse mass M_{T}. The neutrino's p_{T} can be inferred using transverse momentum conservation:

$$\mathbf{p}_{\mathrm{T}}^{\nu} = -\mathbf{p}_{\mathrm{T}}^{\ell} - \sum_i E_i \sin\theta_i \,, \tag{8.4}$$

where the last term is referred to as the *hadronic recoil* u_{T}, and the sum runs over all the clusters of energy deposition in the detector's calorimeter, exclusive of the charged lepton. The neutrino p_{T} can also be referred to equivalently as simply the missing transverse energy $E_{\mathrm{T,miss}}$, since the leptons have negligible mass at these energies and all their energy is effectively in their momenta. The transverse mass is defined as

$$M_{\mathrm{T}} = \sqrt{2\left(p_{\mathrm{T}}^{\ell} p_{\mathrm{T}}^{\nu} - \mathbf{p}_{\mathrm{T}}^{\ell} \cdot \mathbf{p}_{\mathrm{T}}^{\nu}\right)} \,. \tag{8.5}$$

Reconstruction of the latter two variables requires calorimeter coverage of the whole solid angle, which is a typical feature of GPDs, but not for LHCb. All three variables are highly-correlated with m_W: for example the charged lepton p_{T} exhibits a "Jacobian" peak, with a falling edge at $\sim m_W/2$. $E_{\mathrm{T,miss}}$ peaks similarly, while the transverse mass has an edge at $\sim m_W$. At the Tevatron, M_{T} was the most sensitive variable [132], but in the higher pileup environment at the LHC, the charged lepton p_{T} is more suitable. In the aforementioned ATLAS measurement, $E_{\mathrm{T,miss}}$ wasn't considered at all, and the M_{T} fit had only a small impact on their final result. In all cases, measurement of m_W is achieved by template-fitting the m_W-sensitive distributions, with the best-fitting template corresponding to the best-fitting value of m_W, as introduced in Section 2.5. If multiple distributions are fitted, they will be fitted simultaneously, or their measured values averaged to produce one measurement of m_W. For the reasons listed above, and because electron reconstruction is challenging at LHCb, the LHCb measurement focused on the muon transverse momentum in $W \to \mu\nu$ decays.

8.2 Fitting strategy of the present analysis

The overall fitting strategy is more complex than a simple template fit to the muon p_{T}, and was devised (and revised during development) in response to the impact on m_W of a

number of key modelling factors. The first, mostly aesthetic, departure from this simple fit was the choice to instead fit the muon q/p_{T} distribution. This variable provides a way to simultaneously analyse (and plot) both muon charges, and allows the whole high-p_{T} region to be placed on the plot. While the latter region is not used in the fit, it does provide a helpful control region as it is very sensitive to a number of relevant systematic uncertainties. Figure 8.3 illustrates the fitting principle, with two simulated templates differing only by the values of m_W used in their preparation. This figure also illustrates the scale of the modelling challenge: its two values of m_W, differing at the $\mathcal{O}(100\,\mathrm{MeV})$ level, shift the Jacobian edge up and down by around 1%; however the target of this analysis is to measure m_W with a precision that is more than a factor of ten better than this.

The strategy also relies on fitting a number of nuisance parameters to the data in addition to m_W, all of which help to account for imperfect physics predictions. For example, during the development of the analysis, it was discovered that propagating the uncertainties from the predictions of the angular coefficients - particularly of A_3 - to m_W initially yielded a dominating $\sim 30\,\mathrm{MeV}$ uncertainty. The lepton p_{T} distribution can be approximated as

$$p_{\mathrm{T}}^{\mu} \approx \frac{m_W}{2} \sin\theta + p_{\mathrm{T}}^{W} \cos\phi\,, \tag{8.6}$$

which when compared to the A_3 term in the 5D cross section (Eqn. 2.23) shows

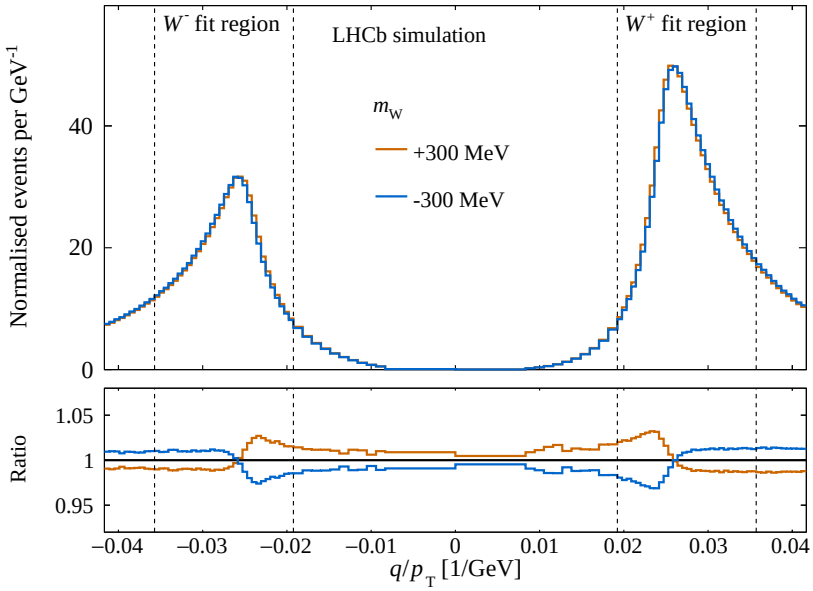


Figure 8.3: Illustration (with LHCb simulation) of how m_W can be determined from the shape of the muon q/p_{T} distribution.

the importance of the A_3 term. This uncertainty was reduced to around 10 MeV by introducing a floating A_3 scale factor (shared between the two W charges) to the fit, and effectively only relying on the kinematic dependence of predictions of A_3, not the overall scale. This significantly reduced the A_i-related uncertainty to around 10 MeV.

Equation 8.6 also highlights the importance of accurately modelling the p_{T}^W distribution in a measurement of the lepton p_{T} distribution. p_{T}^W cannot be directly reconstructed by LHCb, as it requires reconstructing p_{T}^ν ($E_{\mathrm{T,miss}}$), but even in ATLAS and CMS p_{T}^W cannot be measured with sufficient precision to serve as a useful input to a m_W measurement [135, 136]. Previous m_W measurements have therefore relied on tuning event generator predictions of p_{T}^Z to the measured distribution in $Z \to \mu\mu$ decays, with systematic uncertainties assigned to cover the extrapolation to p_{T}^W. Ref. [137] showed that variations in the Quantum Chromodynamics (QCD)-related event-generator tuning parameters $k_{\mathrm{T}}^{\mathrm{intr}}$ and α_s affect the lepton p_{T} distribution differently to variations in m_W, such that these parameters could be floated alongside m_W in a simultaneous fit to W and Z boson data. In this analysis, rather than p_{T}^Z, the angular variable ϕ^* [119] is fitted, defined in Section 6.3 but repeated here for convenience:

$$\phi^* \equiv \tan\phi_{\mathrm{acop}} \sin\theta_\eta^* \approx \frac{p_{\mathrm{T}}^Z}{m_Z} \,. \tag{8.7}$$

It is chosen over p_{T}^Z as it can be measured with greater precision, and is less sensitive to detector mismodelling effects. Floating m_W, the A_3 scale factor, $k_{\mathrm{T}}^{\mathrm{intr}}$ and α_s amount to reweighting the templates during every iteration of the fit, according to the 5D cross section predictions associated with the current parameters values, as described in Section 5.1. A final set of weights are applied each iteration to account for the floating value of m_W:

$$w_{\mathrm{mass}}(m, m_W) = \frac{(m^2 - (m_W^{\mathrm{Pythia}})^2)^2 + m^4(\Gamma_W^{\mathrm{Pythia}})^2/(m_W^{\mathrm{Pythia}})^2}{(m^2 - m_W^2)^2 + m^4(\Gamma_W^{\mathrm{Pythia}})^2/m_W^2} \,, \tag{8.8}$$

where Γ_W is the width of the W boson. This is simply a ratio of relativistic Breit-Wigner functions with a mass-dependent width. The value of the width has negligible impact on the muon p_{T} distribution.

Significant backgrounds are expected, particularly in the $W \to \mu\nu$ sample. Prompt hadrons decaying-in-flight (DIF) have an unknown cross section, so their fit fraction is allowed to float freely. The fit fraction $f_{W\to\tau\nu}$ for $W \to \tau\nu$ (with $\tau \to \mu$ decays) is constrained with respect to the signal fraction as follows. Taking into account the differing reconstruction-and-selection efficiency (which is the sum of weights in the template $\sum w$ divided by the total number of generated fiducial events $N^{\mathrm{gen,fid}}$), this constraint becomes

$$f_{W\to\tau\nu} = f_{W\to\mu\nu} \times \frac{\sum w_{W\to\tau\nu}}{\sum w_{W\to\mu\nu}} \times \frac{N^{\text{gen,fid}}_{W\to\mu\nu}}{N^{\text{gen,fid}}_{W\to\tau\nu}} \times \mathcal{B}(\tau\to\mu\nu\bar{\nu})\,, \tag{8.9}$$

where lepton flavour universality has been assumed, and the PDG world average value of $\mathcal{B}(\tau\to\mu\nu\bar{\nu}) = 0.1739$ was used [88]. The $W\to\tau\nu$ fraction is evaluated separately for each charge such that the charge asymmetry is consistent with the $W\to\mu\nu$ components. The remaining backgrounds, which include $Z\to\mu\mu$, $Z\to\tau\tau$ and heavy-flavour hadrons in the W sample, and rare backgrounds (e.g. low-mass Drell-Yan, and from top quarks and vector boson pairs) in both the W and Z samples, are constrained relative to the observed number of $Z\to\mu\mu$ candidates N^{Z}_{obs} in data via their fiducial cross section $\sigma^{\text{fid}}_{\text{comp}}$. This constraint is

$$f^{\text{term}}_{\text{comp}} = \frac{N^{Z}_{\text{obs}}}{N^{\text{term}}_{\text{obs}}} \times \frac{\sum w^{\text{term}}_{\text{comp}}}{\sum w^{\text{Z}}_{\text{Z}}} \times \frac{\sigma^{\text{fid}}_{\text{comp}}}{\sigma^{\text{fid}}_{\text{Z}}} \times \frac{N^{\text{gen,fid}}_{\text{Z}}}{N^{\text{gen,fid}}_{\text{comp}}}, \tag{8.10}$$

where $N^{\text{term}}_{\text{obs}}$ is the total number of data events selected in either the W or Z data as appropriate ("term" refers to which term they contribute to in the total χ^2, W or Z, and hence the selection cuts applied). For example, $\sum w^{\text{Z}}_{\text{Z}}$ is the sum of weights of Z template events passing the Z selection, as defined in Table 6.1. Note that the weights that are being summed include the dynamically-varying physics weights described above that change iteration-to-iteration. The fiducial cross section for $Z\to\mu\mu$ at $\sqrt{s} = 13\,\text{TeV}$ has been measured [138], while the others are predicted with POWHEG and the NLO NNPDF3.1 PDF sets, using the same fiducial acceptance requirement as the $Z\to\mu\mu$ cross section measurement for all predictions.

The total fit model is a sum of signal and background templates, fitting simultaneously the q/p_{T} in a $W\to\mu\nu$ data sample, and dimuon ϕ^* in a $Z\to\mu\mu$ data sample. The templates were generated with the LHCb simulation (as introduced in Section 3.7), and in the default configuration the important Drell-Yan components ($W\to\mu\nu$, $Z\to\mu\mu$, $W\to\tau\nu$ and $Z\to\tau\tau$) were reweighted to match unpolarized cross section predictions by POWHEGPYTHIA and angular coefficient predictions made by DYTURBO at $\mathcal{O}(\alpha_s^2)$, as motivated in Section 5.1. The templates are also reweighted to match QED FSR predictions by the average of PYTHIA, PHOTOS and HERWIG. The fit region in q/p_{T} was $28 < p_{\text{T}} < 52\,\text{GeV}$, and in ϕ^* was $0 < \phi^* < 0.5$. The fit is achieved by utilizing MINUIT [139] to minimize a sum of two negative log-likelihood terms (W and Z), with the Beeston-Barlow Lite prescription [140] taking account of the finite sample sizes used to make the templates. These terms are multiplied by two to be equivalent to a χ^2, which is how it will be referred to in the following. The χ^2 is minimized with respect to eight floating parameters. For completeness and clarity, these are:

- m_W,

- The $W^+ \to \mu^+\nu_\mu$ fit fraction,
- The $W^- \to \mu^-\overline{\nu}_\mu$ fit fraction,
- The total DIF fraction (charge asymmetry is fixed),
- A_3 scaling factor for both W charges (unnecessary for Z),
- α_s^W shared by both muon charges for all W-boson processes,
- α_s^Z for all Z-boson processes,
- $k_{\mathrm{T}}^{\mathrm{intr}}$ shared amongst W^+, W^- and Z.

The m_W value was blinded during the analysis by adding an offset drawn from a uniform distribution of random values in the range of $[-0.5, 0.5]$ GeV. During the final stages of collaboration review, this blinding offset was removed.

8.3 Pseudodata validation

The fit model features a number of novel ideas, and while it was concluded in Section 5.2 that POWHEGPYTHIA gave the best description of p_{T}^Z data, this may not hold true for p_{T}^W. For these reasons, it is important to validate the fit model against predictions from other p_{T}^W models. To do this, the central model was used to fit *pseudodata* generated with different models of p_{T}^W. For simplicity and to isolate effects due only to the physics modelling, these fits were conducted using a fit model and pseudodata without detector effects included, and backgrounds were also not included. Six different generator configurations - as introduced in Section 5.2 were used to prepare the pseudodata: POWHEGPYTHIA, POWHEGHERWIG, HERWIG, DYTURBO and PYTHIA with two different PDF sets. A spread in the tuning parameters was used to try to give a representative variance between the predictions. The results of these pseudodata fits are shown in Figure 8.4, and the fit results are listed in Table 8.1. Whilst there is large variation in fit quality - highlighted by the χ^2 values and the ratio panels - the floating α_s and $k_{\mathrm{T}}^{\mathrm{intr}}$ successfully adjust for the different predictions of QCD, and the variations of the preferred m_W are no more than 10 MeV. This test demonstrates that the fit model has sufficient flexibility to describe the underlying boson production and simultaneously extract m_W with high precision.

8.4 Signal selection

The Z data sample used for the m_W fit is the same as the `Z` control sample, as described in Section 6.1. The W sample is a sample of 2016 LHCb data with a single, high-p_{T} muon in the final state. The selection requirements for this sample are listed in Table 8.2, and overall it yields 2.4 million W boson candidates.

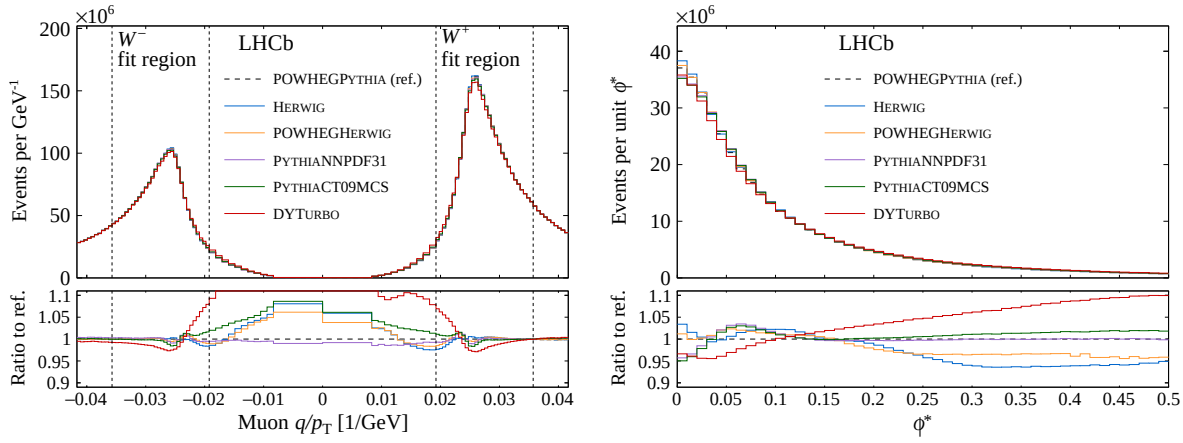


Figure 8.4: Projections of the q/p_{T} (left) and ϕ^* (right) distributions of each pseudo-dataset. The four dashed vertical lines indicate the two fit regions in the q/p_{T} distribution.

Table 8.1: Fit results from the pseudodata validation fits. The contributions to the total χ^2 from the q/p_{T} and ϕ^* distributions are denoted χ^2_W and χ^2_Z, respectively. The shift in the m_W value with respect to the POWHEGPYTHIA pseudodata is denoted δm_W. The uncertainties quoted are statistical.

Data config.	χ^2_W	χ^2_Z	δm_W [MeV]
POWHEGPYTHIA	64.8	34.2	–
HERWIG	71.9	600.4	1.6
POWHEGHERWIG	64.0	118.6	2.7
PYTHIA, CT09MCS	71.0	215.8	−2.4
PYTHIA, NNPDF31	66.9	156.2	−10.4
DYTURBO	83.0	428.5	4.3

The "Z-veto" cut is aptly named: if a second high-p_{T} track (with transverse momentum p'_{T} and pseudorapidity η') is present in the acceptance (as there would be for a Z decay), the event is rejected. No muon identification requirement is made on this second "muon", as that would reduce the effectiveness of the cut. The cut on the isolation variable (introduced in Section 6.1) is particularly effective at removing backgrounds from hadronic decays, the other products of which tend to be boosted to a similar direction as the muon in LHCb. The cut on η avoids the edges of the detector, where in particular the isolation cone will not be fully instrumented. The detector alignment and momentum scale calibration is also expected to be worse near the edges of the detector. A maximum impact parameter significance cut suppresses $W \to \tau\nu$ and heavy-flavour decays. The other cuts reject small amounts of poorly-reconstructed events. The cut with the largest rejection rate (with respect to all other cuts being applied) is the isolation cut.

Table 8.2: Selection requirements applied to select a high-p_{T} single-muon sample for measuring m_W. "TOS" stands for "trigger-on-signal", meaning that the signal muon was responsible for firing the trigger lines. The trigger lines are as described in Section 3.5: `L0MuonEW`, `Hlt1SingleMuonHighPt` and `Hlt2SingleMuonHighPt`.

Description	Requirement
Z-veto	$p'_{\mathrm{T}} < 25\,\mathrm{GeV}$ and $(2.0 < \eta' < 4.5)$
Pile-up	Number of candidates in event = 1
Sanity cut	$p < 2\,\mathrm{TeV}$
Trigger	Muon is TOS at L0, HLT1 and HLT2
η acceptance	$2.2 < \eta < 4.4$
Impact parameter significance	$\chi^2_{\mathrm{IP}} < 9$
Track quality	$\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 1.8$
HCAL activity	$E_{\mathrm{T}}^{\mathrm{HCAL}} < 5.0\,\mathrm{GeV}$
Isolation	$\mathcal{I}^{\mathrm{PF},0.4} < 4.0\,\mathrm{GeV}$
Muon PID	Muon is positively identified

8.5 Preparation of data templates

The majority of the analysis workflow (in terms of lines of code and CPU-time) is taken up by a series of steps preparing the templates and the data for the m_W fit. The strategies behind each of them has been explained in Chapters 5, 6 and 7.

The first three steps of note in the analysis are concerned with accurately reproducing LHCb's momentum measurement. The first step is to determine time-dependent corrections to the momentum scale of LHCb, as described in Section 6.2. Figure 8.5 shows the fill-number distribution and the resulting corrections to the momentum scale, which find systematic variations of $\mathcal{O}(10^{-4})$ in the momentum scale throughout the data-taking period. As a closure test, the procedure was applied on the corrected data, and found corrections that were consistent with zero, which can also be seen in the figure. These corrections are then applied as $p \to p \cdot (1 + \alpha)$ to all muons in the signal and control data samples.

The second step is to effectively re-align the detector offline to remove troublesome track curvature biases using the pseudomass method, as detailed in Section 6.3. Indexed by detector region, the total curvature corrections δ are shown for data and simulation in Figure 8.6. It can be seen that much larger corrections are needed for the data compared to the simulation. Application of these corrections was seen to greatly improve the Z-mass resolution in the data, and largely removes observed biases in e.g. m_W extracted with different magnet polarities.

With the detector "re-aligned" for both data and simulation, the main momentum

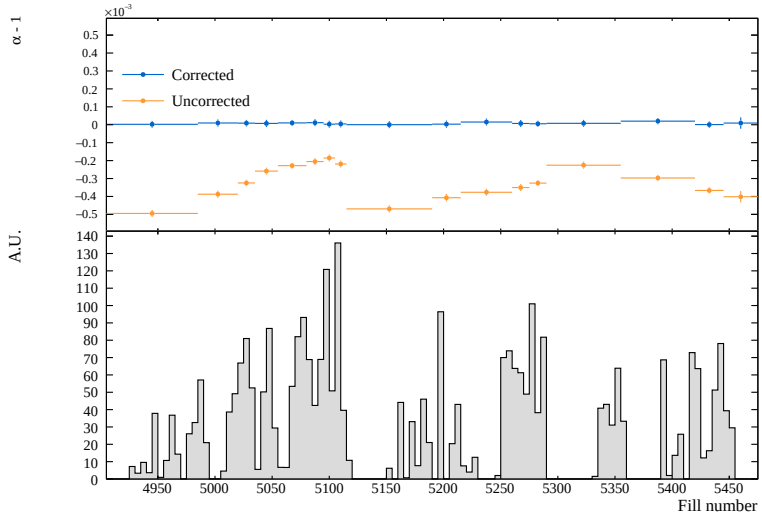


Figure 8.5: The fill number distribution and the momentum scale corrections in coarse bins of fill number.

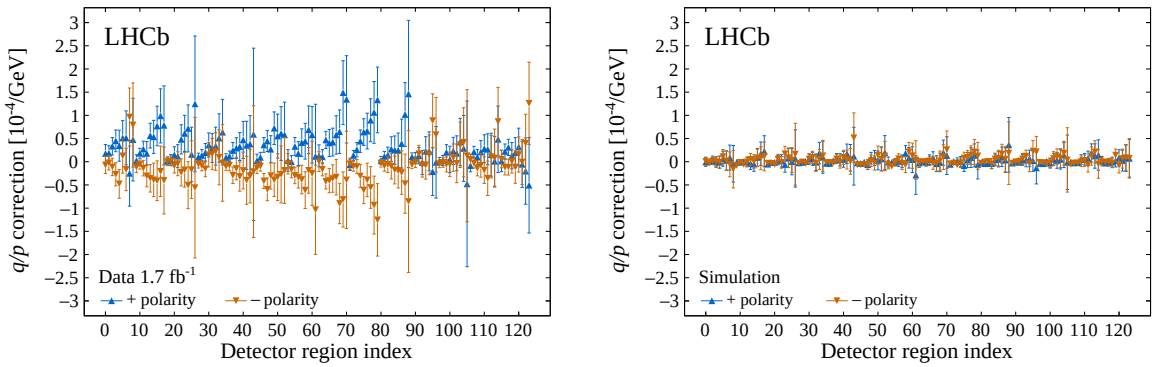


Figure 8.6: Curvature corrections, indexed in terms of detector region (determined by muon η, ϕ, and whether the muon passed through the IT or OT) for (left) data and (right) simulation. The different coloured points correspond to the two different polarities, and the periodicity is due to a pseudorapidity dependence that repeats in each ϕ interval.

scaling and smearing fit is carried out with Z, $\Upsilon(1S)$ and J/ψ data and simulation, as detailed in Section 6.4. Across the 36 fit categories in sample, detector region and magnet polarity, the fit quality was good, with $\chi^2/\text{ndf} = 1862/2082$. Re-combining the fit categories for each sample, the fit result is shown in Figure 8.7. The resulting fit parameter values (and uncertainties) are shown in Table 8.3. A typical signal muon in this dataset has $p \sim 400\,\text{GeV}$, meaning that the residual curvature bias is a few parts per mille, and the curvature resolution smearing is $\mathcal{O}(1\%)$ when the $1/\cosh\eta$ term (≈ 0.1 for $\eta = 3$) is accounted for.

To ensure the efficiencies of the loose χ^2_{IP} and χ^2_{trk} requirements are well-modelled,

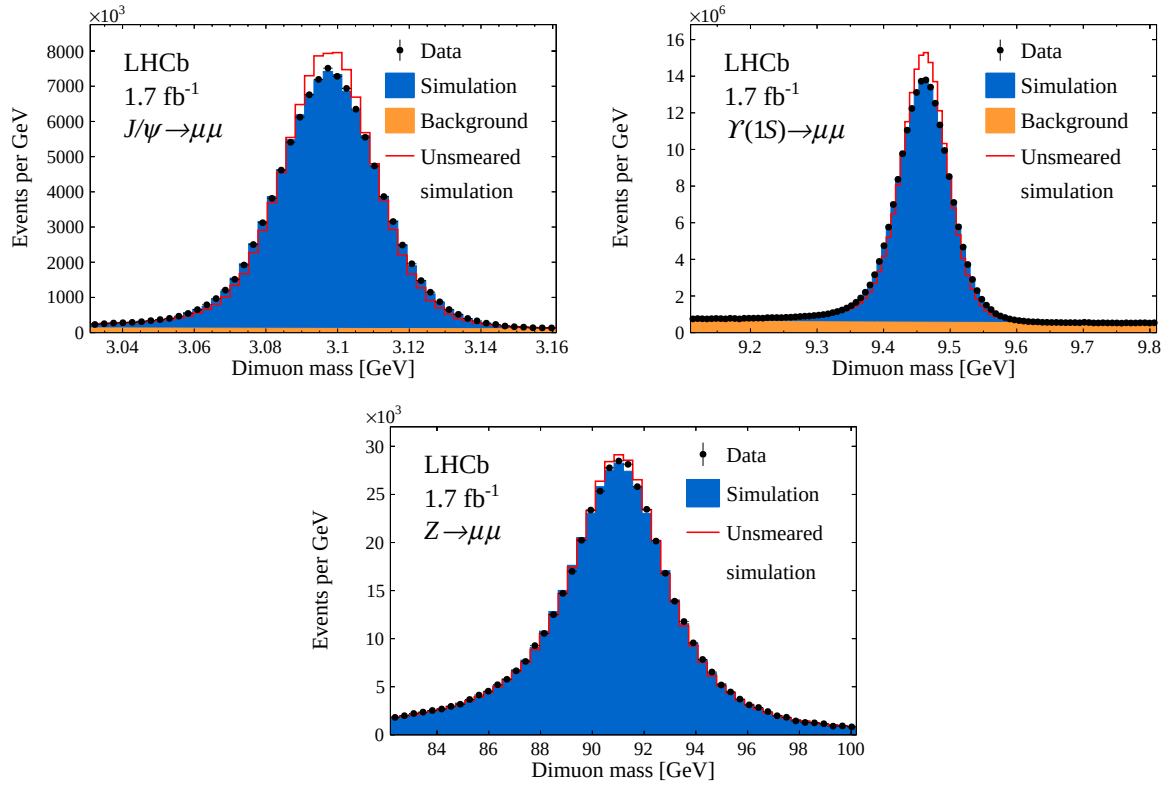


Figure 8.7: Fitted dimuon mass distributions for J/ψ, $\Upsilon(1S)$ and Z boson candidates, combining all η and magnet polarity categories, to determine the smearing parameters to be applied to the simulation. The red histogram indicates the model before the application of the smearing.

Table 8.3: Parameter values (and their errors) of momentum smearing fit as used in the m_W analysis.

Parameter	Value	Error
Momentum scale ($\times 10^{-3}$) $[1.8 < \eta < 2.2]$	0.58	0.10
Momentum-independent smearing ($\times 10^{-3}$)	2.015	0.019
Momentum-dependent smearing ($\times 10^{-6}$) (GeV^{-1}) $[1.8 < \eta < 2.2]$	177	12
Curvature bias ($\times 10^{-6}$) (GeV^{-1})	-0.5	0.4
Momentum scale ($\times 10^{-3}$) $[2.2 < \eta < 4.4]$	-0.005	0.003
Momentum-dependent smearing ($\times 10^{-6}$) (GeV^{-1}) $[2.2 < \eta < 4.4]$	149	9

these variables must be smeared in the simulation to match the data. The smearing fit is described in Section 6.5, and matches the $Z \to \mu\mu$ simulation to data in bins of η and ϕ. The χ^2_{trk} smearing uses the same binning and method. Two examples of the effect of this smearing are shown in Figure 8.8. At this point of the analysis all smearing parameters and corrections have been computed, so the final data processing step - where the smearings and corrections are applied - can take place. In this processing the selection cuts are applied with the "all-but-one-cut" strategy, as introduced in Section 6.1.

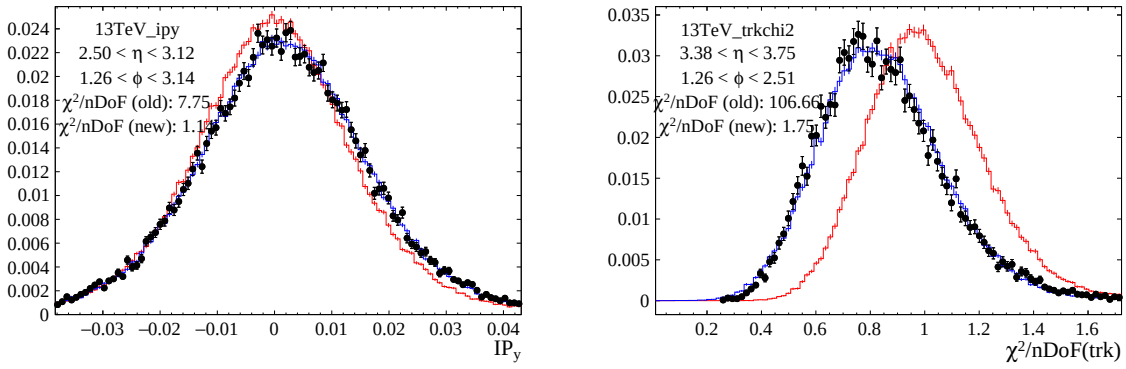


Figure 8.8: Effect of the smearing on the simulation of muon IP_y (left) and muon χ^2_{trk} (right) for two arbitrarily-chosen bins in muon η and ϕ. The black points are $Z \to \mu\mu$ data, and the red (blue) histogram is the simulation before (after) smearing.

The next set of steps are concerned with deriving weight-based corrections to the templates to ensure selection efficiencies are adequately modelled. Muon tracking, trigger and identification efficiency corrections are derived from $Z \to \mu\mu$ (and $\Upsilon(1S) \to \mu\mu$ in the case of the trigger efficiencies) tag-and-probe samples in bins of muon η, ϕ and $1/p_{\mathrm{T}}$ as detailed in Chapter 7. The largest mismodelling and largest variation as a function of $1/p_{\mathrm{T}}$ was found in the trigger efficiencies, which is shown (alongside the fitted parametric forms) in Figure 8.9 for one representative ϕ bin. Muon isolation efficiency scale factors were determined as a function of muon η and recoil projection u, as described in Section 6.6. These binned scale factors are applied as weights to the templates, and the effect of them can be seen as a function of u and $1/p_{\mathrm{T}}$ on $W \to \mu\nu$ simulation in Figure 8.10. Note that, in looking up the signal muon u value to get the appropriate weight, the simulated value is used. This is because u cannot be reconstructed in LHCb for $W \to \mu\nu$ events (in $Z \to \mu\mu$ it can be reconstructed). The efficiencies of the χ^2_{IP} and χ^2_{trk} requirements must also be corrected as described in Section 6.7. The resulting efficiency scale factors $\varepsilon_{\mathrm{D}}/\varepsilon_{\mathrm{MC}}$ were observed to be within 1% of unity, and led to small changes in m_W at the $\mathcal{O}(100\,\mathrm{keV})$ scale when switched on/off.

The final *static* weight (as opposed to the dynamic physics weights) to be applied to the templates is one that ensures they have the same proportion of magnet polarities as the real data. The final correction to the $W \to \mu\nu$ and $Z \to \mu\mu$ templates is a parametric correction to enhance the cross section at high p^V_{T}, as motivated and described in Section 5.2.1. Figure 8.11 shows the p^Z_{T} distribution in data and simulation, and the ratio of the two, the latter fitted with an error function. The undershoot at high p^V_{T} is clearly visible, although the error function only makes any noticeable modification to the $p^Z_{\mathrm{T}} \gtrsim 40\,\mathrm{GeV}$ tail.

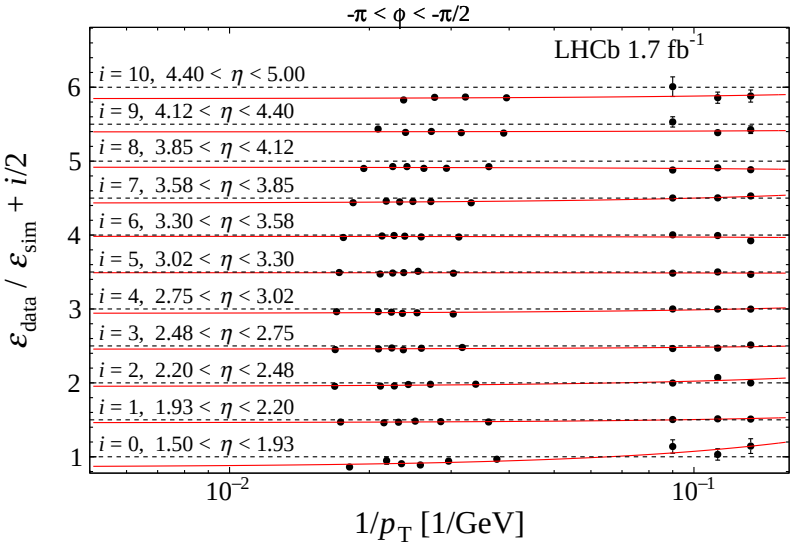


Figure 8.9: Ratios of the trigger efficiency in data to simulation as a function of $1/p_{\mathrm{T}}$, shown for an arbitrary ϕ bin, and all η bins by offsetting the values in each η bin by a constant. The red line is the fitted parametrization in each bin.

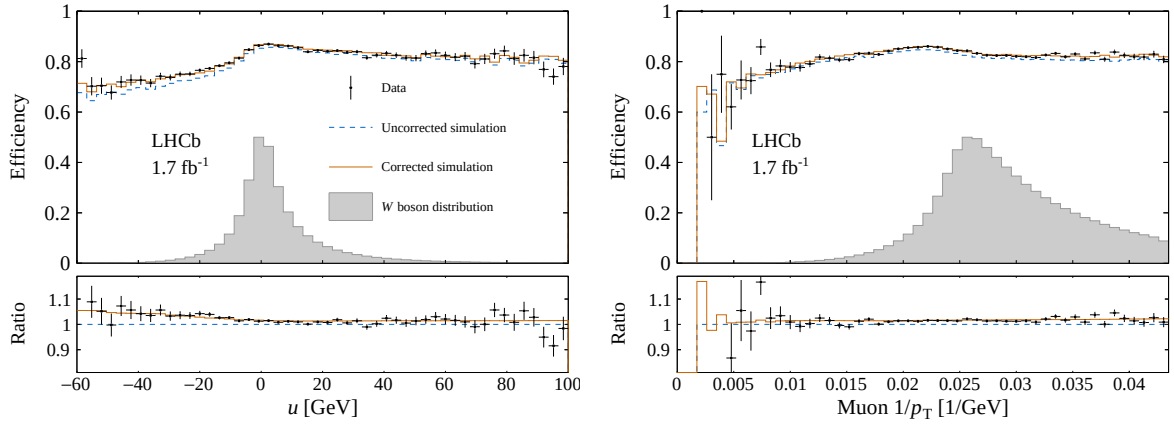


Figure 8.10: Isolation efficiencies as a function of the recoil projection u (left) and muon $1/p_{\mathrm{T}}$ (right). The grey histograms indicate the - arbitrarily normalized - shapes of each distribution in simulated W boson events. In the lower panels the ratios of the isolation efficiency with respect to the uncorrected simulation are shown.

The final step of note before the m_W fit was the preparation of the DIF templates. For this, the `NoMuID` control sample of unidentified muons in data was prepared and fitted in bins of isolation and χ^2_{IP} with a parametric function as detailed in Section 6.8. Figure 8.12 shows the fits to the lowest isolation and lowest χ^2_{IP} bin. The fitted parametric function in this bin was then sampled to generate template events for use in the fitter.

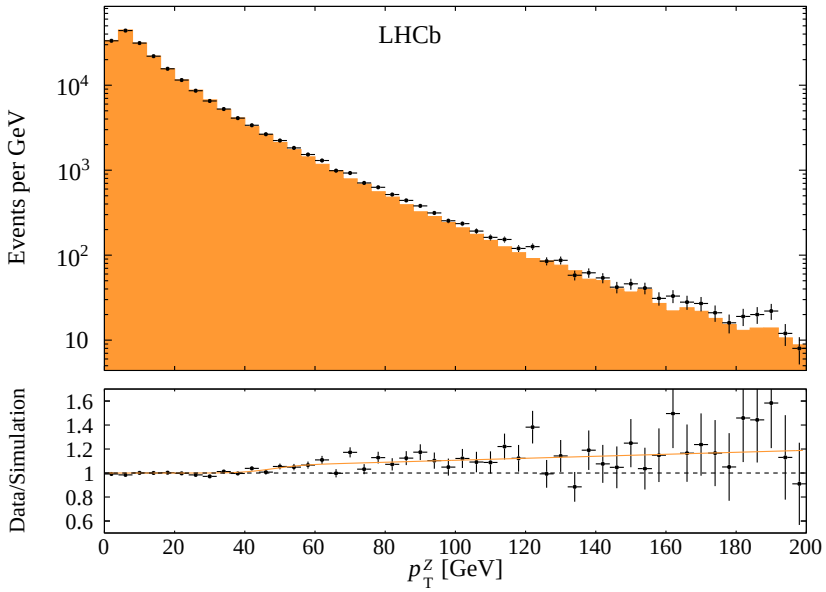


Figure 8.11: The tuned p_{T}^{Z} distribution from POWHEGPYTHIA compared to data. The lower panel shows the ratio overlaid with the parametric weight function.

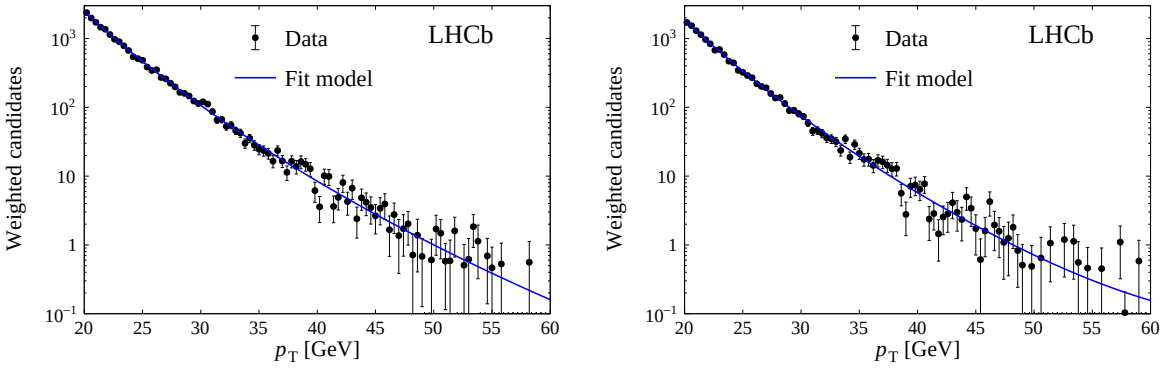


Figure 8.12: Transverse momentum spectra of positively (left) and negatively (right) charged hadrons from the `NoMuID` control sample, weighted to mimic hadronic decays-in-flight. Overlaid are fits with the Hagedorn function, as defined in Eqn. 6.9.

8.6 Fit result

With the templates and data prepared, the m_W fit can proceed. Figure 8.13 shows the data with the central fitted model (NNPDF3.1 PDFs) overlaid. The statistical uncertainty on m_W is 23 MeV. The total fit χ^2 per degree of freedom is 105/102, showing excellent agreement with the data in the fit regions. The modelling is, as expected, poorer outside this region, particular at very-high p_{T} where a number of systematic uncertainties are large. The best-fitting values of the eight free parameters are shown in Table 8.4. The A_3 scaling factor is consistent with 1 at the 1σ uncertainty level, which suggests that

the choice of DYTurbo with its central scales is sufficient to model the data. The value of α_s is noticeably higher than the world average of measurements at the EW scale (0.1179 ± 0.0010 [88]), but α_s is purely a nuisance parameter here, and its floating value (along with $k_{\mathrm{T}}^{\mathrm{intr}}$) is accounting for missing higher-order predictions of QCD. Figure 8.14 shows the model's q/p_{T} and rapidity distributions for the Z sample, which agree well with the data and give confidence in the detector and physics modelling across the two signal samples. A large variety of other post-fit plots were checked regularly as part of the analysis development, and in Figures 8.15, 8.16, 8.17, 8.18 the muon η and ϕ post-fit plots are shown for both the W and Z samples. The good agreement here gives confidence in the detector modelling strategies, particularly of the muon reconstruction and isolation efficiencies.

Table 8.4: Values of the parameters determined in the m_W fit with the NNPDF3.1 PDF set. The uncertainties quoted are statistical.

Parameter	Value
Fraction of $W^+ \to \mu^+\nu$	0.5288 ± 0.0006
Fraction of $W^- \to \mu^-\nu$	0.3508 ± 0.0005
Fraction of hadron background	0.0146 ± 0.0007
α_s^Z	0.1243 ± 0.0004
α_s^W	0.1263 ± 0.0003
$k_{\mathrm{T}}^{\mathrm{intr}}$	$1.57 \pm 0.14\,\mathrm{GeV}$
A_3 scaling	0.975 ± 0.026
m_W	$80362 \pm 23\,\mathrm{MeV}$

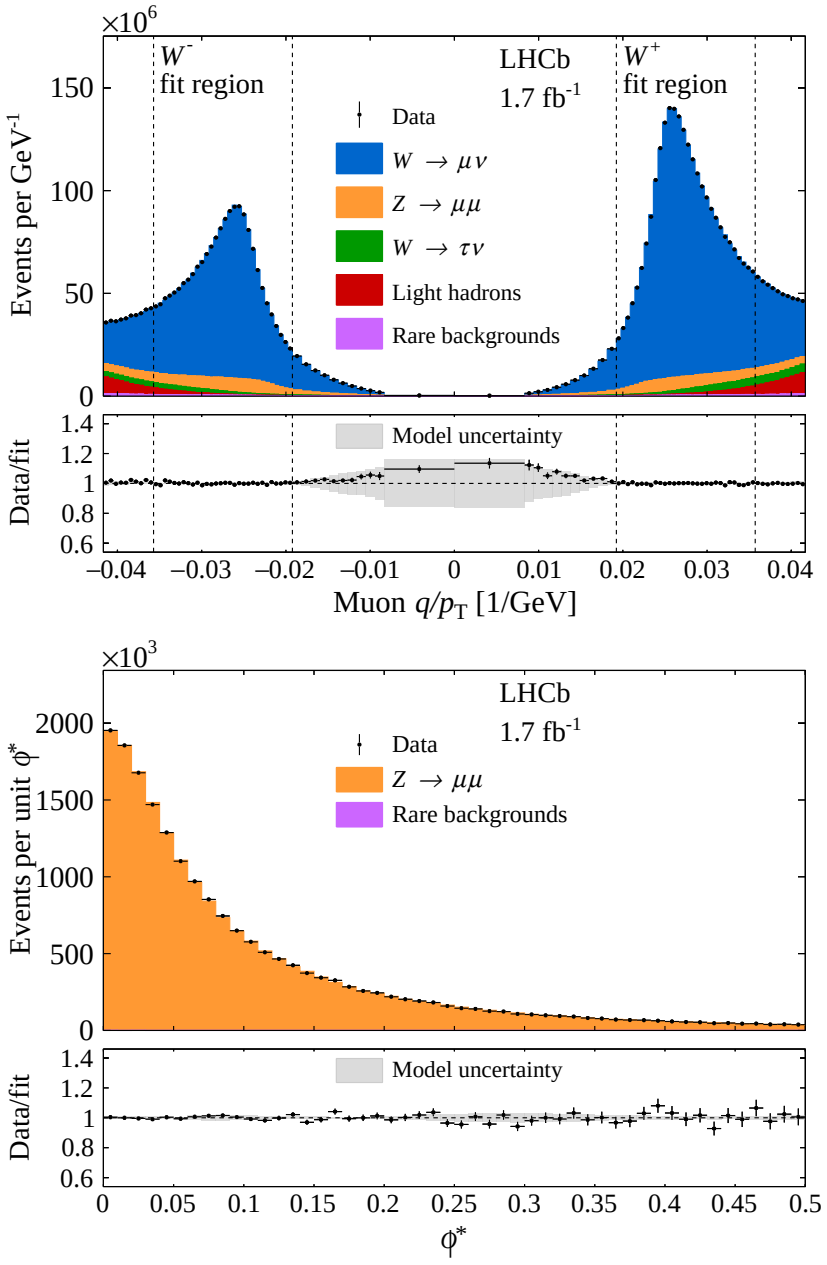


Figure 8.13: Distributions of W q/p_{T} (top) and $Z\phi^*$ (bottom) data distributions compared to the model after the m_W fit.

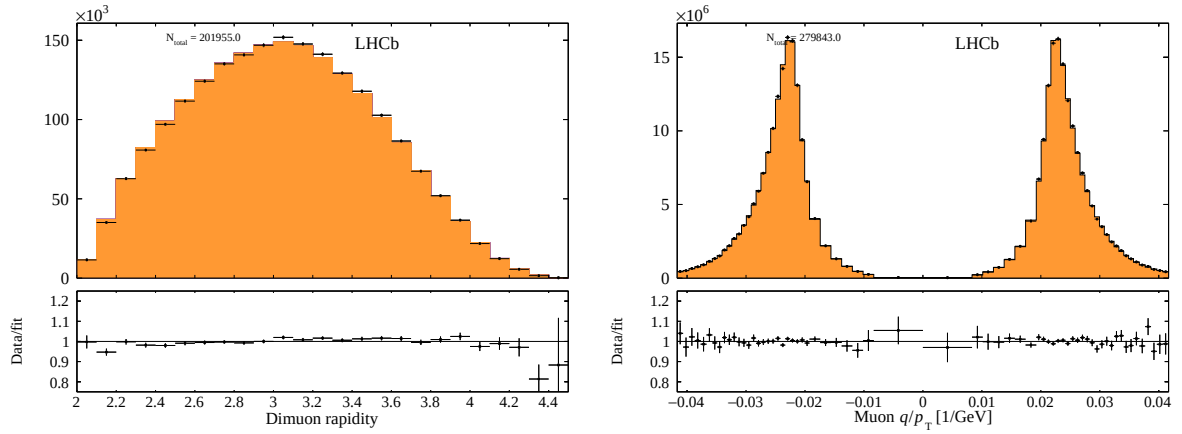


Figure 8.14: Post-fit projection of the model's Z rapidity (left) and q/p_{T} (right) compared to the dimuon data sample.

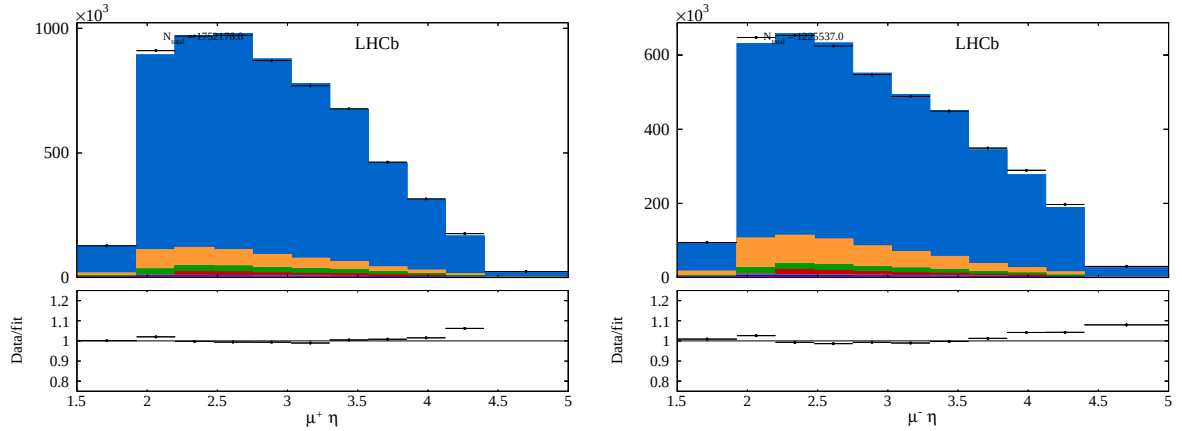


Figure 8.15: Post-fit projection of the model's single muon η distributions compared to the μ^+ (left) and μ^- (right) data samples. The colour scheme is the same as Figure 8.13.

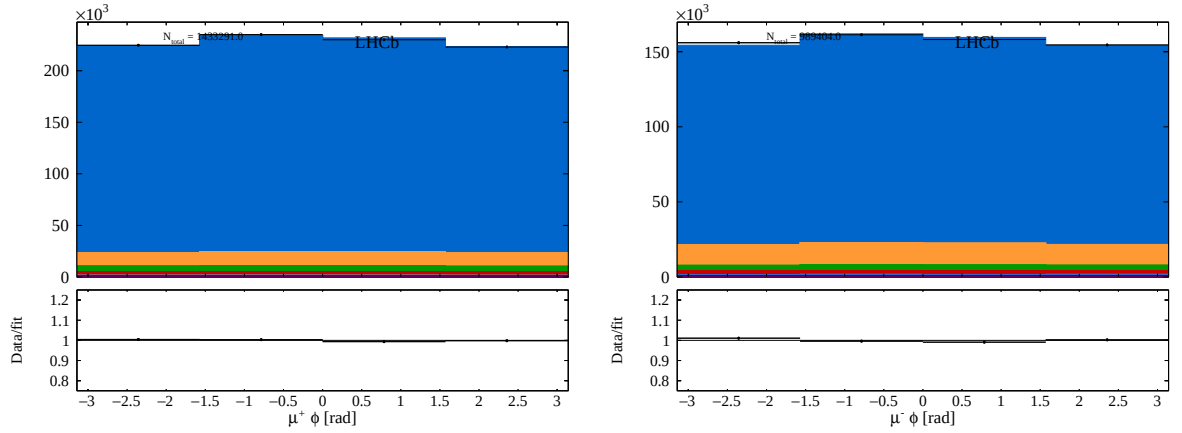


Figure 8.16: Post-fit projection of the model's single muon ϕ distributions compared to the μ^+ (left) and μ^- (right) data samples. The colour scheme is the same as Figure 8.13.

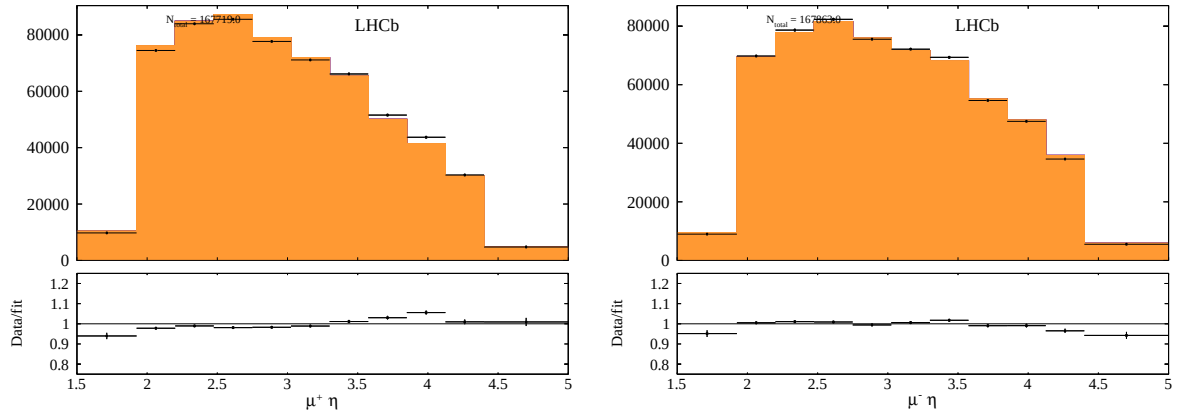


Figure 8.17: Post-fit projection of the model's muon η distributions from $Z \to \mu\mu$ compared to the dimuon data, for μ^+ (left) and μ^- (right).

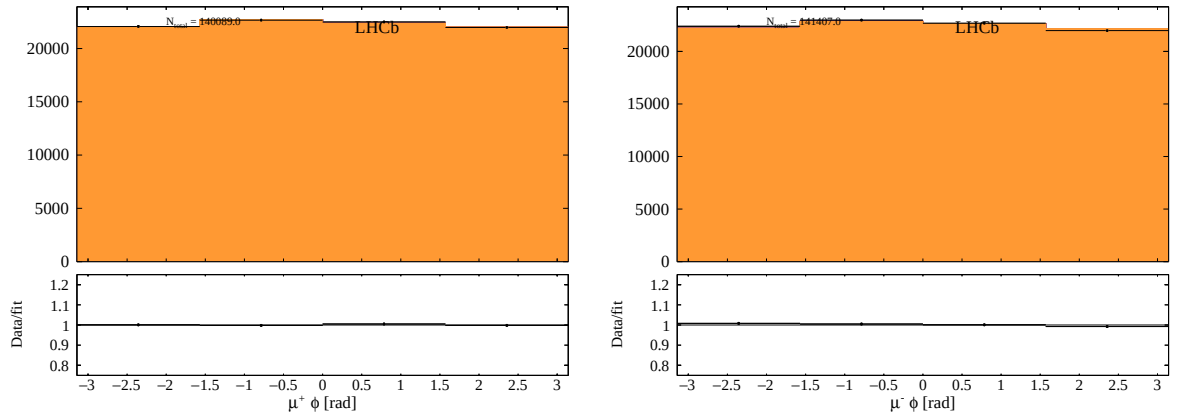


Figure 8.18: Post-fit projection of the model's muon ϕ distributions from $Z \to \mu\mu$ compared to the dimuon data, for μ^+ (left) and μ^- (right).

8.7 Systematic uncertainties

Table 8.5 gives a summary of the relevant systematic uncertainties in this measurement, excluding the PDF uncertainties, which will be discussed at the end of this section. The vast majority of the uncertainties rely on performing several m_W fits, each with a single, different choice in one part of the analysis, and comparing the resulting m_W values to the default value. The final uncertainty from a group of m_W values is then typically calculated by taking either the RMS of the set or the "symmetric envelope" (henceforth "envelope"), which is defined as $\pm\max(|\delta m_{W,i}|)$, where $\delta m_{W,i} = m_{W,\text{default}} - m_{W,i}$. Where other methods are used they will be described in the following.

Table 8.5: Leading systematic uncertainties for this analysis, not including the PDF uncertainty whose value is described in the text. The "TH" and "EXP" types group the uncertainties into those coming from primarily theoretical and primarily experimental sources.

Type	Source	Fits	Δm_W [MeV]
TH	Transverse momentum model	4	11.2
TH	Angular coefficients	31	9.5
TH	QED FSR model	3	7.3
EXP	Momentum smearing details	2	5.0
TH	Missing Higher-order EW corrections	N/A	5.0
EXP	Trigger & Muon ID efficiency	11	3.9
EXP	Trigger & Muon ID control sample size	10	3.9
EXP	Energy offset correction	2	3.1
EXP	Momentum smearing parameter uncertainty	50	2.9
EXP	Isolation efficiency	4	3.1
EXP	Isolation control sample size	9	2.6
EXP	Quarkonia QED FSR	4	2.4
EXP	$\Upsilon(1S)$ mass uncertainty	2	2.0
EXP	Tracking efficiency	12	1.9
EXP	DIF background (p_T range)	8	1.3
EXP	DIF background (p_T and `PIDmu` correlation)	9	1.3
EXP	Tracking control sample size	10	0.8
EXP	DIF background (kaon and pion abundances)	1	0.2
EXP	IP cuts	1	0.1
TH	High p_T^V parametric correction	2	0.0
Theory (excl. PDFs)		40	17.2
Experimental		135	10.3
Statistical		1	22.7

8.7.1 Theoretical uncertainties

The leading systematic uncertainties primarily come from theoretical inputs, which sum in quadrature to 17 MeV. The uncertainty on the transverse momentum model is evaluated by performing fits using "hybrid" configurations of the model, where the y and M shapes of the unpolarized cross section are taken from POWHEGPYTHIA, but the p_{T} shape is taken from a different model. The envelope of the fits using hybrid templates produced with POWHEGHERWIG, HERWIG (both with the NNPDF3.1 PDF set) and PYTHIA with both the CT09MCS and NNPDF3.1 PDF sets gives an 11 MeV uncertainty. DYTURBO was excluded due to its inability to fit the p_{T}^{Z} distribution, as described in Section 5.2. Reassuringly, this uncertainty is similar to the spread of the values obtained in the pseudodata validation tests described in Section 8.3. The angular coefficient uncertainty is evaluated by varying the factorization and renormalization scales in the angular coefficient calculations. The conservative strategy proposed by Ref. [42] was adopted, where the scale choices are varied *incoherently* in the numerator and denominator of the angular coefficients (recall that they are ratios of helicity cross sections). The total uncertainty is taken to be the symmetric envelope of the 31 possible scale variations in this procedure.

The QED FSR uncertainty is taken from an envelope of fits using FSR weights from PYTHIA, HERWIG and PHOTOS. The default fit model is LO in EW emissions in the hard process, so an uncertainty is assigned to cover for missing higher orders with the help of the `POWHEGewPythia` generator, which includes NLO QCD (same as POWHEGPYTHIA) and NLO EW corrections. For technical reasons, a full m_W fit with this generator configuration was not performed, as was the case for the other p_{T} models, but instead the nominal fit model was used to fit pseudodata generated by `POWHEGewPythia` with and without NLO EW corrections. Before this test, it was agreed that if the difference between the two were less than two standard deviations, then the statistical uncertainty on the test would be taken as the related systematic uncertainty. The difference between the two was found to be -4 ± 5 MeV, giving a resulting systematic uncertainty of 5 MeV. The final theoretical uncertainty of significance is that attributed to the largely-aesthetic parametric correction made at high p_{T}^{V}, as described in Section 5.2.1. Two discrete transformations were made to the form of the correction, both of which make differences to m_W of less than 0.1 MeV.

8.7.2 Experimental uncertainties

The leading experimental uncertainties were smaller than their theoretical counterparts, summing in quadrature to 10 MeV. The largest source is the momentum smearing model, which is split into a number of different sources added in quadrature. Firstly, two discrete smearing model changes are made: switching the $1/\cosh\eta$ term for a simple factor of 2.5 for muons in $2.2 < \eta < 3.3$; and shifting the edge at which the multiple-scattering width

is increased (to account for extra material at the edges of the detector) from $\eta = 3.3$ to $\eta = 3.5$. Summing in quadrature, these two contribute a 5 MeV uncertainty. There is a 2.4 MeV uncertainty coming from the envelope of variations to the FSR modelling of the quarkonia, where weights based on an approximate functional form have been applied to increase and decrease the amount of radiative losses for both the J/ψ and $\Upsilon(1S)$ independently. The energy loss due to the amount of material in the detector, and its corresponding effect on the muon 4-momenta, is varied up and down by 10%, leading to a 3 MeV uncertainty. The statistical uncertainty in the momentum smearing parameters is evaluated by sampling the smearing fit's covariance matrix, and the RMS of variations gives a 3 MeV uncertainty. Finally, the external input of the $\Upsilon(1S)$ mass is varied by $\pm 1\sigma$ about the world average, contributing a 2 MeV uncertainty[1]. In total, the uncertainty related to the momentum smearing model is around 7 MeV.

The second-largest experimental uncertainties come from the treatment of the trigger and muon-ID efficiencies. These uncertainties come from the envelope of a series of variations in binnings, tag-and-probe selections and parametrizations used in the efficiency fits. These are described fully (along with the tracking efficiency uncertainties) in Section 7.6. The uncertainty is dominated by the changing of the parametrization, which gives the envelope value of 4 MeV. The statistical component coming from the limited size of the `ZMuID` control sample contributes a further 4 MeV uncertainty. The uncertainties related to the tracking efficiencies are calculated in a very similar way, contributing approximately 2 MeV due to treatment choices (again driven by choice of parametrization) and 1 MeV due to the size of the `ZTrkEff` control sample.

The isolation efficiency treatment also has a statistical component and one based on the method choices. The former is evaluated with Gaussian random variations of the efficiency ratios binned in (η, u) before smoothing, with the RMS of the resulting m_W values taken to compute the uncertainty. For the latter, fits were performed with twice the number of u, and twice and half the number of η bins[2], and a second pass of the smoothing was performed. Both the statistical and choice-based components of this uncertainty contribute around 3 MeV to the total systematic uncertainty.

A variety of choices were made in the preparation of the DIF templates that contribute a small systematic uncertainty to m_W. The upper and lower edges of the p_{T} range of `NoMuID` data used to train the model were varied about the default of $[20, 60]$ GeV, contributing a 1.3 MeV uncertainty to m_W. A further uncertainty is present from the extrapolation from the `NoMuID` control sample to the positively ID-ed DIF muons in the signal sample, which is evaluated using the related continuous `PIDmu` variable by

[1]The J/ψ mass is known much better, and the Z mass is known at a similar precision to the $\Upsilon(1S)$, but this is less important for the Z due to the smaller sample size.
[2]Half the number of u bins was unreasonably coarse.

performing the fits again in different PIDmu values up to zero (above zero significant signal leaks into the control sample). As this does not fully account for the extrapolation, the conservative choice is made to take the maximum variation of m_W across all fits with different PIDmu requirements as the systematic uncertainty, which is a reassuringly small 1.3 MeV. Finally, the $1/p$ weights in the DIF model were adjusted to account for the abundances of particle species as predicted by PYTHIA, which contributed a 0.2 MeV uncertainty. Altogether, the DIF background model contributes just 1.8 MeV of uncertainty to m_W, reflecting its small fit fraction. The final systematic of note is due to the efficiency of the χ^2_{IP} cut, where a m_W fit is performed with the related weights turned off, which causes just a 0.1 MeV deviation in m_W. Since the χ^2_{trk} cut has even less effect on the data and the templates, no uncertainty is assigned.

8.7.3 PDF uncertainties

The PDF uncertainty is evaluated following the recommendation of each PDF group and PDF4LHC [41], with the methods described in Sections 2.4.1 and 5.4. Interpolation histograms were prepared for 100 NNPDF3.1 PDF replicas, and 58 (64) CT18 (MSHT20) PDF eigenvector variations. The α_s-related uncertainty for each PDF group was evaluated by using PDFs with α_s values higher or lower by 0.002. The resulting m_W values are combined as described in Section 2.4.1 via a master formula or by taking the RMS, giving the values summarized in Table 8.6. The total PDF uncertainty from each PDF set is the quadrature sum of the parametric replica/eigenvector uncertainty and the α_s-related uncertainty. In Figure 8.19 the fit χ^2 is plotted against the m_W values corresponding to all the PDF variations considered.

The three PDF sets have been extracted from almost identical datasets, and therefore it was chosen to make no preference between them and take their uncertainties as fully correlated. This means that the resulting PDF uncertainty on the final, PDF-averaged result is just the arithmetic mean of the PDF uncertainties from each group. This gives a total PDF uncertainty on the central m_W result of 9 MeV.

Table 8.6: Uncertainties for the NNPDF3.1, CT18 and MSHT20 sets. The contributions from the PDF uncertainty with fixed α_s and from the α_s variation are quoted separately as is their sum in quadrature, which defines the total uncertainty for each PDF set.

Set	$\sigma_{\text{PDF,base}}$ [MeV]	$\sigma_{\text{PDF},\alpha_s}$ [MeV]	σ_{PDF} [MeV]
NNPDF3.1	8.3	2.4	8.6
CT18	11.5	1.4	11.6
MSHT20	6.5	2.1	6.8

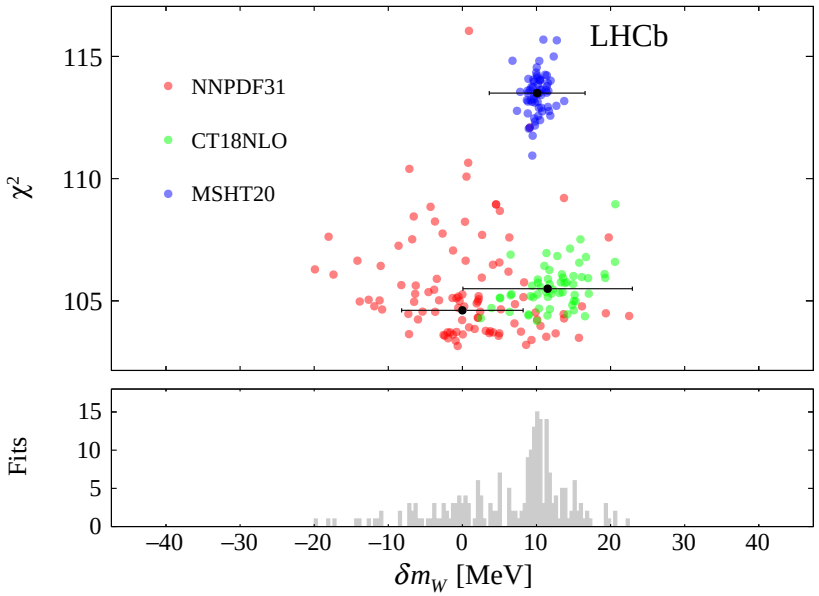


Figure 8.19: The χ^2 versus m_W distribution including all PDF variations considered.

8.8 Cross-checks

A variety of cross-checks are performed to give confidence in the analysis strategy:

- **Splitting of the data into orthogonal subsets**: the data was split by magnet polarity, muon ϕ, muon η and charge $\times$ polarity, and the values of m_W obtained from these discrete subsets were compared for consistency. All differences within the subsets were within or close to two standard deviations, which is considered to be consistent with random statistical fluctuation.
- **W-like m_Z measurement**: The fitter application was configured to perform a "W-like" fit of the Z q/p_{T} distribution to extract the Z mass. The extracted value is consistent with the PDG value of m_Z within one standard deviation, and the values obtained from each muon charge are consistent with one another within one standard deviation.
- **Fit range**: The minimum and maximum p_{T} values in the m_W fit were varied by $\pm 2,4\,\mathrm{GeV}$ either side of their nominal edges at 28 and 52 GeV. The resulting m_W values are consistent with one another within statistical uncertainties.
- **Fit freedom**: The number of nuisance parameters in the nominal (POWHEGPYTHIA) fit configuration is varied, the results of which are shown in Table 8.7. Overall, the fitter is shown to be broadly insensitive to these changes in terms of fit quality and extracted m_W.
- **NNLO PDF sets**: using NNLO, rather than NLO, PDF sets from NNPDF3.1 led to a shift in m_W of 1 MeV,

- **Separate m_W^+ and m_W^- values**: the fitter was configured to fit an additional m_W asymmetry parameter, therefore allowing the mass of W^+ and W^- to differ. This asymmetry parameter was found to be consistent with zero within one standard deviation.

Table 8.7: Fit results with variations in which physics parameters are varying freely.

Configuration change	χ^2_{tot}/ndf	δm_W [MeV]	$\sigma(m_W)$ [MeV]
$2 \to 3$ α_s parameters	103.4/101	-6.0	± 23.1
$2 \to 1$ α_s and $1 \to 2$ $k_{\text{T}}^{\text{intr}}$ parameters	116.1/102	$+13.9$	± 22.4
$1 \to 2$ $k_{\text{T}}^{\text{intr}}$ parameters	104.0/101	$+0.4$	± 22.7
$1 \to 3$ $k_{\text{T}}^{\text{intr}}$ parameters	102.8/100	-2.7	± 22.9
No A_3 scaling	106.0/103	$+4.4$	± 22.2
Varying QCD background asymmetry	103.8/101	-0.7	± 22.7

8.9 Conclusion and outlook

With the systematic uncertainties included, the central results for the NNPDF3.1, CT18 and MSHT20 PDF sets respectively are

$$m_W = 80362 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 9_{\text{PDF}}\,\text{MeV},$$
$$m_W = 80350 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 12_{\text{PDF}}\,\text{MeV},$$
$$m_W = 80351 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 7_{\text{PDF}}\,\text{MeV},$$

where the first uncertainty is statistical, and the others are systematic uncertainties deriving from experimental, theoretical (excluding PDF) and PDF-related sources respectively. None of the three PDF sets were preferred over the others, so it was chosen to take the arithmetic average of the three central values as the central result:

$$m_W = 80354 \pm 23_{\text{stat}} \pm 10_{\text{exp}} \pm 17_{\text{theory}} \pm 9_{\text{PDF}}\,\text{MeV}.$$

This measurement agrees well with the PDG average [88], the SM prediction from the global EW fit [128] and is compatible with previous direct measurements of the W boson mass. This is summarized in Figure 8.20. A total uncertainty of approximately 32 MeV is achieved, despite using only approximately one third of the LHCb Run 2 dataset. This result was published in Ref. [124].

This measurement serves as a proof-of-principle of the strategy to measure m_W with the LHCb detector, and paves the way towards a further measurement using the

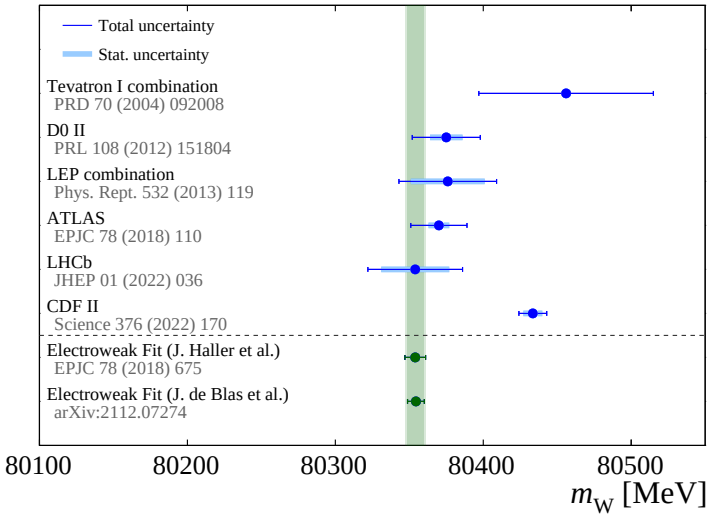


Figure 8.20: Measured value of m_W compared to those from the Tevatron [141–143], LEP [131] and ATLAS [130] experiments. The measurements are listed chronologically, with the latest measurement from CDF coming after publication of the present measurement. The latest predictions of m_W from the global electroweak fit [128, 129] are also included.

entire Run 2 dataset. With approximately 3× more data, a total uncertainty of around 20 MeV is targeted and appears to be achievable given the measurement presented here. At the time of publication, the recent measurement of m_W by the CDF collaboration [143] had not yet been reported, and its large tensions with the predictions of global fits and all other modern direct measurements make further measurements of m_W from the LHC even more important and anticipated. The challenge is to further reduce all the systematic uncertainties to be compatible with the 20 MeV target. Some uncertainties largely scale with the size of the control sample; others arguably have conservative treatments here. Some sources clearly require new ideas and further collaboration, particularly on the theory-uncertainty side. Ref. [133] has shown that the PDF uncertainty can be substantially reduced in two ways: by profiling the PDFs; and by fitting the doubly differential p_{T} and η distributions. Particular effort is needed to reduce the dominating systematic uncertainties in the modelling of the boson production and decay. At the time of writing, strategies are beginning to take shape towards understanding and reducing these key systematic uncertainties.

CHAPTER 9

Test of lepton flavour universality in W-boson decays

This chapter describes the measurement of the ratio $R_{\tau\mu}$, which is a precision test of the Standard Model (SM)'s lepton flavour universality (LFU). At the time of writing, this measurement is not complete, and so this chapter shows its current status. This measurement has been led by the author and constitutes the majority of the author's work towards this PhD. It relies upon the modelling strategies described in the preceding chapters, many of which were developed by collaborators, but the configuration and application of them to the $R_{\tau\mu}$ analysis here is the author's work.

The chapter begins with an introduction to the context and aims of the measurement. Section 9.2 then details the signal selection used to pick out high-p_{T} muons from $W \to \mu\nu$ and $W \to \tau\nu$ decays from the 2016 LHCb dataset. The analysis fitting strategy and description of the purpose-built fitter is presented in Section 9.3, which is followed by its validation studies in Section 9.4. Section 9.5 outlines the specialized impact parameter (IP) modelling strategy developed for this analysis. Section 9.6 then reviews the preparation of the templates for fit, including results of the application of various modelling strategies. The current status of the nominal fit result is shown in Section 9.7. Next is the status of the systematic uncertainties and cross-checks in Sections 9.8 and 9.9 respectively. A summary of the status and outlook of this measurement is given in Section 9.10.

9.1 Introduction

A fundamental axiom of the SM is its lepton flavour universality; that all three families of leptons have identical interactions with the electroweak (EW) gauge bosons. Mathematically, this means that in any SM interaction, if e.g. a muon were to be swapped for a tau lepton, the form of the equations describing that interaction will be the same - one simply needs to swap the μ symbols for a τ. However, this doesn't necessitate that the *value* of

an observable, e.g. a cross section, will be invariant under this swap: the charged leptons have different masses, and so a difference often arises purely from the kinematics of the interaction. In processes such as $W \to \ell\nu_\ell$ at the LHC, where the energy scale is much greater than the rest-mass energy of the lepton masses, this effect is very small. This is the case for the observable of interest, $R_{\tau\mu}$, defined as

$$R_{\tau\mu} \equiv \frac{\mathcal{B}(W \to \tau\nu)}{\mathcal{B}(W \to \mu\nu)}, \tag{9.1}$$

which is precisely predicted to be 0.999 in the SM [144–146], with uncertainties typically at the per-mille level. The differing tau lepton and muon masses ($m_\tau/m_\mu \sim 16$) cause the slight departure from 1.

Broadly-speaking, LFU in the SM is well-verified by experiment. At LEP, measurements of the Z branching fractions to leptons aligned with LFU at per-mille level precision [147]. At low-energy, measurements of the tau lepton lifetime (along with the extremely well-known muon mass and lifetime, and tau lepton mass) give a strong constraint on any hypothetical difference between the W boson's couplings to the tau lepton and muon [9]. However, in high-energy W decays, the historic picture of LFU has been less clear. The first measurements of the W branching fractions were carried out at UA1 and UA2, and supported LFU, albeit at around 25% precision [148]. At LEP, ratios of branching fractions between electrons and muons were consistent with LFU, however the ratio $R_{\tau\mu}$ was measured to be 1.070 ± 0.026 - an intriguing 2.7σ from the SM expectation [131]. Recent measurements of electron/muon ratios in W decays by CDF [149], D0 [150], ATLAS [151] and LHCb [152] have continued to be consistent with LFU at the 1% level, but at the beginning of this study in 2019 there had been no further competitive measurements addressing the so-called "LEP anomaly". In flavour physics, a number of unresolved LFU anomalies have also been observed. The most relevant here are $R(D)$ and $R(D^*)$, which are 1.4σ and 2.8σ from the SM expectation respectively [153], from a combination of results by LHCb [154],Belle [155], and BaBar [156], with the discrepancies in the direction of enhanced $W \to \tau$ couplings. These are shown in Figure 9.1. Larger still are the discrepancies in $b \to s\ell\ell$ transitions, with recent measurements of the LFU ratios R_K and R_{K^*} being 2.1-2.5 [157] and 3.1 [158] standard deviations from SM predictions respectively. Angular analyses in these channels also give large discrepancies [159]. Although the $b \to s\ell\ell$ anomalies are from processes mediated by Z decays (in the SM) and concern electrons and muons, it is interesting to wonder if a broad and connected picture of beyond-the-Standard-Model (BSM) LFU violation is emerging.

Since the beginning of this study, however, the ATLAS and CMS experiments have both produced measurements of $R_{\tau\mu}$ that are in conflict with the LEP measurement and consistent with LFU, as can be seen in Figure 9.2. ATLAS used an innovative

tag-and-probe approach in dileptonic $t\bar{t}$ decays to enable them to analyse muons from $W \to \mu\nu$ and $W \to \tau\nu$ down to $p_{\mathrm{T}} = 5\,\mathrm{GeV}$. With a 2D fit to the p_{T} and "transverse impact parameter" distributions, they measured $R_{\tau\mu} = 0.992 \pm 0.013$ [160]. CMS also used primarily $t\bar{t}$ data with leptonic W decays, simultaneously fitting the p_{T} distributions of many different decay categories (e.g. number and identity of leptons, number of jets etc.) to measure the branching ratios of $W \to \tau\nu$, $W \to \mu\nu$ and $W \to e\nu$, from which they could derive $R_{\tau\mu} = 0.985 \pm 0.020$ [161]. The goal of this study therefore switched to verifying these results at a similar precision. It is also complementary, as it involves purely on-shell W production rather W bosons from the decay of top quarks.

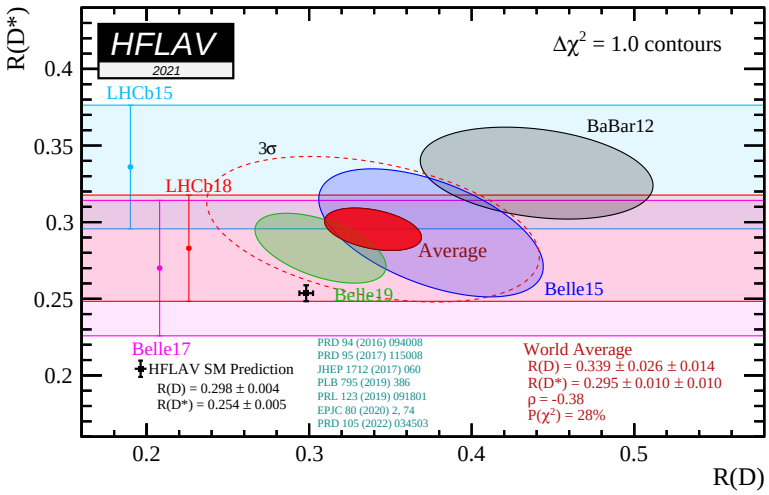


Figure 9.1: Summary (at the time of writing) of the experimental measurements of the LFU ratios $R(D)$ and $R(D^*)$, alongside the SM prediction. Taken from Ref. [153].

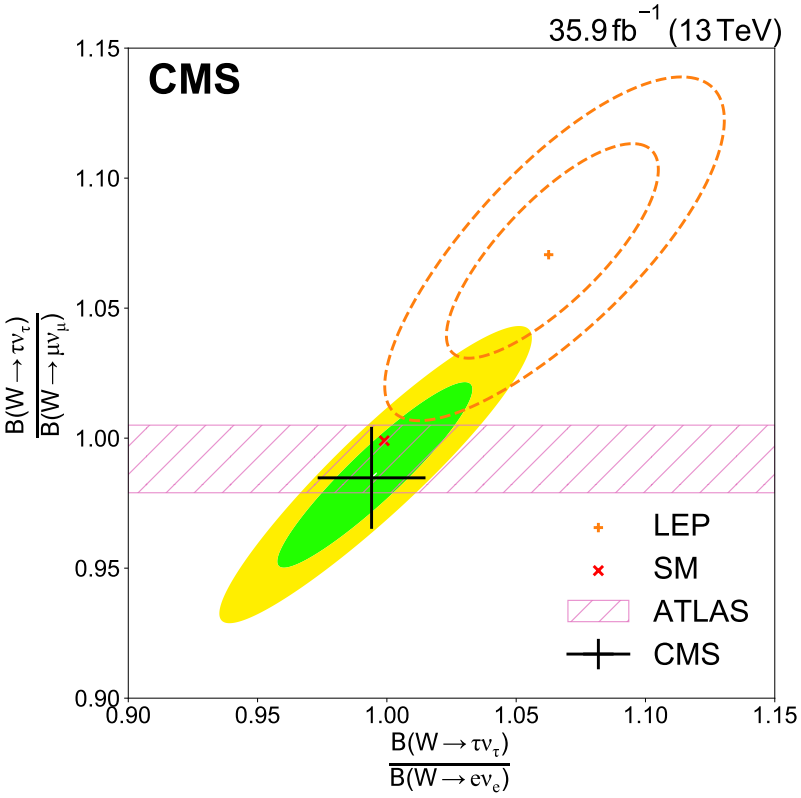


Figure 9.2: Summary (at the time of writing) of the experimental measurements of LFU rations of W branching fractions to leptons, alongside the SM prediction. $R_{\tau\mu}$ is the quantity plotted on the y axis of the figure. Taken from Ref. [161].

9.2 Signal selection

To measure $R_{\tau\mu}$, a data sample of single, high-p_{T} muon decays is required. The selection requirements for this sample are listed in Table 9.1. The key differences in the selection with respect to the m_W measurement are the removal of isolation and impact parameter requirements, the motivation for which will become clear shortly. The same `Z` sample, as described in Section 6.1, is heavily used as a control sample.

The Z-veto cut for rejecting $Z \to \mu\mu$ was described in Section 8.4. The η cut is tighter than m_W to maximize the charge asymmetries[1] of the signal components in the hope of making them more distinct from the broadly charge-symmetric backgrounds, and to keep the isolation cone further away from the edges of the detector. The cut on the number of tracks coming from the best primary vertex (BPV), or PV track multiplicity $N_{\mathrm{tracks}}^{\mathrm{BPV}}$, rejects events where the PV and therefore IP is poorly reconstructed. The other

[1]The charge asymmetry of W production changes sign at high η, being positive at lower values due mainly to there being two up quarks in the proton and one down quark, while the $V-A$ structure of the weak interaction changes this at high η.

Table 9.1: The $R_{\tau\mu}$ selection cuts applied to 2016 LHCb data and simulation. "TOS" stands for "trigger-on-signal", meaning that the signal muon was responsible for firing the trigger lines. The trigger lines are as described in Section 3.5: `L0MuonEW`, `Hlt1SingleMuonHighPt` and `Hlt2SingleMuonHighPt`.

Description	Requirement
Z-veto	$p'_{\mathrm{T}}\,(2.0 < \eta' < 4.5) < 25$ GeV
Pile-up	Number of candidates in event = 1
Momentum error	$\Delta p/p < 0.06$
Sanity cut	$p < 2\,\mathrm{TeV}$
Trigger	Muon is TOS at L0, HLT1 and HLT2
η acceptance	$2.2 < \eta < 3.7$
Track quality	$\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 1.3$
HCAL activity	$E_{\mathrm{T}}^{\mathrm{HCAL}} < 5.0\,\mathrm{GeV}$
Muon PID	Muon is positively identified
PV track multiplicity	$N_{\mathrm{tracks}}^{\mathrm{BPV}} > 10$

cuts, as with m_W, reject small amounts of poorly-reconstructed events.

9.3 Analysis fitting strategy

At the most basic level, measuring $R_{\tau\mu}$ amounts to counting the number of $W \to \tau\nu$ and $W \to \mu\nu$ decays from a dataset of events where a W boson is produced. Tau leptons primarily decay hadronically, however many other physics processes can give similar final states in LHCb. It is therefore chosen to analyse tau leptons decaying to a high-p_{T} muon and two neutrinos, which provide a much cleaner event signature. LHCb does not reconstruct neutrinos, so ostensibly $W \to (\tau \to \mu\nu\nu)\,\nu$ and $W \to \mu\nu$ have the same final state, however they can be separated as follows. Firstly, muons from $W \to (\tau \to \mu\nu\nu)\,\nu$ decays have a lower p_{T} on average, due to the momentum of the tau lepton being shared between the muon and the two neutrinos, while in $W \to \mu\nu$ the muon has the peaking structure at $p_{\mathrm{T}} \approx m_W/2$ that is familiar from earlier chapters. Furthermore, at these energies in LHCb, tau leptons typically fly $\mathcal{O}(1\,\mathrm{cm})$ before decaying, and so the muon produced has a typical IP of a few tens of microns. Muons from $W \to \mu\nu$ have a true IP of zero. The two signal channels are therefore referred to as "displaced" and "prompt" muons.

There are a number of relevant and non-negligible background processes in this dataset. Like m_W, light hadrons decaying-in-flight to muons (DIF) are important and must be modelled. In the m_W measurement, a tight cut on χ^2_{IP} was applied to remove displaced muons, both from $W \to \tau\nu$ and heavy-flavour (HF) decays. This is obviously

not desirable here, and so HF backgrounds with a single muon become very significant. The LHCb IP resolution, as shown in Section 3.6, plays a large part here in broadening the peaks in IP, causing significant overlap between all the components of interest. Fortunately, muons from HF and DIF are typically less isolated than prompt and displaced signal muons. A tight isolation cut - as used in the m_W analysis - therefore seems motivated, although during the development of this analysis it was found that the high-isolation region gives a strong constraint on the yields of the HF and DIF yields in the low-isolation region used to fit the signals. Without this, the uncertainty propagated to $R_{\tau\mu}$ associated with these yields is intolerably large. It was thus decided to consider the isolation as a third fit variable.

For the reasons outlined above, the measurement of $R_{\tau\mu}$ is a template fit of the single, high-p_{T} muon data, binned in muon $1/p_{\mathrm{T}}$, IP and isolation. The isolation variable used here is the vector sum of the p_{T} of charged particles in a cone of radius $\Delta R = 0.5$, plus the vector sum of the p_{T} of neutral particles in the same cone, and is denoted $\mathcal{I}$. This is a slightly different isolation variable to that used in the m_W analysis, as it was found to be slightly better modelled in the simulation than its particle-flow counterpart. The simulated shape of the signals and the most important backgrounds at the reco-level

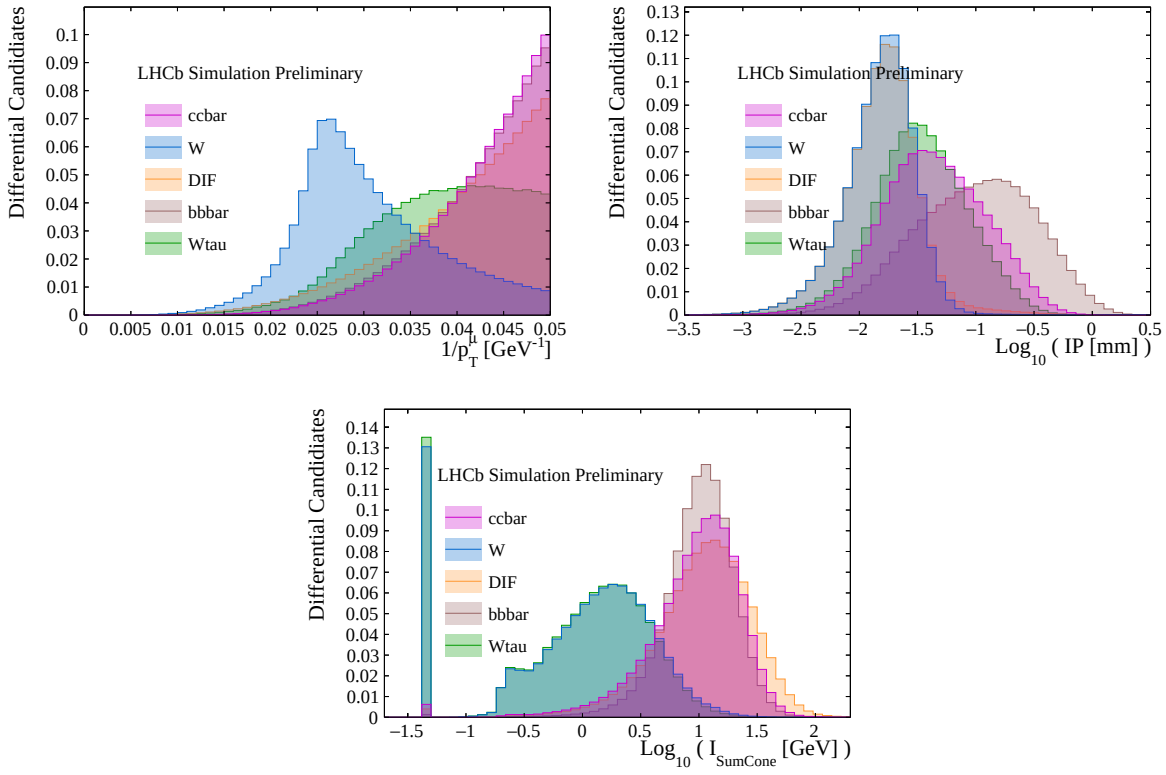


Figure 9.3: Simulated (normalized) shapes of the muon $1/p_{\mathrm{T}}$ (top left), $\log_{10}$(IP) (top right) and $\log_{10}(\mathcal{I})$ (bottom) for $W \to \tau\nu$, $W \to \mu\nu$, DIF, $c\bar{c}$ and $b\bar{b}$ decays. Positive muons are used here; the picture is qualitatively the same for negative muons.

and shown in Figure 9.3, which illustrate the conclusions made above. Only the 2016 data is used to make full use of the software and studies made for the m_W measurement.

The template fit measures the yields of positively- and negatively-charged prompt and displaced muons in data, denoted by $N_{\mu^\pm}$ and $N_{\tau^\pm}$ respectively. It is assumed that the charge asymmetry is the same for prompt and displaced muons, since it derives from the W production, so there are therefore three degrees of freedom. It is chosen to associate these to $R_{\tau\mu}$ and the two *effective* $W \to \mu\nu$ fiducial cross sections $\sigma'(W^\pm \to \mu^\pm\nu_\mu)$. These cross sections are labelled as such because the sample's integrated luminosity is not well-known, and the signal yields are proportional to both the cross sections and the luminosity, so the cross sections absorb the uncertainty in the luminosity by floating freely. As such they cannot be expected to be accurate measurements of the true fiducial cross sections. Denoting the poorly-known luminosity as $\mathcal{L}_{\text{dummy}}$, these four signal yields can be connected back to $R_{\tau\mu}$ and the two effective cross sections as

$$N_{\mu^\pm} = \mathcal{L}_{\text{dummy}} \cdot \sigma'(W^\pm \to \mu^\pm\nu_\mu) \cdot \frac{\sum w(\mu^\pm)}{N^{\text{gen,fid}}(\mu^\pm)}\,, \tag{9.2}$$

$$N_{\tau^\pm} = \mathcal{L}_{\text{dummy}} \cdot R_{\tau\mu} \cdot \sigma'(W^\pm \to \mu^\pm\nu_\mu) \cdot \mathcal{B}\left(\tau \to \mu\nu\nu\right) \cdot \frac{\sum w(\tau^\pm)}{N^{\text{gen,fid}}(\tau^\pm)}\,, \tag{9.3}$$

where the final fractions are the reconstruction and selection efficiencies of each process. The dummy luminosity is fixed to a semi-arbitrary $1\,\text{fb}^{-1}$, which - along with the effective cross sections - cancels in the $R_{\tau\mu}$ ratio. The value of the branching ratio $\mathcal{B}(\tau \to \mu\nu\nu) = 0.1739$ is taken from the 2022 version of the PDG particle listings [162].

The total fit model is a sum of templates coming from each signal and background process. The following backgrounds are included: $Z \to \mu\mu$; $Z \to \tau\tau$; DIF; HF decays (split into $c\bar{c}$ and $b\bar{b}$ production); low-mass Drell-Yan; top quarks and WW pairs. The so-called "EW backgrounds" (all except $c\bar{c}$, $b\bar{b}$ and DIF) have their yields $N^{\text{W}}_{\text{comp}}$ constrained relative to the observed number of $Z \to \mu\mu$ candidates $N^{\text{Z}}_{\text{obs}}$ in a manner similar to Equation 8.10, except there is only a W "term" here:

$$N^{\text{W}}_{\text{comp}} = \sum w^{\text{W}}_{\text{comp}} \times \frac{N^{\text{Z}}_{\text{obs}}}{\sum w^{\text{Z}}_{\text{Z}}} \times \frac{\sigma^{\text{fid}}_{\text{comp}}}{\sigma^{\text{fid}}_{\text{Z}}} \times \frac{N^{\text{gen,fid}}_{\text{Z}}}{N^{\text{gen,fid}}_{\text{comp}}}, \tag{9.4}$$

where the superscripts refer to the selection (either `W`, referring to Table 9.1; or `Z`, referring to Table 6.1) and the subscripts refer to the physics process. The symbols here are otherwise the same as Equation 8.10, as are the fiducial cross sections and fiducial requirements used. This equation effectively scales the simulated samples to have the same integrated luminosity as the data, as determined from the pure $Z \to \mu\mu$ data sample. The starting values of the $W \to \mu\nu$ and $W \to \tau\nu$ yields are also constructed using this

equation and the LFU assumption of $R_{\tau\mu} = 1$.

9.3.1 Nuisance parameters in the fit

The $c\bar{c}$ and $b\bar{b}$ inclusive cross sections with a single, high-p_{T} muon in the final state at LHCb have not been measured, meaning it would be difficult to sufficiently validate any event-generator prediction of their cross sections. The situation for DIF is similar. The yields of these components are therefore allowed to float freely in the fit for each charge, amounting to six freely-floating yields (three components, two charges). The $c\bar{c}$ and $b\bar{b}$ yields start from a predicted value, while the DIF yields are initially set to the remaining data yield once all other components have been predicted.

The final floating parameters in the $R_{\tau\mu}$ fit are assigned to cover for the reasonable case that the simulated p_{T} and isolation shapes of the $b\bar{b}$, $c\bar{c}$ and DIF templates are incorrect; the full LHCb simulation was used to make these templates, and robust validation studies of them have not been made. The IP shape is expected to be better-controlled by the dedicated IP smearing model, and mostly dependent on the underlying well-understood physics. The $R_{\tau\mu}$ fitter therefore has the capacity to float form factors which morph the p_{T} and isolation shapes on-the-fly in the fit. These form factors rely on the assumption that the p_{T} (isolation) shapes of these backgrounds can be adequately described by a parametric function within the fit region. Figure 9.4 shows the comparison of each background's simulated $1/p_{\mathrm{T}}$ and isolation shape, overlaid with the fitted function. The function chosen to model the p_{T} (isolation) shape is a Hagedorn (Crystal-Ball) function. A reasonable agreement between the parametric models and the simulated shapes is observed within the fit region. Using the set of fitted values $\overrightarrow{\alpha_{\mathrm{def\overline{l}}}}$ from these fits to the default simulated shapes, a form factor can be constructed from a ratio of Hagedorn functions $f\left(\overrightarrow{\alpha}\right)$ that morphs the simulated shape to different parameter values of the underlying function. For each bin (indexed with subscript i) with lower (upper) edges at x_i (x_{i+1}), the bin content N_i is morphed as

$$N_i \to N_i' = \frac{N_i}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} \frac{f\left(\overrightarrow{\alpha_{\mathrm{new}}}\right)}{f\left(\overrightarrow{\alpha_{\mathrm{def\overline{l}}}}\right)}\,, \tag{9.5}$$

i.e. the original bin content is multiplied by the average of the form factor over that bin, and the form factor is a ratio of the parametric function with the new values in the numerator, and the default in the denominator. Each form factor acts only on one axis of the three-dimensional templates: for example, for the k-th isolation bin, all (i, j) $1/p_{\mathrm{T}}$ and IP bins will be morphed according to the same value of the form factor. Once all bins have been morphed in this way, the template is re-normalized, so the morphing only affects the shape.

Giving form factors for the isolation and p_{T} shape for all six fit components (two

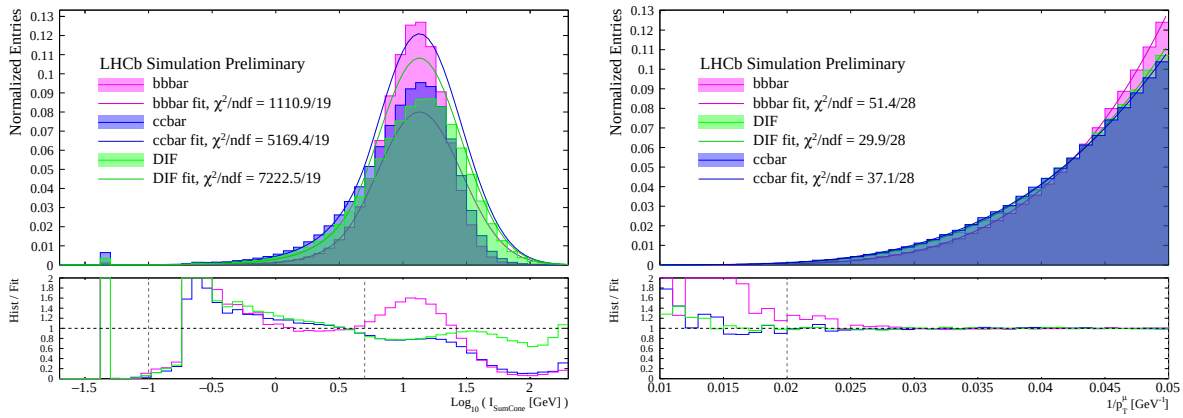


Figure 9.4: Fits to the (left) isolation and (right) $1/p_{\mathrm{T}}$ shapes of the various simulated background samples with the form factor numerator functions (the Crystal-Ball and Hagedorn functions, respectively). The dashed vertical lines indicate the fit range, which corresponds also to the limits of the $R_{\tau\mu}$ fit. In the case of isolation, an arbitrary low limit is placed to avoid the "zero" bin.

charges and three physics processes) would be far too much freedom for the fitter to reliably handle. It is therefore assumed that any p_{T} mismodelling is charge-independent, and is the same for the two HF species, meaning one Hagedorn-based form factor shared between both charges of $c\bar{c}$ and $b\bar{b}$. Similarly, one shared p_{T} form factor is floated for DIF. The two free parameters in the Hagedorn (see Eqn. 6.9) are very correlated, and it is therefore difficult to float both, so the function's $p_{\mathrm{T},0}$ parameter is fixed to a rough average of best-fitting values of $p_{\mathrm{T},0}$ to the DIF, $c\bar{c}$ and $b\bar{b}$ as displayed in Figure 9.4. The value choice is varied as a systematic uncertainty. Altogether, this amounts to two floating p_{T} form factor parameters in the $R_{\tau\mu}$ fit. Similar arguments can be applied to the isolation form factors, and the key Crystal-Ball parameter here is the tail parameter α, since it is only the tail of the Crystal-Ball shape that lies in the fit region. The other three Crystal-Ball parameters are fixed and also treated as potential systematic uncertainty sources. This means one isolation form factor parameter floating for HF, and one for DIF.

Altogether, there are 13 floating parameters in the default configuration of the $R_{\tau\mu}$ fit. These are, for clarity and completeness:

- $R_{\tau\mu}$,
- $\sigma'(W^+ \to \mu^+\nu)$,
- $\sigma'(W^- \to \mu^-\nu)$,
- $N\left(c\bar{c} \to \mu^+ + X\right)$,
- $N\left(c\bar{c} \to \mu^- + X\right)$,
- $N\left(b\bar{b} \to \mu^+ + X\right)$,
- $N\left(b\bar{b} \to \mu^- + X\right)$,
- $N\left(\mathrm{DIF} \to \mu^+ + X\right)$,

- $N\,(\mathrm{DIF} \to \mu^- + X)$,
- HF p_T form factor parameter (Hagedorn exponent β_HF),
- DIF p_T form factor parameter (Hagedorn exponent β_DIF),
- HF isolation form factor parameter (Crystal-Ball α_HF),
- DIF isolation form factor parameter (Crystal-Ball α_DIF).

In Section 9.9 a set of $R_{\tau\mu}$ fits is performed with variations in the number of free nuisance parameters to cross-check that there is no large effect on $R_{\tau\mu}$ coming from this choice. The fit is achieved by minimization (using MINUIT [139]) of a binned χ^2. The Beeston-Barlow Lite prescription [140] is used to account for the finite size of the simulated samples.

9.3.2 Blinding

The value of $R_{\tau\mu}$ has been (and currently is) blinded while the analysis is under development. Blinding is achieved by drawing a deterministic random number from a uniform distribution in the interval $[-0.1, 0.1]$ which is added to $R_{\tau\mu}$ in each iteration of the fitter. MINUIT therefore floats the true value of $R_{\tau\mu}$, but unbeknownst to it the value is shifted by the time the χ^2 is evaluated, meaning that the true value of $R_{\tau\mu}$ is in fact the fitted value minus the offset. The blinding offset will be removed once approval of the analysis has been given by the review committee assigned by LHCb. All fit results shown in the remainder of this thesis are currently blinded by the same random offset, meaning the quantitative effect on $R_{\tau\mu}$ of any changes can be seen.

9.3.3 Fit binning

The default fit configuration has the following binning: ten bins in $1/p_\mathrm{T}$, linearly-spaced in $[1/50, 1/22]$ GeV^{-1}; three bins in IP with edges at $[0, 80, 180, 300]$ μm; and three bins in isolation with edges at $[0, 1.5, 3.0, 5.0]$ GeV.

For ease of plotting, and also to minimize any dependence in the fitter's code design on the number of dimensions in the fit[2], the 3D templates are "flattened" to 1D histograms as the final step before they are passed into the fitting code. An illustration of this is given in Figure 9.5, which shows the flattened pre-fit model (the sum of all the flattened templates constructed with the starting values of the floating parameters). Some regions of the $(1/p_\mathrm{T}, \mathrm{IP}, \mathcal{I})$ space are much more sparsely populated than others, so at the plotting stage some regions are scaled up (by the factors written at the top of the plot) for ease of visualization. Moving from left to right, the first steps are through the ten $1/p_\mathrm{T}$ bins in the lowest IP and lowest isolation slice. At the dot-dashed line, the IP

[2]Fitting in 1D or 2D is still technically possible, and this flexibility was built-in when the fitter was upgraded from an earlier version that was 2D only (q/p_T and IP) to the current 3D binning.

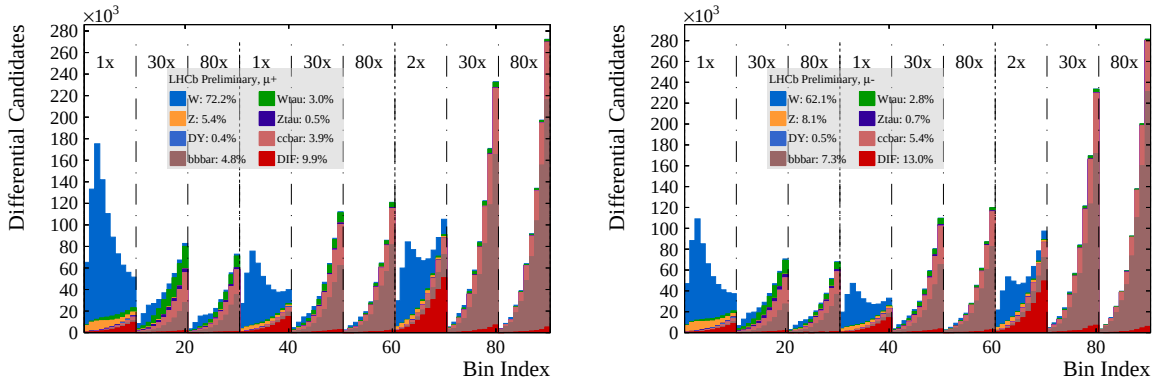


Figure 9.5: Pre-fit $R_{\tau\mu}$ fit model for positively-charged (left) and negatively-charged (right) muons. The dot-dashed lines delineate the different IP bins, and the double-dashed line delineates the isolation bins. The numbers at the top of the plot are the scale factors applied to the templates in that $(\mathrm{IP}, \mathcal{I})$ bin to give visualizable bin populations in all bins.

bin increments by one, and the ten $1/p_{\mathrm{T}}$ bins in the second IP slice (still in the lowest isolation slice) are stepped through. In this way the IP bins are stepped through, before the isolation bin increments by one at the double-dashed line, and again the $1/p_{\mathrm{T}}$ and IP start incrementing again. Mathematically, the flattened bin index is therefore given by

$$\text{Bin Index} = 1 + \left(i_{1/p_{\mathrm{T}}} - 1\right) + N_{1/p_{\mathrm{T}}}\left(j_{\mathrm{IP}} - 1\right) + N_{1/p_{\mathrm{T}}} N_{\mathrm{IP}}\left(k_{\mathrm{iso}} - 1\right) , \tag{9.6}$$

where $\left(i_{1/p_{\mathrm{T}}}, j_{\mathrm{IP}}, k_{\mathrm{iso}}\right)$ are the 3D bin indices for each axis, and $N_{1/p_{\mathrm{T}}}$ and N_{IP} are the number of $1/p_{\mathrm{T}}$ and IP bins respectively.

9.4 Testing of the fitter

It is imperative to ensure that the custom-built fitter measures what it is designed to measure: it must not have any bias on the floating parameters it extracts; and it should have correct *coverage* i.e. that the parameter uncertainties it quotes are indeed the average one-sigma variations in the extracted value. The first test of both these properties is a coverage test, for which toy datasets are constructed, which look approximately like the real data under the SM assumption of $R_{\tau\mu} = 1$. This toy data is generated by constructing the total pre-fit model and then varying each bin's content randomly to mimic Poissonian statistical variations. Since this toy dataset was prepared with $R_{\tau\mu} = 1$, a fit to it with the model should return $R_{\tau\mu} = 1$, within a statistical uncertainty commensurate with the size of the statistical fluctuations applied via the Poissonian variations. Generating many such toy datasets with different random number sequences and performing many such toy fits should therefore yield a Gaussian distribution of $R_{\tau\mu}$ values, with a mean $R_{\tau\mu}$

($\overline{R_{\tau\mu}}$) value of 1. The *pull* distribution, which is $\left(R_{\tau\mu,i} - \overline{R_{\tau\mu}}\right)/\delta R_{\tau\mu,i}$ where $R_{\tau\mu,i}$ and $\delta R_{\tau\mu,i}$ are respectively the fitted values and uncertainties from each toy fit, should also be Gaussian with a mean of 0 and a standard deviation of 1. A non-zero mean would be an indicator of bias in the fitter, and a standard deviation below (above) 1 would indicate an under- (over-)coverage of the uncertainty. The $R_{\tau\mu}$ values and their pulls from 100 toy experiments are shown in Figure 9.6. Within the uncertainty on the parameters relating to a finite number of toy fits, the correct coverage is observed and there is no evidence of bias on the fitted value.

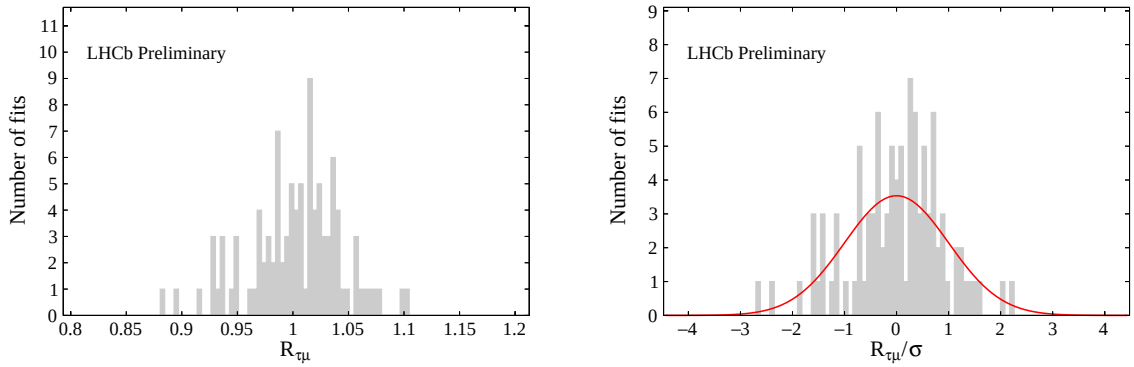


Figure 9.6: Distribution of $R_{\tau\mu}$ results (left) and pulls (right) for 100 fits to toy datasets to test the statistical coverage of the fit.

The coverage test does not rule out all sources of potential bias. For example, if there is a mistake in the equations connecting the floating parameters to the template yields, this will not show up in the coverage test as both the toy data and fit model have been constructed using those same equations. A *closure test* helps to verify that no such biases are present. This test involves fitting a *pseudodata* template, generated as follows (with the SM assumption of $R_{\tau\mu} = 1$). Firstly, all the simulated samples are split in two. One half of each sample is used to make the templates in the usual manner; the other half to construct the pseudodata. The overall size of the pseudodata sample is chosen to have the same integrated luminosity as the data, which can be approximately measured by counting the number of events in the extremely-pure Z sample and with knowledge of measured $Z \to \mu\mu$ cross section at $\sqrt{s} = 13\,\text{TeV}$. Using this integrated luminosity, predictions of the SM cross sections, and the reconstruction-and-selection efficiency for each process, the remaining one-half of each simulated sample can be randomly pre-scaled such that the relative yields of each process are as expected by their relative cross sections. Adding them all together after this pre-scaling produces a "pseudo" dataset. The partitioning of the samples, and pseudodata creation is done in an independent software process *before* the template preparation and fitting code, making this test sensitive to errors in the latter code. The pseudodata and the remaining one-half of the simulated samples are then

passed into the template preparation and fitting code as if they were regular data and the usual simulated samples respectively, and a single $R_{\tau\mu}$ fit is conducted, which should return $R_{\tau\mu} = 1$ within the statistical uncertainty of the fit. The closure test fit projection is shown in Figure 9.7. The fit quality is good ($\chi^2/\text{ndf} = 167/167$) and the fitted $R_{\tau\mu}$ is 0.950 ± 0.038, which is again indicative of no bias in the fitter.

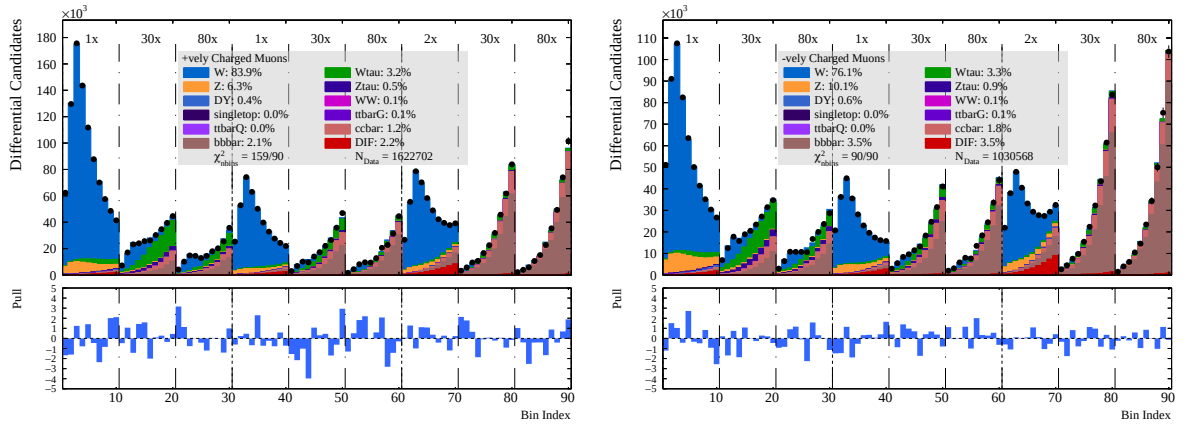


Figure 9.7: Post-fit projection of the fit model to the pseudodata (the black points) to test the closure of the $R_{\tau\mu}$ fitter. Positively- (negatively-)charged muons are shown on the left (right).

9.5 Impact parameter modelling strategy

In the m_W measurement, only a comparatively rough understanding of the LHCb's IP resolution was necessary, as this variable was only used for a cut on χ^2_{IP} to remove displaced muons. Here however, far greater control is needed to be able to reliably *fit* the IP distribution. The IP modelling strategy for the $R_{\tau\mu}$ measurement has three main steps: a weight-based correction of the simulated PV track multiplicity distribution; smearing of IP_x and IP_y (reweighted in $N^{\text{BPV}}_{\text{tracks}}$) by an extension of the smearing tool developed for the m_W measurement (described in Section 6.5); and correction of residual mismodelling at very high IP with weights. As described previously in Section 3.6, the IP used here is as the length of a 2D vector of closest approach (not 3D), as the z component of the 3D IP is negligible for tracks traversing the LHCb angular acceptance. The motivation for this strategy is as follows.

Figure 9.8 shows the IP_x and IP_y distributions before any smearing in $Z \to \mu\mu$ data and simulation control sample (passing the `Z` selection; see Table 6.1). The distributions are roughly Gaussian, but have enhanced tails, with larger tails in the simulation than the data. The tails are crucial, because the muons here overlap significantly in IP with displaced muons from $W \to \tau\nu$. This problem is compounded by the much larger overall yield of prompt muons than displaced muons (owing to the $\tau \to \mu$ branching fraction

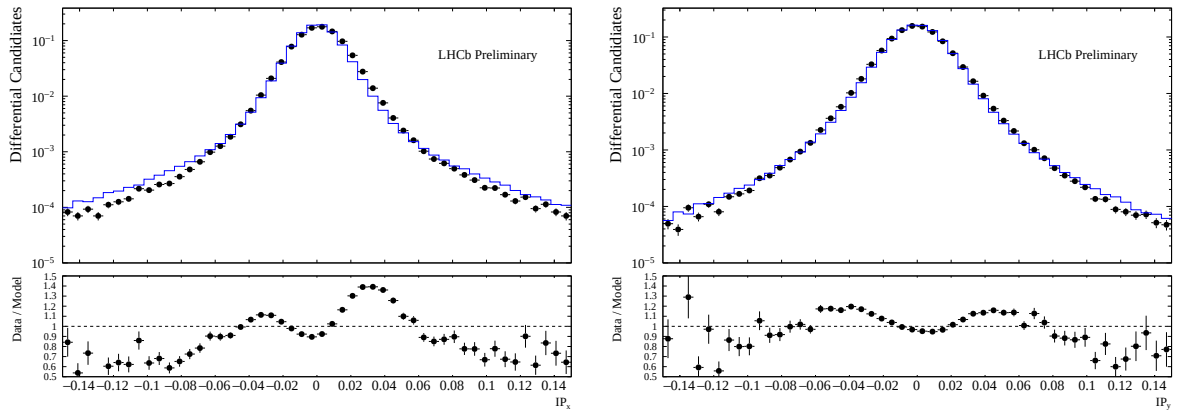


Figure 9.8: Normalized IP$_x$ (left) and IP$_y$ (right) distributions in data (black points) and simulation (blue line), with no smearing or corrections applied to the simulation.

and lower kinematic acceptance for $W \to (\tau \to \mu\nu\nu)\,\nu$ events), which mean that prompt muons are not at all negligible here. The basic smearing tool previously used for m_W is not sufficient to model these tails: indeed a smearing is more appropriate for enhancing tails in the simulation, not weakening them.

The primary cause for the overestimation of the tails in IP was found to be mismodelling of the PV track multiplicity distribution. Studies of the LHCb reconstruction performance (Section 3.6; Figure 3.22 in particular) have shown that the PV coordinate resolution is dependent upon the number of tracks originating from the PV. Since the IP is calculated with respect to a PV's position, it should be expected that the IP resolution of a given track is therefore also dependent on the number of tracks pointing back to its PV. There may be more than one reconstructed PV in an event, so the track is assigned to the PV that it best fits to, which is therefore known as the "best primary vertex" (BPV). This dependency of IP resolution on $N^{\text{BPV}}_{\text{tracks}}$ was indeed observed, and it can lead to mismodelling of the IP resolution if the $N^{\text{BPV}}_{\text{tracks}}$ is itself not properly simulated. During the m_W analysis, it was observed that the LHCb simulation severely underestimates $N^{\text{BPV}}_{\text{tracks}}$ on average, as can be seen for $Z \to \mu\mu$ data and simulation in Figure 9.9.

Weights can be derived to correct the $N^{\text{BPV}}_{\text{tracks}}$ mismodelling by taking the ratio of the normalized data and simulation $N^{\text{BPV}}_{\text{tracks}}$ distributions. However, while these weights (by design) correct the Z simulation's $N^{\text{BPV}}_{\text{tracks}}$, a residual underestimate was observed after their application to $W \to \mu\nu$ events. This non-universality between $Z \to \mu\mu$ and $W \to \mu\nu$ may be related to their subtly different boson production processes. To try to account for this, the $Z \to \mu\mu$ events are binned in reconstructed boson rapidity, which is flatter for the W^+ at forward rapidities in LHCb than the W^- and Z, meaning that low-y Z events are hopefully more "W^--like" [6]. The resulting $N^{\text{BPV}}_{\text{tracks}}$ weights are normalized in each rapidity slice to ensure that they do not change the rapidity distribution. To mitigate any

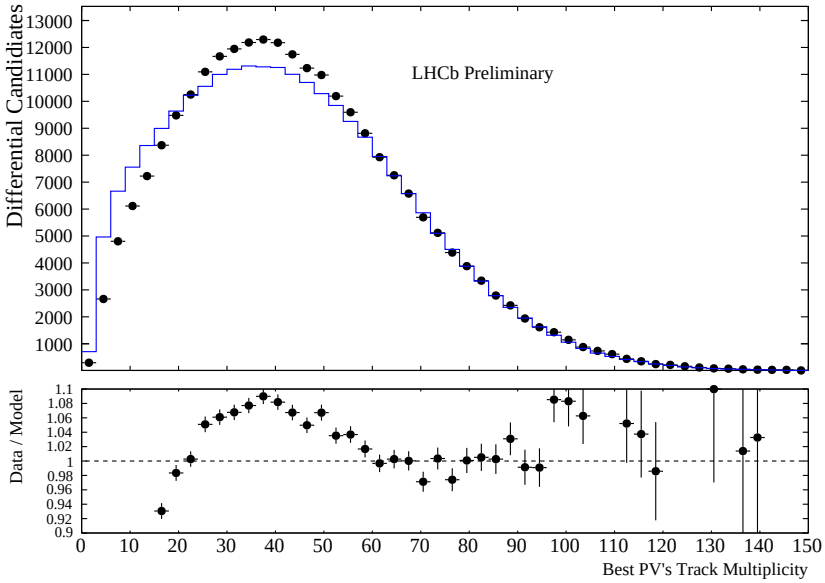


Figure 9.9: Normalized $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ distributions in the $Z \to \mu\mu$ data (black points) and simulation (blue line) without any corrections applied.

statistical uncertainty due to the limited size of the $Z \to \mu\mu$ control sample, the $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ weights are fitted as a function of rapidity within each $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ bin, and weights are taken from this function's value, thereby smoothing out the weight distribution. The function is also used to extrapolate out to higher and lower rapidity values, as the W boson samples cover a wider range in rapidity than the Z. The fitted parametric function is a linear function by default.

After these weights are applied to the $Z \to \mu\mu$ simulation, a more familiar scenario is reached: the simulation now *underestimates* the tails in the data's IP component distributions, and can be smeared. However, the data's non-Gaussian, enhanced tails would still be difficult to model with a smearing based on a Gaussian, therefore the smearing tool (from Section 6.5) was extended to be able to fit for a "double-Gaussian" smearing model. The hitherto-used "single-Gaussian" model shifts the unsmeared data by adding a mean μ, and broadens it by adding a Gaussian-distributed random number of width σ. This is also the case for the double-Gaussian function, except that it adds on a larger (by a multiplicative width factor a) broadening term for a small fraction f_2 of events, thus enhancing the tails. Mathematically, the "double-Gaussian" smearing function is implemented as

$$x \to x' = x + \mu + \sigma \cdot \mathcal{N}(0,1) \times \begin{cases} a & \mathcal{U}_{[0,1]} > 1 - f_2 \\ 1 & \mathcal{U}_{[0,1]} \leq 1 - f_2 \end{cases}, \tag{9.7}$$

where $\mathcal{U}_{[0,1]}$ is a uniformly-distributed random number between 0 and 1. The smearing tool is also configured to reweight the simulation according to the $N^{\text{BPV}}_{\text{tracks}}$ weights, and the smearing fits are binned in $N^{\text{BPV}}_{\text{tracks}}$ and η to capture the observed dependencies. There is a large degeneracy in the four smearing parameters, and thus the parameter f_2 is fixed to 0.1, which was observed to lead to the best fit qualities. As before, the remaining smearing parameters μ, σ and a are fitted to the data by minimizing an unbinned Anderson-Darling statistic. Two examples of these smearing fits are shown in Figure 9.10.

The final piece of the IP modelling strategy is to reweight the IP distribution of prompt muons at very high IP. After application of the $N^{\text{BPV}}_{\text{tracks}}$ weights and the double-Gaussian IP smearing, and integrating back over muon η and $N^{\text{BPV}}_{\text{tracks}}$, an overestimate at very high IP values is now seen. This is possibly caused by a missing higher-order term in the smearing function, or poor statistics to control the smearing function in this area when binned in η and $N^{\text{BPV}}_{\text{tracks}}$. Since prompt muons with such high IP should have little influence on determining $R_{\tau\mu}$, the solution to this to directly reweight the IP distribution here based upon the ratio of normalized data and (smeared and $N^{\text{BPV}}_{\text{tracks}}$-weighted) simulation in $Z \to \mu\mu$. There is a very limited number of muons in $Z \to \mu\mu$ at these highest IP values, so to smooth out the statistical fluctuations the ratio histogram is fitted with an error function, which is observed to describe the shape of the histogram well. The resulting IP weights for prompt signals ($Z \to \mu\mu$, $W \to \mu\nu$ and DIF) are taken from the function value.

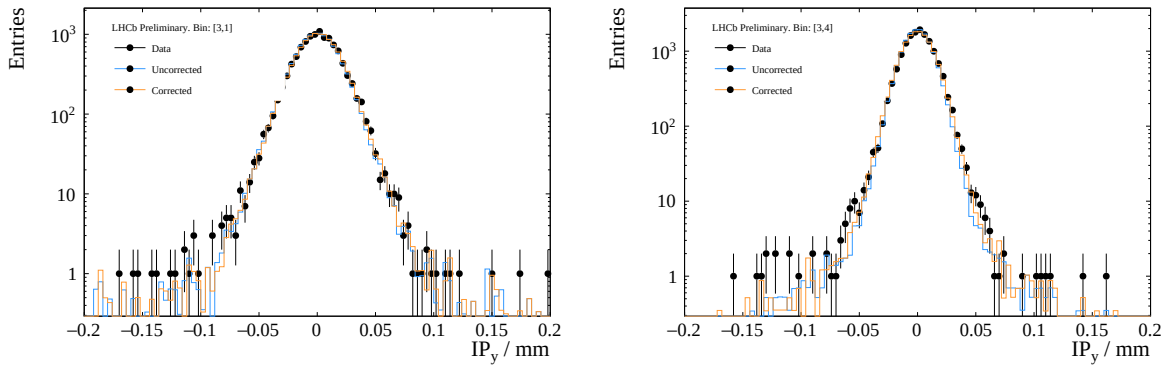


Figure 9.10: IP$_y$ distribution for the $Z \to \mu\mu$ data, with the unsmeared and double-Gaussian-smeared simulation overlaid. The bin indices indicate that these muons are from $2.5 < \eta < 2.8$, and $10 < N^{\text{BPV}}_{\text{tracks}} < 25$ (left) and $40 < N^{\text{BPV}}_{\text{tracks}} < 55$ (right).

9.6 Overview of the analysis/template preparation

The templates for all fit components were made using the full LHCb simulation. This includes the DIF component, which comes from an inclusive simulation of hard QCD events that have been weighted by $1/p$ to mimic decays-in-flight. As with the m_W analysis, several steps and the majority of time in the analysis workflow are required to prepare the templates and data for use in the $R_{\tau\mu}$ fit. These steps are focussed on trimming down the events to only the variables used in the final analysis, deriving corrections from control samples, and applying those corrections to the data and simulation where appropriate. The strategies, methods and motivations of each of these steps has been explained in detail in Chapters 6, 5 and 7, and the preceding section.

As this measurement is more of a precise counting exercise rather than a precise measurement of the p_{T} shape, it is expected that the overall momentum scale does not have to be quite so stringently controlled as for the m_W analysis. Therefore, no fill-dependent momentum scale corrections are applied. The first step is therefore to re-align the detector offline and remove any curvature biases in the data using the pseudomass method, as detailed in Section 6.3, applied on Z data and simulation. Next, the distributions of the PV's track multiplicity $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ are compared for Z data and simulation, in order to derive a set of weights that will correct the underestimation of $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ in the simulation, as described in Section 9.5. By default, eight $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ bins (above $N^{\mathrm{BPV}}_{\mathrm{tracks}} = 10$) are used. Five y bins are used from $2 < y < 4.5$, which are then extrapolated and interpolated by a linear function to 30 bins in a wider y range. The ratio of $Z \to \mu\mu$ data to simulation as a function of $N^{\mathrm{BPV}}_{\mathrm{tracks}}$, in slices of Z rapidity, is shown in Figure 9.11 before and after the linear-fit-based extrapolation is applied to smoothen the histogram and extend to higher rapidities. The latter is used to derive weights. The expected dependence with $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ is observed, with a ratio of 0.5 or lower being observed in the lowest $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ and the ratio becoming positive for large $N^{\mathrm{BPV}}_{\mathrm{tracks}}$. A weak dependence is observed with rapidity.

With these weights available to be used to correct the Z simulation, and the detector precisely aligned (and the tracks momenta recomputed accordingly) the smearing of the IP_x and IP_y distributions with the double-Gaussian smearing model can proceed. The smearing model and fits are described in Section 9.5. A summary plot of the IP_x and IP_y fits is shown in Figure 9.12. The values of σ show that, in a typical $\left(\eta, N^{\mathrm{BPV}}_{\mathrm{tracks}}\right)$ bin, the IP_x and IP_y distributions have been widened by the smearing, and the χ^2/ndf panel shows that the fit quality of the smearing fits is good.

Three further smearing steps are then performed. The first two of these steps also use smearing method described in the previous section. Firstly, the χ^2_{trk} distribution in simulation is smeared to match the data using the single-Gaussian smearing model (similarly to m_W - see Section 6.5) in three bins of muon η. The next smearing is of

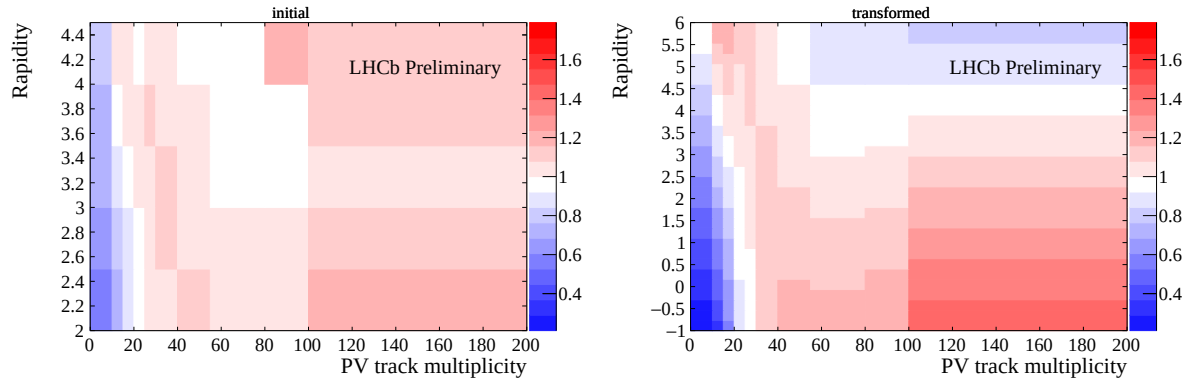


Figure 9.11: Ratio of data to simulation of the $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ distribution sliced in rapidity. Each rapidity slice is normalized. Left: Raw, coarse histogram. Right: Final histogram used for weights, after interpolation and extrapolation from the edges of the histogram domain to high and low rapidity.

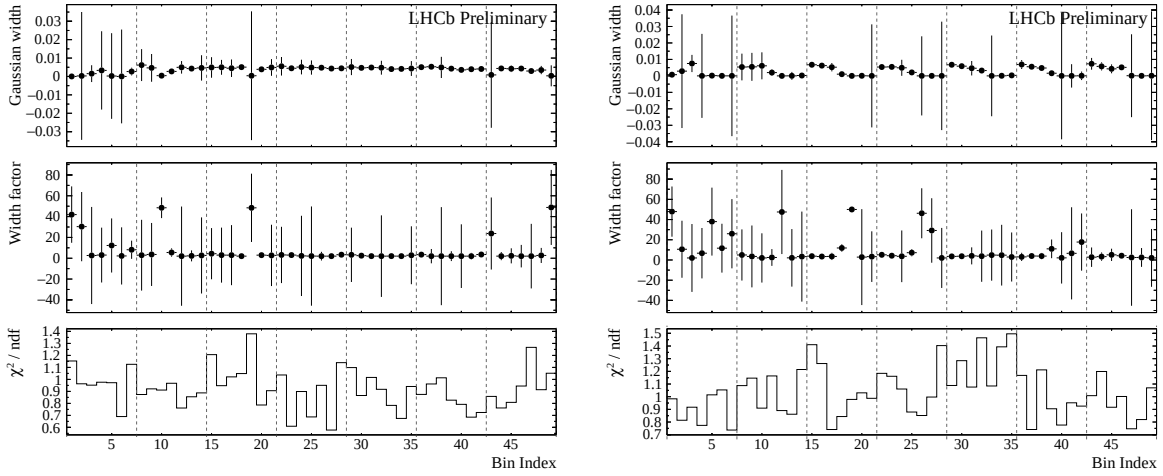


Figure 9.12: Summary of the fitted smearing parameters and fit quality for IP_x (left) and IP_y (right). The nested binning is in η, while after each dashed line, the $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ bin increments by one.

the muon isolation distribution. Although the $\mathcal{I}$ isolation variable was seen to be quite well-modelled in the simulation, the $\log_{10}(\mathcal{I})$ distribution for $Z \to \mu\mu$ motivated a simple scaling $\mathcal{I} \to k \times \mathcal{I}$ to improve the modelling, so the smearing tool described above was configured to fit for such a scaling. The Z data and simulation are here divided into seven bins of muon η and seven bins of the recoil projection u to extract the scaling parameter k, and a different scaling fit is derived for both the neutral and charged components of the $\mathcal{I}$ variable. The scaling parameters and fit quality in each bin is shown in Figure 9.13.

The final smearing step in the $R_{\tau\mu}$ analysis is of the momentum scale. The momentum scale is expected to be less important for this measurement than for m_W, so the simplest option is taken: exactly the same configuration of the momentum scaling

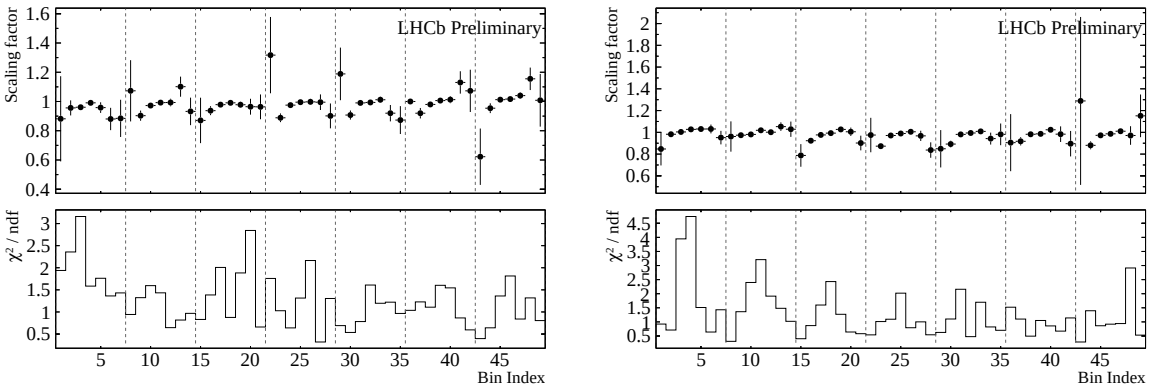


Figure 9.13: Summary of the scaling parameters and fit quality for the charged (left) and neutral (right) isolation variables. The nested binning is in the recoil projection variable u, while after each dashed line, the η bin increments by one.

procedure as the m_W analysis (as detailed in Section 6.4) is used here. The resulting smearing parameters are shown in Table 9.2, and - after integrated the categories back together - the post-smearing Z, $\Upsilon(1S)$ and J/ψ mass peaks are compared for data and simulation in Figure 9.14. The total fit quality is good, with $\chi^2/\text{ndf} = 2063/2082$. A typical signal muon in this dataset has $p \sim 400\,\text{GeV}$, meaning that the residual curvature bias is a few parts per mille, and the curvature resolution smearing is $\mathcal{O}(1\%)$ when the $1/\cosh\eta$ term (≈ 0.1 for $\eta = 3$) is accounted for.

With the fitted smearing parameters for the IP, χ^2_{trk}, isolation and momentum measurement to hand, all simulated samples required in the remainder of the analysis are processed and smeared accordingly. All the samples needed to construct templates for the fit are processed and required to pass the `W` selection. The `Z` control sample is re-processed with the up-to-date smearing parameters, and the `ZMuID`, `ZTrkEff` and `U1S` control samples are processed for use in the muon reconstruction efficiency treatment. The "all-but-one-cut" strategy, as introduced in Section 6.1, is applied here.

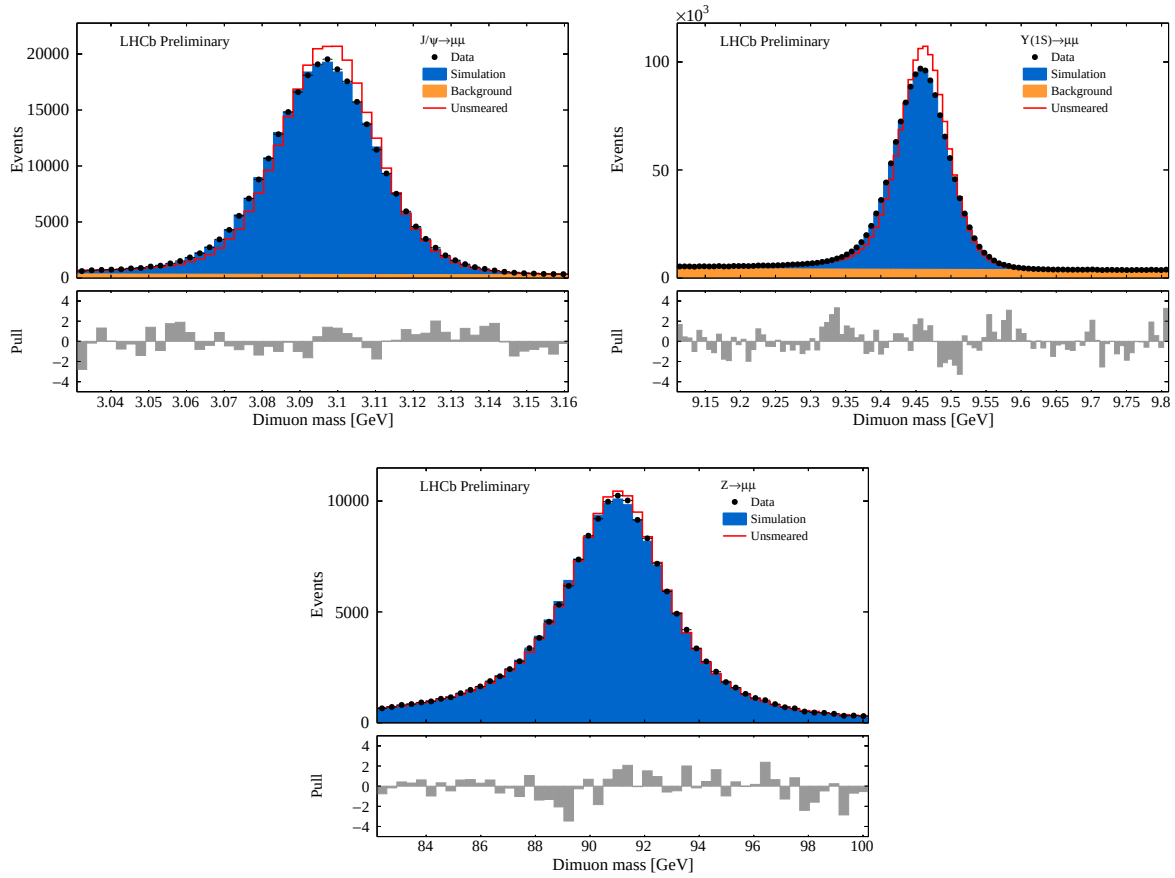


Figure 9.14: Fitted dimuon mass distributions for J/ψ, $\Upsilon(1S)$ and Z boson candidates, combining all η and magnet polarity categories, to determine the smearing parameters to be applied to the simulation. The red histogram indicates the model before the application of the smearing.

Table 9.2: Parameter values (and their errors) of momentum smearing fit.

Parameter	Value	Error
Momentum scale ($\times 10^{-3}$) $[1.8 < \eta < 2.2]$	0.45	0.10
Momentum-independent smearing ($\times 10^{-3}$)	2.02	0.02
Momentum-dependent smearing ($\times 10^{-6}$ GeV^{-1}) $[1.8 < \eta < 2.2]$	208.4	15.9
Curvature bias ($\times 10^{-6}$ GeV^{-1})	-0.33	0.39
Momentum scale ($\times 10^{-3}$) $[2.2 < \eta < 4.4]$	-0.16	0.00
Momentum-dependent smearing ($\times 10^{-6}$ GeV^{-1}) $[2.2 < \eta < 4.4]$	142.1	7.4

After processing of the Z simulation and with all the cuts of the `Z` selection applied, the muon $\log_{10}(\text{IP})$ and $\log_{10}(\mathcal{I})$ distributions in Z data and simulation (integrated over all other variables) are shown in Figure 9.15. The double-Gaussian structure of the IP_y component can be seen to have a clear effect on the $\log_{10}(\text{IP})$ distribution at high IP, where a kink in the tail is clearly visible. Unfortunately, the double-Gaussian smearing has

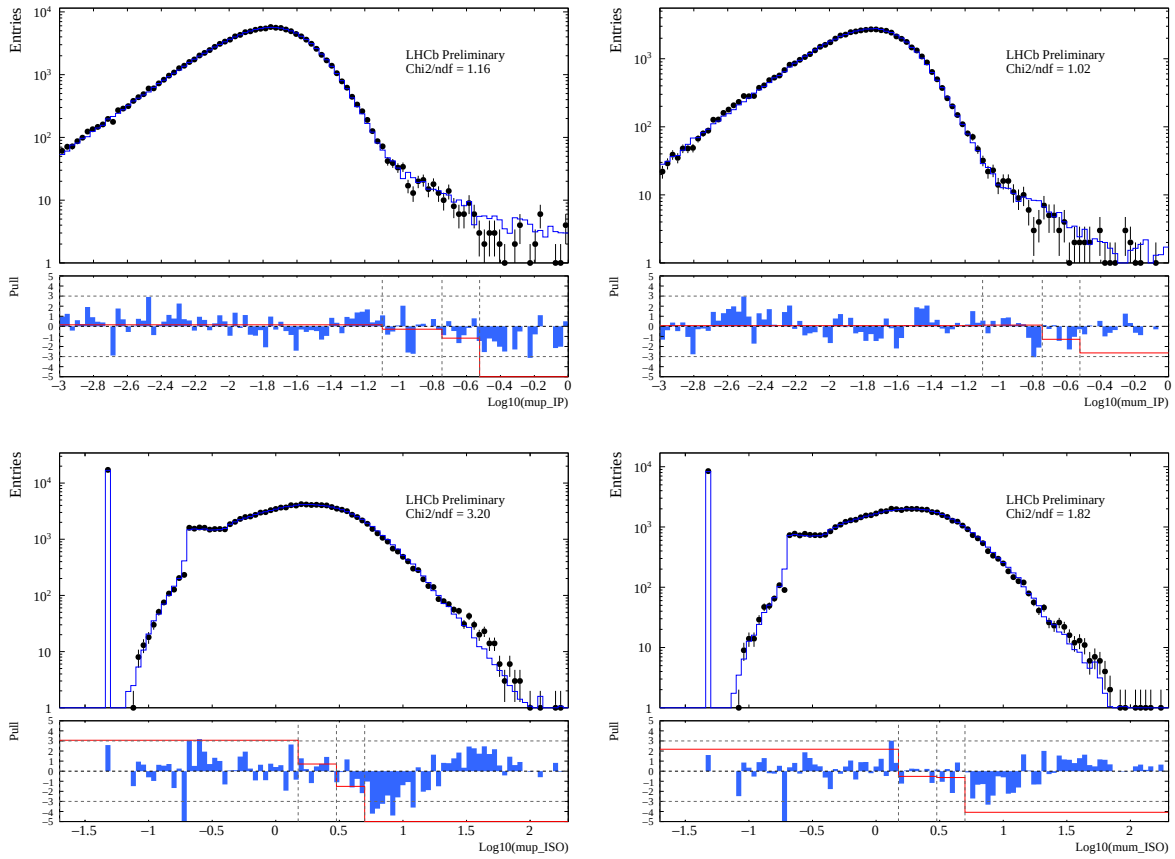


Figure 9.15: Data and simulation (normalized to data) $\log_{10}(\text{IP})$ (top) and $\log_{10}(\mathcal{I})$ (bottom) distributions for μ^+ (left) and μ^- (right). A χ^2/ndf is shown for each, which evaluates the compatibility of the two shapes. In the pull plots, the red line signifies what the pulls would be if the $R_{\tau\mu}$ fit binning was used, the edges of which are indicated by the dashed vertical lines. The rightmost vertical line in each pull plot is therefore the upper edge of the $R_{\tau\mu}$ fit binning. Both the IP and $\mathcal{I}$ lower edges are zero, so they cannot be shown, but the $\mathcal{I}$ is truncated to show the "zero" bin.

slightly over-inflated the tail at the very-highest IP. Aside from this region, the simulation is seen to model the data extremely well. The modelling of the isolation distribution is seen to be poorer overall, although this is mostly driven by the highest isolation values, which are outside the $R_{\tau\mu}$ fit region. Within the fit region, the modelling is better, although the simulation is still slightly biased towards lower isolation values. Some mismodelling of the isolation shape in data is therefore still to be expected.

The next set of steps in the $R_{\tau\mu}$ workflow are concerned with correcting any mismodelling in the simulation of the effect of the selection requirements. Corrections in all cases come via event weights. Firstly, the muon tracking, trigger and identification efficiencies are calculated in data and simulation via the tag-and-probe method, leading to a weights-based correction as detailed in Chapter 7. There is currently no difference in the strategy here with respect to that used for the m_W measurement. The resulting

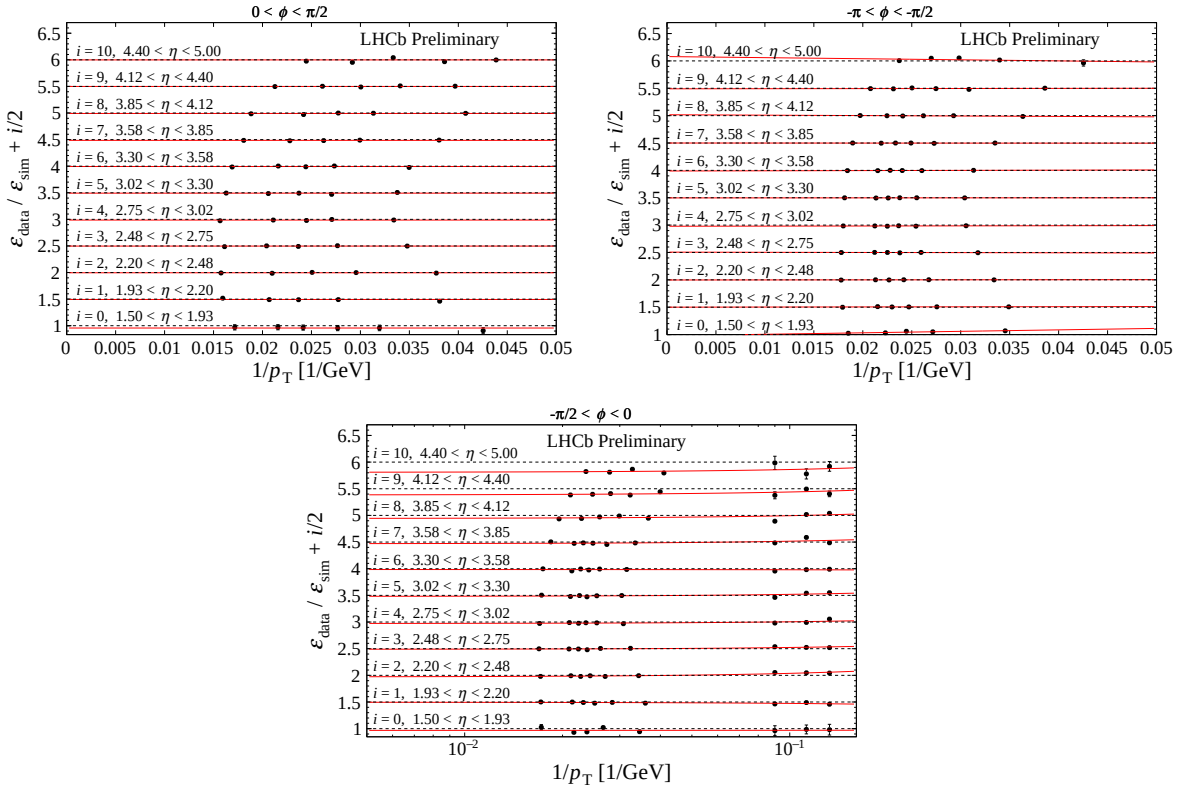


Figure 9.16: Muon tracking (top left), identification (top right) and trigger efficiency ratios for arbitrary ϕ bins. All η bins are included by offsetting the graphs by a constant. The red lines indicate the fitted parametrization to the points in each bin. For the trigger efficiencies, the three columns of points at the far right come from $\Upsilon(1S) \to \mu\mu$, while all other points are from Z decays.

parametrizations of the efficiency ratios for the muon tracking, identification and trigger efficiencies are shown for one (arbitrary) bin in ϕ in Figure 9.16. As observed in the m_W analysis, the largest trends here are of the trigger efficiency in η and near the turn-on of the L0 trigger at low p_{T}. A correction to the χ^2_{trk} distribution is not deemed necessary for this analysis.

No corrections are required with respect to an χ^2_{IP} or isolation cut since substantial effort has been made to make sure the distributions are well-modelling, and the only cut in any case is at the upper edge of the fit binning. The penultimate correction to be applied to the templates via weights is the high-IP parametric correction, described in Section 9.5 and shown to be necessary in Figure 9.15. The fit to the ratio of IP distributions from data and simulation is shown in Figure 9.17. The overshooting of the simulation at very high IP is seen, but it is well-modelled by the error function. Reassuringly, this fit supports the conclusion that the simulation is well-modelled in the bulk of the IP distribution. Weights are derived from this fit and applied to the $W \to \mu\nu$, $Z \to \mu\mu$ and DIF templates to improve mismodelling at the highest values of IP. The final weight-based correction to be

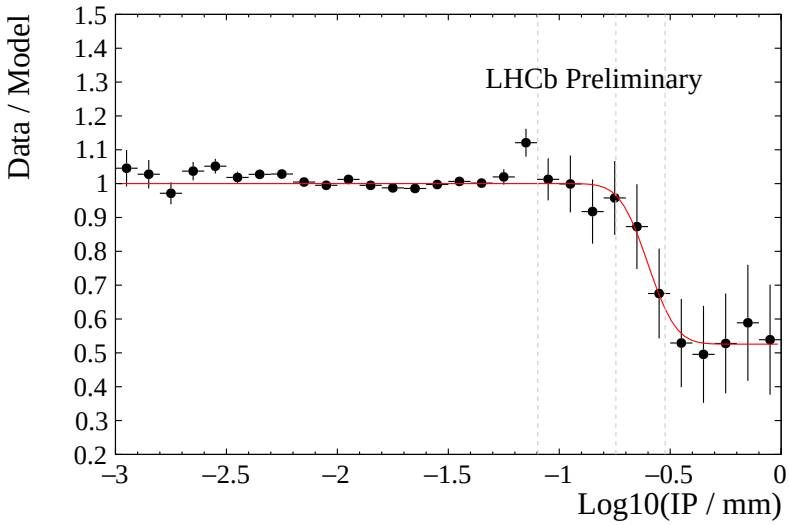


Figure 9.17: Normalized ratio of data to simulation of muon P from $Z \to \mu\mu$, post-smearing. Both muon charges are used in this plot. Superimposed is a fit to the ratio with an error function from which weights are derived for the key prompt EW components. The dashed lines indicate the bin edges in IP in the $R_{\tau\mu}$ fit; the rightmost line is the upper edge of the fit binning.

applied ensures that the simulated samples have the same proportion of magnet polarities as the real data.

Finally, weights are derived to transform the templates from the LO Pythia physics description to the default physics model as described in Section 5.1: angular coefficients at $\mathcal{O}(\alpha_s^2)$ from DYTurbo, unpolarized cross section from POWHEGPythia at NLO in α_s, and the central replica of the NNPDF3.1 NLO PDF set. Since the $R_{\tau\mu}$ fitter does not float the tuning parameters α_s and $k_{\mathrm{T}}^{\mathrm{intr}}$, fixed values are taken. The α_s^Z and $k_{\mathrm{T}}^{\mathrm{intr}}$ values used are the best-fitting values from the m_W fitter, while a 2% lower value than that found for m_W is taken for α_s using with $W \to \mu\nu$ and $W \to \tau\nu$. This was seen to improve the modelling of the Jacobian peaks in the $R_{\tau\mu}$ fit at low IP and low isolation, where $W \to \mu\nu$ dominates. The A_3 scaling factor is left equal to 1 by default. These physics weights are only applied to $W \to \tau\nu$, $W \to \mu\nu$, $Z \to \mu\mu$ and $Z \to \tau\tau$ - their effect to the other EW components would be negligible in terms of $R_{\tau\mu}$. Template events also have a default final-state radiation (FSR) weight corresponding to the average of predictions from Pythia, Herwig and Photos.

9.7 Status of fit results

Figure 9.18 shows the data and the current status of the default fitted model used to determine $R_{\tau\mu}$. The 3D templates have been flattened to 1D as described in Section 9.3. The blinded $R_{\tau\mu}$ value has an uncertainty of 0.035, or 3.5%. The current fit χ^2 per degree of freedom is 256/167, which is driven by small pockets of mismodelling: there appears to be too much prompt signal at high IP; and there are hints of mismodelling of the isolation shape of the signal components, which cannot be seen here as it at the $\mathcal{O}(1\%)$ level. Both of these features are somewhat expected given the control plots seen earlier. The low- and mid-IP regions are well-modelled at the few per-cent level, and larger discrepancies elsewhere should not be drive a significant bias in $R_{\tau\mu}$. The current fitted values of the floating parameters are shown in Table 9.3. It is interesting to note that the HF components are charge-symmetric within their uncertainties, while DIF shows a significant asymmetry towards positive muons. The negative values of the Hagedorn exponents β_{HF} and β_{DIF} indicate that the fit preferred a softer momentum distribution for both DIF and HF, although the overall correction was small. The denominator value of the Crystal-Ball α parameter was 0.8, so the higher (lower) value for HF (DIF) corresponds to a steeper (shallower) slope towards higher isolation values. In Figure 9.19 the fitted model has been projected onto the data's 1D $1/p_{\text{T}}$, IP and isolation distributions. Within the $R_{\tau\mu}$ fit regions - as indicated by the vertical dashed lines - the modelling of all these distributions is shown to be good by the ratio plots.

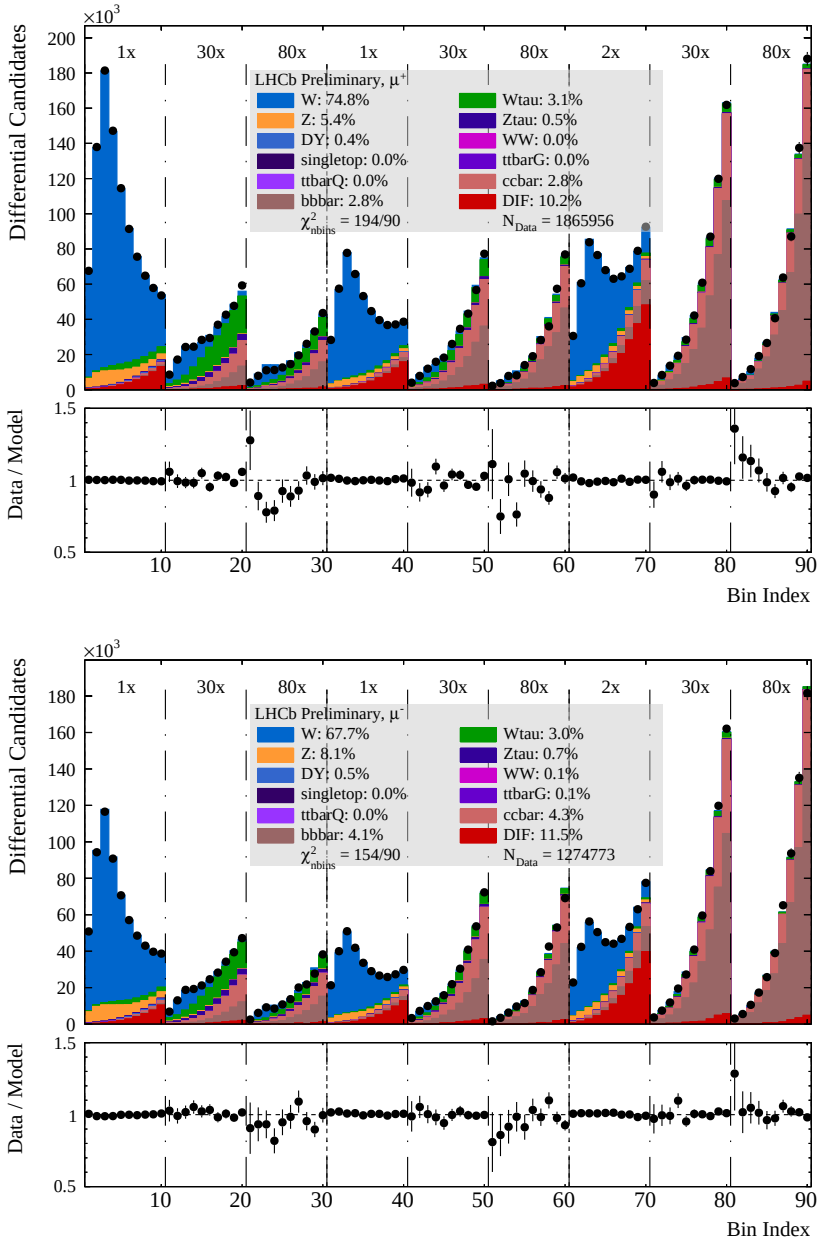


Figure 9.18: Post-fit status of the fit model overlaid on the data. The dot-dashed lines delineate the different IP bins, and the double-dashed line delineates the isolation bins. The numbers at the top of the plot are the scale factors applied to the templates in that $(\mathrm{IP}, \mathcal{I})$ bin to give visualizable bin populations in all bins. The χ^2 here is evaluated from the plot and per the number of bins, and is not the same as the fitted χ^2, as the latter is determined with the Beeston-Barlow Lite method. The lower panels show the ratio of the data to the model in each bin.

Table 9.3: Parameter values of the default fit.

Parameter	Value	Error
$R_{\tau\mu}$	0.997	0.035
$\sigma'(W^+ \to \mu^+\nu)$	3427.3 pb	4.6 pb
$\sigma'(W^- \to \mu^-\nu)$	2405.2 pb	4.0 pb
$N\,(c\bar{c} \to \mu^+ + X)$	51446	3847
$N\,(c\bar{c} \to \mu^- + X)$	54544	3409
$N\left(b\bar{b} \to \mu^+ + X\right)$	52600	2178
$N\left(b\bar{b} \to \mu^- + X\right)$	51698	2038
$N\,(\mathrm{DIF} \to \mu^+ + X)$	190912	2113
$N\,(\mathrm{DIF} \to \mu^- + X)$	146983	1842
β_{HF}	-0.022	0.069
β_{DIF}	-0.292	0.091
α_{HF}	0.989	0.016
α_{DIF}	0.091	0.007

A variety of other post-fit plots are made to check the accuracy of the physics and detector modelling in the templates. Firstly, the muon η and ϕ distributions are shown in Figure 9.20. Many corrections, particularly the muon reconstruction efficiencies, have a dependence in these variables, so their adequate modelling within the fit region is reassuring. A slope with η can be observed, which is to be investigated and understood. Finally, the muon $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ and χ^2_{IP} distributions are shown in Figures 9.21. Although the majority of the mismodelling of $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ has been corrected by the weights derived from Z events, a small over-estimate at low $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ still remains in the simulation of the single-muon components. Since low $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ correlates so strongly with poor IP resolution, it can be inferred that the excess of prompt muons at high IP may be related to this remaining mismodelling. At the time of writing, this is still under study, with PV misassociation being one of the leading hypotheses for a $N^{\mathrm{BPV}}_{\mathrm{tracks}}$ mismodelling source that has not been accounted for. A slope can also be seen in the ratio plot of χ^2_{IP}, which is unsettling given the variable's correlation to IP and suggests that the modelling of impact-parameter-related variables is not fully understood yet.

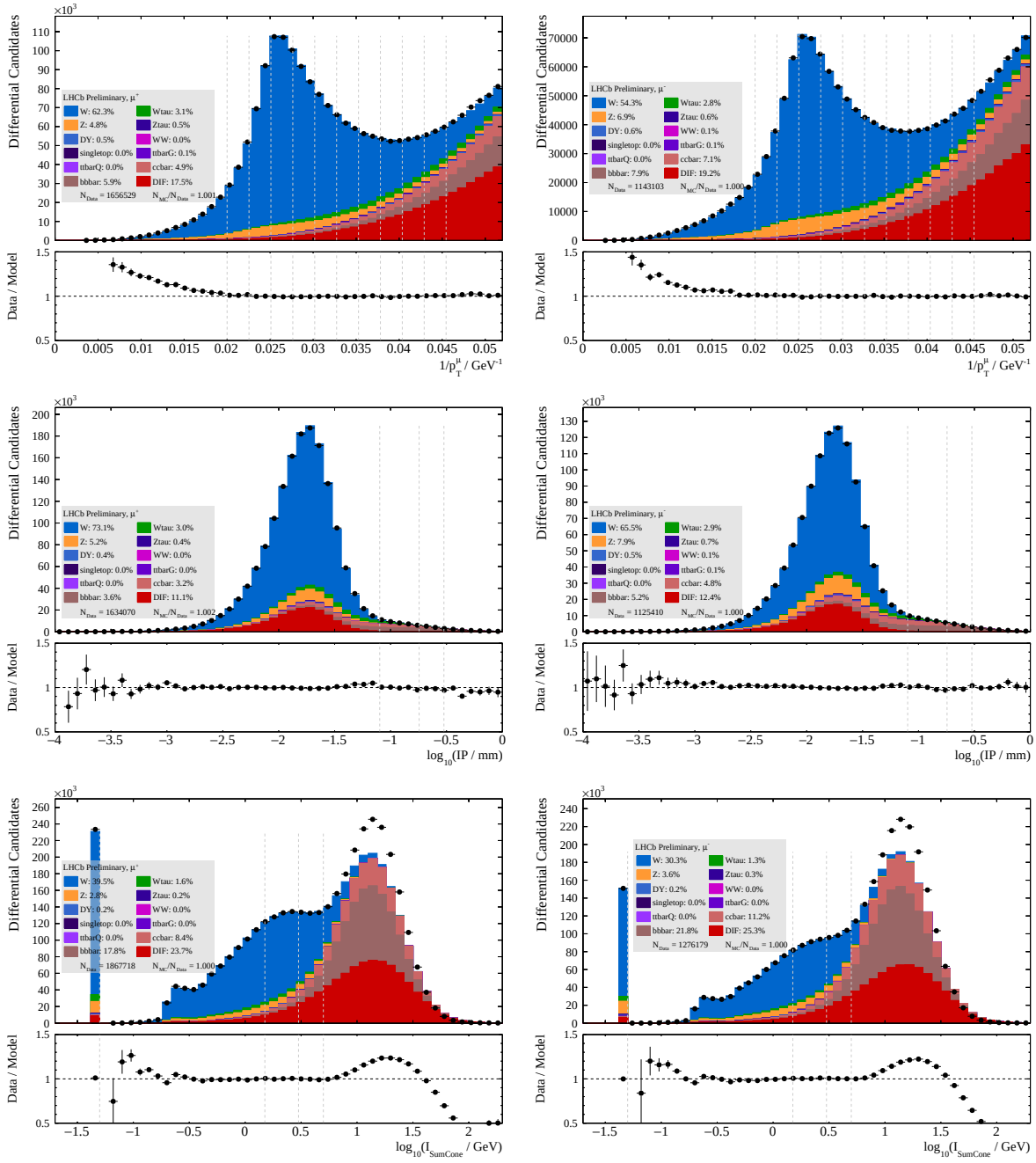


Figure 9.19: Post-fit projections of the muon $1/p_{\mathrm{T}}$ (top), $\log_{10}(\mathrm{IP})$ (middle) and $\log_{10}(\mathcal{I})$ (bottom) distributions, where all other variables have been integrated over. The vertical dashed lines indicate the bin edges of the $R_{\tau\mu}$ fit binning in each variable. For IP and isolation, the lowest edge is zero, so this line is missing and the lowest bin propagates to (and past) the left edge of the x axis. Those muons with an isolation of zero have been shifted up to an arbitrary value to be included. The rightmost vertical dashed line indicates the upper limit of the $R_{\tau\mu}$ fit binning in all plots.

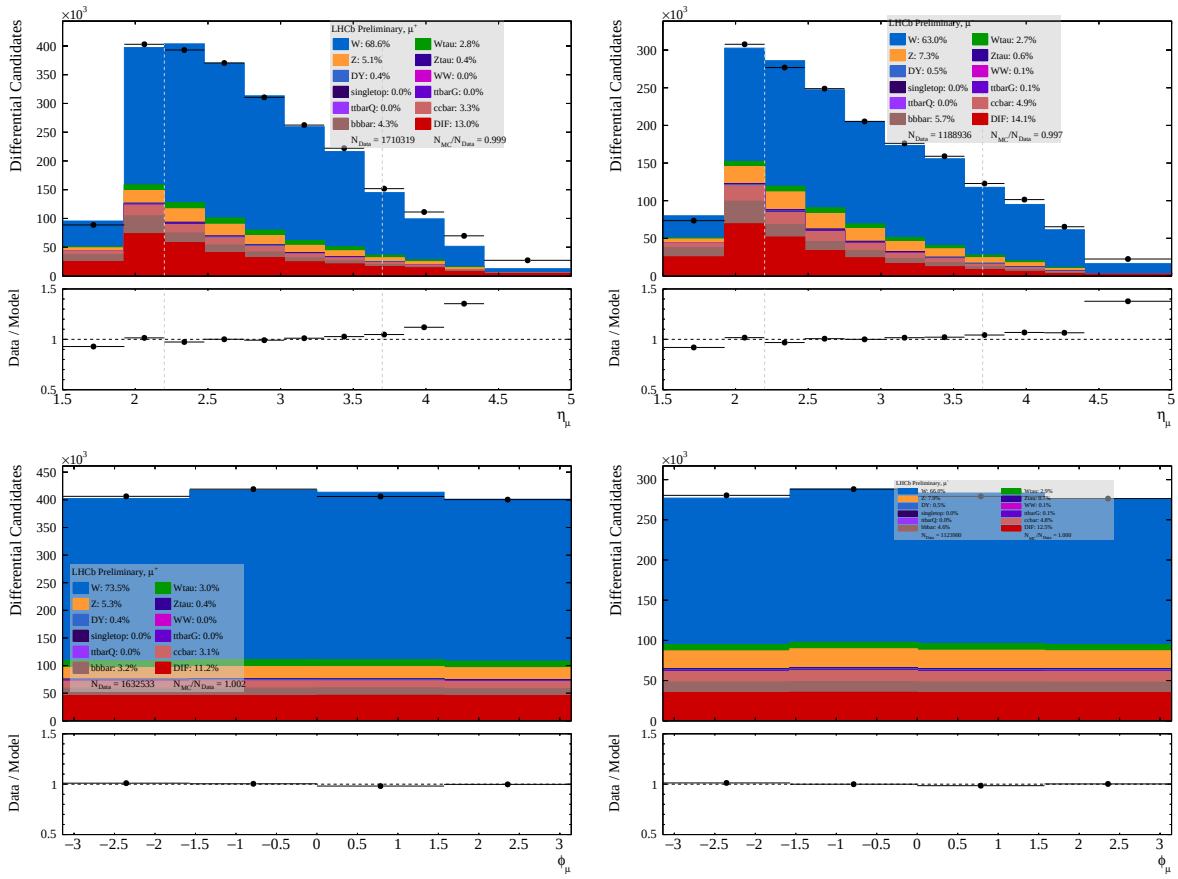


Figure 9.20: Post-fit projections of the muon η (top) and ϕ (bottom), where all other variables have been integrated over. In the former case, the range between the vertical dashed lines is what is included in the $R_{\tau\mu}$ fit. The binning is chosen to match the binning of the muon reconstruction efficiency parametrizations.

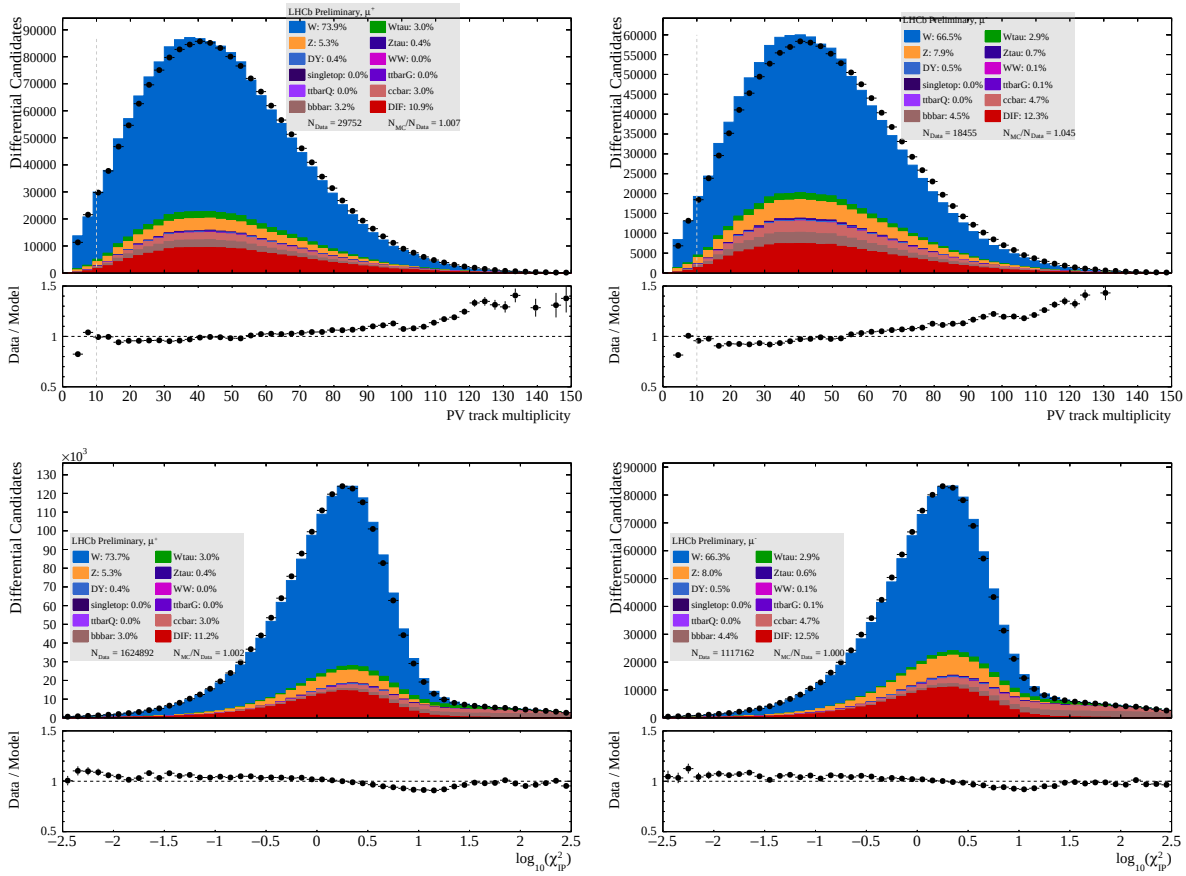


Figure 9.21: Post-fit projections of the muon $N^{\text{BPV}}_{\text{tracks}}$ (top) and χ^2_{IP} (bottom), where all other variables have been integrated over. In the former case, only the range above the vertical dashed line is included in the $R_{\tau\mu}$ fit.

9.8 Systematic uncertainties

Table 9.4 summarizes the status of the systematic uncertainties in this measurement. All systematic uncertainties are measured by performing fits for $R_{\tau\mu}$ with one variation made to the analysis. The systematic uncertainty is then calculated by comparing this value - or a set of such $R_{\tau\mu}$ values - to the default value. Systematic uncertainties are calculated by taking an RMS of a set of $R_{\tau\mu}$ values, the absolute difference between two values, or by calculating the symmetric envelope (henceforth "envelope"). Analogous to m_W, the envelope is defined as $\pm\max(|\delta R_{\tau\mu,i}|)$, where $\delta R_{\tau\mu,i} = R_{\tau\mu,\text{default}} - R_{\tau\mu,i}$.

The leading systematic uncertainties primarily come from experimental sources. The variations made to evaluate each systematic uncertainty, in the order of the table, are as follows. In parentheses after the name, the method of evaluation (envelope, absolute difference or RMS) is also given.

- **IP smearing parameters** (Envelope): The fixed f_2 parameter is varied up (down) from 0.1 to 0.15 (0.05).
- $N_{\text{tracks}}^{\text{BPV}}$ **extrapolation function** (Abs. Diff.): Interpolation/extrapolation is flat as a function of rapidity.
- **Isolation form factor** (Envelope): Each of the fixed parameters in the Crystal-Ball form factors are varied up and down by sensible amounts, chosen such that the modelling of the simulated isolation shapes was still reasonable, as checked by reproducing plots like Figure 9.4.
- **PDF set** (Envelope): The central PDF replicas from CT18 and MSHT20 are used.
- **High-IP parametric correction** (Envelope): The number of bins used to fit the error function (see Figure 9.17) has been made increased and decreased by a factor of two.
- $N_{\text{tracks}}^{\text{BPV}}$ **statistical uncertainty** (RMS): Before any fitting and extrapolation, the bin contents of the $N_{\text{tracks}}^{\text{BPV}}$ ratio histogram (Figure 9.11, left) are adjusted to Gaussian random numbers of mean equal to the original ratio and width equal to the error in that mean.
- **Angular coefficients** (Envelope): The A_3 fudge factor is set to 0.9 and 1.1.
- **Muon ID & trigger efficiencies** (Envelope): As described in Section 7.6.
- **IP smearing η binning** (Envelope): The number of η bins (within the η range of the $R_{\tau\mu}$ fit) used in the IP smearing is increased (decreased) from 5 to 8 (3).
- $N_{\text{tracks}}^{\text{BPV}}$ **weights y binning** (Envelope): The number of initial y slices used in the $N_{\text{tracks}}^{\text{BPV}}$ weights (before fitting and extrapolating in y) is increased (decreased) from 5 to 10 (2).
- **Residual curvature biases** (Envelope): A residual curvature bias is inserted at the template-drawing stage (just before the fit) that is $\pm 3\times$ the uncertainty on the

Table 9.4: Systematic uncertainties for this analysis, listed in descending order. The "TH" and "EXP" types group the uncertainties into those coming from primarily theoretical and primarily experimental sources.

Type	Source	Fits	$\Delta R_{\tau\mu}$
EXP	IP smearing parameter uncertainty	2	3.21%
EXP	$N_{\text{tracks}}^{\text{BPV}}$ extrapolation function	1	2.12%
EXP	Fixed isolation form factor parameters	8	1.33%
TH	Choice of PDF set	1	1.09%
EXP	High-IP parametric correction	2	1.13%
EXP	$N_{\text{tracks}}^{\text{BPV}}$ statistical uncertainty	10	1.02%
TH	Angular coefficients	2	0.80%
EXP	Trigger & Muon ID efficiency	11	0.24%
EXP	IP smearing η binning	2	0.65%
EXP	$N_{\text{tracks}}^{\text{BPV}}$ weight y binning	2	0.67%
EXP	Residual curvature biases	2	0.52%
TH	Transverse momentum model ($k_{\text{T}}^{\text{intr}}$)	2	0.36%
EXP	$N_{\text{tracks}}^{\text{BPV}}$ binning	2	0.36%
EXP	Isolation scaling binning	2	0.27%
EXP	Isolation recoil binning	2	0.26%
TH	FSR model	2	0.26%
EXP	Trigger & Muon ID control sample size	10	0.24%
EXP	Fixed Hagedorn parameter	2	0.17%
EXP	Tracking efficiency	12	0.11%
TH	Transverse momentum model (α_s)	2	0.22%
EXP	Tracking control sample size	10	0.02%
EXP	Residual momentum scaling	2	0.03%
Theory		9	1.4%
Experimental		82	4.5%
Statistical		1	3.5%

curvature bias parameter derived from the momentum smearing fit (see Table 9.2).

- **Transverse momentum model ($k_{\text{T}}^{\text{intr}}$)** (Envelope): $k_{\text{T}}^{\text{intr}}$ is varied up and down by 30% from the default value.
- $N_{\text{tracks}}^{\text{BPV}}$ **weights binning** (Envelope): The number of $N_{\text{tracks}}^{\text{BPV}}$ bins used in the ratio histogram (Figure 9.11) is increased and decreased by approximately a factor of 2.
- **Isolation scaling binning** (Envelope): The number of η and recoil bins is increased and decreased by approximately a factor of 2.
- **FSR model** (Envelope): Weights are applied to the templates based on the FSR

model of HERWIG and PHOTOS.

- **Trigger & Muon ID control sample size** (RMS): As described in Section 7.6.
- **Fixed Hagedorn parameter** (Envelope): The Hagedorn form factor $p_{\mathrm{T},0}$ parameter is (conservatively) varied up and down by a factor 5, at which point visible degradation in the modelling of the HF and DIF p_{T} shapes (Figure 9.4) is seen.
- **Tracking efficiency** (Envelope): As described in Section 7.6.
- **Transverse momentum model (α_s)** (Envelope): α_s is varied up and down by 2% from the default.
- **Tracking control sample size** (RMS): As described in Section 7.6.
- **Residual momentum scale** (Envelope): A residual momentum scaling bias $p \to p \cdot (1 + \alpha)$ is inserted at the template-drawing stage (just before the fit) that is $\pm 3\times$ the uncertainty on the momentum scaling parameter derived from the momentum smearing fit (see Table 9.2).

The size of the variations for the isolation form factor parameters is crude and conservative, and can be refined. It is currently dominated by the variations in the Crystal-Ball N tail parameter, which can be reduced. The variations to $k_{\mathrm{T}}^{\mathrm{intr}}$ and α_s are in place of using different transverse momentum models such as DYTURBO or PYTHIA, as was done in the m_W analysis, and is a conservative simplification given that the uncertainty on these parameters in the m_W fit was around 10% and 0.5% respectively. A similar argument is made for the angular coefficients. The uncertainty due to the fixed f_2 IP smearing parameter dominates, as there is not enough control on the modelling of the double-Gaussian tails in the IP resolution function, and any variation here directly effects $R_{\tau\mu}$ by enhancing/diminishing the amount of prompt muons in the mid-IP region, where displaced muons from $W \to \tau\nu$ are most abundant. Further studies are needed to understand the IP resolution mismodelling and constrain the f_2 parameter, thereby reducing this uncertainty.

9.9 Cross-checks

A series of cross-checks are made to give confidence in the overall analysis strategy:

- **Splitting into orthogonal datasets**: the data and simulation are split by magnet polarity; charge multiplied by magnet polarity; into different η regions either side of $\eta = 2.95$; and into different ϕ regions (detector top and bottom halves, and A- and C-sides).
- **Adjusting the fit range**: A variety of semi-arbitrarily different binning schemes in IP and isolation are used whilst keeping the default $1/p_{\mathrm{T}}$ binning. The binning schemes and the resulting blinded $R_{\tau\mu}$ values are shown in Table 9.5. The number

of $1/p_{\mathrm{T}}$ bins is also increased/decreased by a factor of two, and extended with two extra bins at high p_{T}.

- **Fit freedom**: The number of floating parameters controlling the yields and form factors for HF and DIF was varied, as is shown in Tables 9.6 and 9.7 for HF and DIF respectively.
- **Effect of IP weights**: Turn off the weight-based correction at very high IP.
- **Beeston-Barlow Lite implementation**: To test that the Beeston-Barlow Lite procedure is correctly implemented, a fit variation is performed with it turned off.

Splitting by η gives two fits that are, respectively, around 2σ above and below the nominal result. This means there is an uncomfortable $\sim 4\sigma$ difference between the two fits. Splitting by polarity and charges $\times$ polarity gives less than 1σ shifts. The A- and C-side fits are below and above the nominal result by around 1σ, as are the fits to the top and bottom ϕ regions. The fit quality in all the above cases is similar. These cross-checks suggest that further study may be needed in the modelling of factors related to the η distribution, and indeed the η post-fit plot shows a trend of mismodelling outside the fit region which is slightly unsatisfactory.

Variations to the $1/p_{\mathrm{T}}$ fit range all cause deviations of less than 1σ, as do the variations to the isolation binning, as can be seen in Table 9.5. The variations to the fit's IP binning cause quite large shifts to $R_{\tau\mu}$: just shifting the edges within the existing upper and lower limits shifts $R_{\tau\mu}$ by $\sim 3\sigma$, while adding another bin in the mid-IP region causes around a 5σ shift (accounting for the statistical uncertainty reduction in this variation). All the variations to the number of isolation or IP bins produce similar or worse fit qualities, while the variation increasing the number of $1/p_{\mathrm{T}}$ bins by a factor of two does improve the fit quality markedly, to around $\chi^2/\mathrm{ndf} = 1.2$. The variations to the isolation and $1/p_{\mathrm{T}}$ bins have a reassuringly small effect, although the IP-related cross-checks lead to the same conclusions as made when discussing the IP-related post-fit plots and systematic uncertainty - more study and validation is needed here.

The variations to the number of HF and DIF free nuisance parameters have, in general, a reassuringly small effect. The small variation caused when turning off the HF and DIF $1/p_{\mathrm{T}}$ form factors suggests that it may be reasonable to run the nominal fit without these form factors, thereby simplifying the fit and decreasing the statistical uncertainty. Further validation of the HF $1/p_{\mathrm{T}}$ shape would be necessary in this case. Similarly, fixing the charge asymmetry of the HF components make little difference, and the number of nuisance parameters could perhaps also be reduced here. On the other hand, the fits show that the isolation form factors are absolutely necessary for both HF and DIF, and indeed the large difference that can be seen in $R_{\tau\mu}$ suggests further validation of HF and DIF isolation shape is necessary.

Finally, switching off the Beeston-Barlow Lite procedure, and the high-IP weight-based correction, make differences to $R_{\tau\mu}$ at only the 0.1-0.2% level, which is negligible and gives confidence in the methods. In the case of the IP weights, a large degradation in the fit quality - particularly in the highest IP bin - is seen. This is to be expected and motivates the usage of the weights while there is a large mismodelling of the signal components at high IP.

Table 9.5: Fit results with variations to the isolation and IP fit binning edges. The default edges are $[0, 1.5, 3.0, 5.0]$ GeV in isolation, and $[0, 0.08, 0.18, 0.3]$ mm in IP.

IP binning / mm	Isolation binning / GeV	χ^2_{tot}/ndf	$\Delta R_{\tau\mu}$	$R_{\tau\mu} - R_{\tau\mu,\text{def}}$
$[0, 0.08, 0.18]$	$[0, 1.5, 3.0, 5.0]$	170/107	0.044	-0.052
$[0.0, 0.05, 0.1, 0.18, 0.3]$	$[0, 1.5, 3.0, 5.0]$	430/227	0.026	0.135
$[0.0, 0.06, 0.15, 0.3]$	$[0, 1.5, 3.0, 5.0]$	304/167	0.029	0.096
$[0.0, 0.08, 0.18, 0.3]$	$[0, 1.5, 3.0]$	163/107	0.052	-0.017
$[0.0, 0.08, 0.18, 0.3]$	$[0, 1.0, 2.0, 3.0, 4.0, 5.0]$	423/287	0.032	0.004
$[0.0, 0.08, 0.18, 0.3]$	$[0, 1.5, 4.0, 6.0]$	289/167	0.031	-0.013

Table 9.6: Fit results with variations to number of free nuisance parameters concerning the heavy-flavour part of the model.

Variation	χ^2_{tot}/ndf	$\Delta R_{\tau\mu}$	$R_{\tau\mu} - R_{\tau\mu,\text{def}}$
$4 \to 2$ free yields: shared between charges	258/169	0.034	-0.010
$1 \to 2$ p_{T} form factors, shared between charges	256/166	0.036	0.001
$1 \to 2$ p_{T} form factors; shared between $c\bar{c}$ and $b\bar{b}$	250/166	0.035	-0.005
$1 \to 2$ iso. form factors, shared between charges	378/166	0.023	-0.299
$1 \to 2$ iso. form factors; shared between $c\bar{c}$ and $b\bar{b}$	255/166	0.035	-0.004

Table 9.7: Fit results with variations to number of free nuisance parameters concerning the DIF part of the model.

Variation	χ^2_{tot}/ndf	$\Delta R_{\tau\mu}$	$R_{\tau\mu} - R_{\tau\mu,\text{def}}$
$2 \to 1$ free yield: shared between charges	417/168	0.035	-0.007
$1 \to 2$ p_{T} form factors, shared between charges	240/166	0.035	-0.007
$1 \to 2$ iso. form factors, shared between charges	253/166	0.035	0.009
$1 \to 0$ p_{T} form factors	266/168	0.035	-0.010
$1 \to 0$ iso. form factors	732/168	0.031	0.345

9.10 Conclusion and outlook

In this chapter, the current status of the measurement of $R_{\tau\mu}$, a high-precision test of the Standard Model's lepton flavour universality, is presented. The measurement is close to completion. The current blinded result is

$$R_{\tau\mu} = 0.997 \pm 0.035_{\text{stat}} \pm 0.047_{\text{syst}} = 0.997 \pm 0.059$$

In many ways, the analysis strategy has been driven by the size of the systematic uncertainties that have been observed during development. This is best exemplified by the fitting strategy itself: fitting into the high-isolation and high-IP regions where there are hardly any signal muons is necessary to control the HF and DIF backgrounds. These backgrounds are now largely understood, as was shown in the previous sections. The analysis strategy has been tested thoroughly, and the evaluation of systematic uncertainties is in an advanced state.

To give confidence in the modelling choices and that any residual biases due to them are small, it would be desirable to have a statistically-limited measurement i.e. that the total systematic uncertainty is smaller than the statistical uncertainty. At the time of writing, the analysis is therefore being held back from completion by the incomplete understanding of LHCb's IP resolution and how it is mismodelled in the simulation, as quantified by the leading IP smearing parameter uncertainty. The post-fit projections of $N_{\text{tracks}}^{\text{BPV}}$ and χ^2_{IP} currently also show significant mismodelling, the importance and origins of which are not yet understood. It would also be desirable to have a better fit quality, although it has been argued that this is mostly aesthetic and will not have a large effect on $R_{\tau\mu}$ alone. Fixing modelling issues in IP and isolation, which themselves may have an effect on $R_{\tau\mu}$, may lead to an improvement in fit quality.

If the IP smearing parameter uncertainty can be reduced, the rest of the systematic uncertainties have been shown to be small in comparison. The measurement will therefore currently be limited by its statistical uncertainty of 3.5%. A number of choices can be made that will affect this: the $(\text{IP}, \mathcal{I})$ binning could be extended; the number of floating nuisance parameters could be reduced; and the η cut could be relaxed. In principle, fitting in χ^2_{IP} should give more separation between the signals and HF components, which will also reduce the statistical uncertainty, but the aforementioned mismodelling issues would need to resolved first. The analysis is set up to be able to make of these changes easily. Notwithstanding these potential improvements, a statistically-limited measurement at around 3-4% accuracy would be around a factor of two less precise than the recent measurements by ATLAS and CMS, but nonetheless still an important verification (or otherwise) of their measurements in a complementary channel. Looking forward, the

LHCb 2017 and 2018 data is also available for a potential extension to this analysis that could proceed in parallel to a m_W measurement using the full Run 2 dataset. Inclusion of this data should reduce the statistical uncertainty by around a third. In contrast, ATLAS has already used all of their Run 2 data.

CHAPTER 10

Conclusion

This thesis has presented two high-precision measurements of the electroweak sector of the Standard Model using data collected in 2016 by the LHCb experiment: the measurement of the W boson mass; and a test of its lepton flavour universality via the ratio $R_{\tau\mu}$. In addition, the author's contribution to the LHCb trigger in Run 3 was described.

The software tool `HltEfficiencyChecker` developed by the author is crucial to the effort of building an optimized High Level Trigger for LHCb in Run 3; one that fully exploits the higher collision luminosity and new detector to have the best possible physics reach, whilst also conforming to the constraints placed on a trigger by a modern particle physics experiment. It gives the crucial metrics of trigger rate and efficiency in a variety of outputs that easily configurable and well-documented. The underlying code is transparent and automated, and allows for test-driven development of trigger lines by their authors. In many ways, this tool greatly improves the process of trigger line authorship in LHCb. This has been showcased here, both in how the tool was used to aid the collaboration decision towards a GPU-based first-level software trigger, and in the author's development of trigger selections for electroweak physics in Run 3.

The m_W measurement - the first at LHCb - successfully showcased that LHCb is not just a flavour physics experiment, but also that it is very much a player in precision electroweak physics. Although alone the precision of 32 MeV achieved does not rival the world's best measurements of this important quantity, it lights the way for a further measurement using all of LHCb's Run 2 data, where a 20 MeV uncertainty is targeted. The author's primary contribution to this ground-breaking study was a complex and comprehensive treatment of the muon reconstruction efficiencies and their modelling in the simulation, which was a key aspect in the measurement's success. The follow-up m_W measurement is already underway, and remarkably, it would make LHCb a rival for the most precise measurement at the LHC, if the predicted precision can be delivered.

There is also large potential value in combining the LHCb measurement of m_W with those from ATLAS and CMS, as their respective parton-distribution-function-related uncertainties will anti-correlate to a degree. The value of m_W obtained is consistent with previous measurements, and the Standard Model via the predictions of global electroweak fits, which means it stands in conflict with the recent measurement of m_W by the CDF collaboration. After the arrival of this ostensibly Standard-Model-defying measurement, the focus now turns back to the LHC experiments to see if they can agree or disagree at a similar level of precision. With this 2016 measurement recently completed, LHCb is perhaps in the best position to be able to provide answers to the latest high-profile quandary of particle physics.

Today, the Standard Model's axiom of lepton flavour universality is under great threat by the many hints of discrepancies found in measurements from experiments like LHCb. At the beginning of the work undertaken towards this thesis, the so-called "LEP anomaly" - a noticeable deviation in the assumed universality of the W bosons decay rates to tau leptons and muons - had also stood amongst these discrepancies, and measurements from the LHC were needed to understand the nature of it. Although unfinished at time of writing, the measurement of $R_{\tau\mu}$ weighs in its perspective to the LEP anomaly, with the potential to do so at a competitive precision to recent measurements by ATLAS and CMS published during the work, if the small number of remaining modelling issues can be satisfactorily resolved. It is an important validation of those measurements, in a complementary decay channel, and another example of an analysis thought hitherto beyond the reach of the LHCb experiment. This analysis will be finalized in the coming months, at which point it will join the rich history of LHC (and LHCb) investigations into this area of the Standard Model, which perhaps is the closest to the precipice of unveiling new physics.

References

[1] Super-Kamiokande Collaboration, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562.

[2] 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.

[3] 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.

[4] LHCb, J. Alves, A.Ãugusto *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[5] W. Barter, *Z boson and associated jet production at the LHCb experiment*, PhD thesis, Emmanuel College, University of Cambridge, 2014, CERN-THESIS-2014-178.

[6] S. Glazov, *Experimental Results from the LHC*, Lecture Notes. CTEQ/MCnet School 2021 (remote), 2021.

[7] T. Schörner-Sadenius, ed., *The Large Hadron Collider: Harvest of Run 1*, Springer, Berlin, 2015.

[8] C. Lester, *Part III Particle Physics*, Lecture Notes. Cavendish Laboratory, University of Cambridge, 2017.

[9] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[10] B. Gripaios, *Part III Gauge Field Theories*, Lecture Notes. Cavendish Laboratory, University of Cambridge, 2018.

[11] I. J. R. Aitchison and A. J. G. Hey, *Gauge theories in particle physics: A practical introduction. Vol. 1: From relativistic quantum mechanics to QED*, CRC Press, Bristol, UK, 4th ed., 2012.

[12] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA, 1995.

[13] T. Lancaster and S. J. Blundell, *Quantum Field Theory for the Gifted Amateur*, Oxford University Press, 2014.

[14] L. Harland-Lang, *Collider Phenomenology*, Lecture Notes. STFC HEP Summer School 2019, IPPP, Durham, 2019.

[15] S. Prestel, *MC event generator introduction*, Lecture Notes. CTEQ/MCnet School 2021 (remote), 2021.

[16] V. Bertone, *Theoretical and Phenomenological introduction to Drell-Yan production in QCD*, Lecture Notes. CTEQ/MCnet School 2021 (remote), 2021.

[17] D. E. Soper, *Basics of QCD Perturbation Theory*, Lecture Notes. CTEQ/MCnet School 2021 (remote), 2021.

[18] T. Potter, *Part II Particle and Nuclear Physics*, Lecture Notes. Cavendish Laboratory, University of Cambridge, 2019.

[19] C. S. Wu *et al.*, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413.

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

[21] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[22] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367.

[23] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, [,160(1964)].

[24] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, [,157(1964)].

[25] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585, [,162(1964)].

[26] D. Binosi and L. Theußl, *Jaxodraw: A graphical user interface for drawing feynman diagrams*, Computer Physics Communications **161** (2004) 76.

[27] T. D. Lee and M. Nauenberg, *Degenerate systems and mass singularities*, Phys. Rev. **133** (1964) B1549.

[28] T. Kinoshita, *Mass singularities of Feynman amplitudes*, J. Math. Phys. **3** (1962) 650.

[29] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*, Cambridge University Press, Cambridge, 2003.

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

[31] 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.

[32] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126** (1977) 298.

[33] NNPDF, R. D. Ball *et al.*, *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77** (2017) 663, arXiv:1706.00428.

[34] I. Hinchliffe and A. Kwiatkowski, *Parton model sum rules*, Annual Review of Nuclear and Particle Science **46** (1996) 609.

[35] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[36] CMS, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[37] S. Farry, *Electroweak Physics at LHCb*, Nuclear and Particle Physics Proceedings **273-275** (2016) 2181, 37th International Conference on High Energy Physics (ICHEP).

[38] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189, arXiv:0901.0002.

[39] J. Pumplin *et al.*, *Uncertainties of predictions from parton distribution functions. II. The Hessian method*, Physical Review D **65** (2001) .

[40] O. Behnke, K. Kröninger, T. Schörner-Sadenius, and G. Schott, eds., *Data analysis in High Energy Physics*, Wiley-VCH, Weinheim, Germany, 2013.

[41] J. Butterworth *et al.*, *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv:1510.03865.

[42] R. Gauld *et al.*, *Precise predictions for the angular coefficients in Z-boson production at the LHC*, JHEP **11** (2017) 003, arXiv:1708.00008.

[43] S. D. Drell and T.-M. Yan, *Massive lepton-pair production in hadron-hadron collisions at high energies*, Phys. Rev. Lett. **25** (1970) 316.

[44] L. Barze *et al.*, *Neutral current Drell-Yan with combined QCD and electroweak corrections in the POWHEG BOX*, Eur. Phys. J. C **73** (2013) 2474, arXiv:1302.4606.

[45] L. Barze *et al.*, *Implementation of electroweak corrections in the POWHEG BOX: single W production*, JHEP **04** (2012) 037, arXiv:1202.0465.

[46] T. Sjöstrand *et al.*, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[47] M. Bahr *et al.*, *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv:0803.0883.

[48] S. Camarda *et al.*, *DYTurbo: fast predictions for Drell–Yan processes*, Eur. Phys. J. C **80** (2020) 251, arXiv:1910.07049, [Erratum: Eur.Phys.J.C 80, 440 (2020)].

[49] J. C. Collins and D. E. Soper, *Angular distribution of dileptons in high-energy hadron collisions*, Phys. Rev. D **16** (1977) 2219.

[50] L. R. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001. 164 p, This report is an abridged version of the LHC Design Report (CERN-2004-003).

[51] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013.

[52] M. Bajko *et al.*, *Report of the Task Force on the Incident of 19th September 2008 at the LHC*, LHC-PROJECT-Report-1168, CERN-LHC-PROJECT-Report-1168, CERN, Geneva, 2009.

[53] J. Wenninger, *Operation and Configuration of the LHC in Run 2*, CERN-ACC-NOTE-2019-0007, 2019.

[54] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[55] O. S. Brüning *et al.*, *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004.

[56] LHCb, S. Amato *et al.*, *LHCb technical proposal: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays*, CERN-LHCC-98-04, CERN-LHCC-98-4, CERN-LHCC-P-4, 1998.

[57] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030, 2003.

[58] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[59] DELPHI, P. A. Aarnio *et al.*, *The DELPHI detector at LEP*, Nucl. Instrum. Meth. A **303** (1991) 233.

[60] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[61] LHCb collaboration, C. Elsässer, *$\bar{b}b$ production angle plots*, Online; accessed August 2022.

[62] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007, 2000.

[63] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, arXiv:1708.00819.

[64] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[65] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346.

[66] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[67] LHCb Collaboration, R. Aaij *et al.*, *Precise determination of the B^0_s–$\overline{B}^0_s$ oscillation frequency*, Nature Phys. **18** (2021) 1, arXiv:2104.04421.

[68] LHCb Silicon Tracker Group, C. Abellan Beteta *et al.*, *Monitoring radiation damage in the LHCb Tracker Turicensis*, JINST **15** (2020) P08016, arXiv:1809.05063.

[69] C. Elsasser *et al.*, *The LHCb Silicon Tracker*, CERN-LHCb-PROC-2013-056, 2013, On behalf of the LHCb Silicon Tracker Group.

[70] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[71] C. Lippmann, *Particle identification*, Nucl. Instrum. Meth. A **666** (2012) 148, arXiv:1101.3276.

[72] G. F. Knoll, *Radiation Detection and Measurement*, John Wiley and Sons, New York, 3rd edition ed., 2000.

[73] T. Bellunato, *Development of Ring Imaging Cherenkov Detectors for LHCb*, PhD thesis, Milan, Università degli Studi di Milano, 2003, CERN-THESIS-2009-038.

[74] R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2*, LHCb-DP-2021-004, 2022.

[75] LHCb, *LHCb: RICH technical design report*, CERN-LHCC-2000-037, 2000.

[76] TA2, M. Alemi *et al.*, *First operation of a hybrid photon detector prototype with electrostatic cross-focussing and integrated silicon pixel readout*, Nucl. Instrum. Meth. A **449** (2000) 48.

[77] J. B. Birks, *The Theory and practice of scintillation counting*, 1964.

[78] C. Abellán Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, arXiv:2008.11556.

[79] LHCb collaboration, *LHCb muon system: second addendum to the Technical Design Report*, Technical design report. LHCb, CERN, Geneva, 2005. CERN-LHCC-2005-012. Submitted on 9 Apr 2005.

[80] G. Dujany and B. Storaci, *Real-time alignment and calibration of the LHCb Detector in Run II*, J. Phys. Conf. Ser. **664** (2015) 082010.

[81] LHCb, S. Borghi, *Novel real-time alignment and calibration of the LHCb detector and its performance*, Nucl. Instrum. Meth. A **845** (2017) 560.

[82] 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.

[83] E. Aslanides *et al.*, *The Level-0 muon trigger for the LHCb experiment*, Nucl. Instrum. Meth. A **579** (2007) 989, arXiv:0705.0310.

[84] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, LHCb-PUB-2011-001, CERN-LHCb-PUB-2011-001, 2011.

[85] R. Aaij *et al.*, *A comprehensive real-time analysis model at the LHCb experiment*, JINST **14** (2019) P04006, arXiv:1903.01360.

[86] LHCb collaboration, *Computing Model of the Upgrade LHCb experiment*, CERN-LHCC-2018-014, 2018.

[87] LHCb Collaboration, R. Aaij *et al.*, *Impact of the $Z \to \mu^+\mu^-$ alignment on the Z mass resolution at LHCb*, LHCB-FIGURE-2020-009, 2020.

[88] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.

[89] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[90] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[91] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[92] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[93] LHCb Collaboration, R. Aaij *et al.*, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016. LHCB-TDR-016, 2014.

[94] A. Piucci, *The lhcb upgrade*, Journal of Physics: Conference Series **878** (2017) 012012.

[95] Hunter, Ross *et al.*, *An automated tool to facilitate consistent test-driven development of trigger selections for LHCb's Run 3*, EPJ Web Conf. **251** (2021) 04024.

[96] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[97] LHCb collaboration, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016, 2014.

[98] LHCb collaboration, *LHCb Upgrade GPU High Level Trigger Technical Design Report*, CERN-LHCC-2020-006, 2020.

[99] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039, CERN-LHCb-PUB-2014-039, 2014.

[100] R. Aaij and et al. *Measurement of the track reconstruction efficiency at lhcb*, Journal of Instrumentation **10** (2015) P02007–P02007.

[101] R. Aaij *et al.*, *Allen: A high level trigger on GPUs for LHCb*, Comput. Softw. Big Sci. **4** (2020) 7, arXiv:1912.09161.

[102] R. Aaij *et al.*, *A Comparison of CPU and GPU Implementations for the LHCb Experiment Run 3 Trigger*, Computing and Software for Big Science **6** (2021) 1.

[103] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[104] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO vector-boson production matched with shower in POWHEG*, JHEP **07** (2008) 060, arXiv:0805.4802.

[105] J. Bellm *et al.*, *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv:1512.01178.

[106] ATLAS, G. Aad *et al.*, *Measurement of the angular coefficients in Z-boson events using electron and muon pairs from data taken at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, JHEP **08** (2016) 159, arXiv:1606.00689.

[107] LHCb Collaboration, R. Aaij *et al.*, *First measurement of the* $Z \to \mu^+\mu^-$ *angular coefficients in the forward region of pp collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Rev. Lett. **129** (2022) 091801. 11 p, arXiv:2203.01602.

[108] E. Barberio, B. van Eijk, and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. **66** (1991) 115.

[109] S. Höche and S. Prestel, *The midpoint between dipole and parton showers*, Eur. Phys. J. C **75** (2015) 461, arXiv:1506.05057.

[110] N. Fischer, S. Prestel, M. Ritzmann, and P. Skands, *Vincia for Hadron Colliders*, Eur. Phys. J. C **76** (2016) 589, arXiv:1605.06142.

[111] T.-J. Hou *et al.*, *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, Phys. Rev. D **103** (2021) 014013, arXiv:1912.10053.

[112] S. Bailey *et al.*, *Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs*, arXiv:2012.04684.

[113] H.-L. Lai *et al.*, *Parton Distributions for Event Generators*, JHEP **04** (2010) 035, arXiv:0910.4183.

[114] LHCb collaboration, R. Aaij *et al.*, *Study of forward Z +jet production in pp collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **01** (2014) 033, arXiv:1310.8197.

[115] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2018) 1, arXiv:1803.00824.

[116] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[117] W. Barter, M. Pili, and M. Vesterinen, *A simple method to determine charge-dependent curvature biases in track reconstruction in hadron collider experiments*, Eur. Phys. J. C **81** (2021) 251, arXiv:2101.05675.

[118] M. Pili, *Measurement of the W boson mass with the LHCb experiment*, PhD thesis, 2021, Presented 08 Oct 2021.

[119] A. Banfi *et al.*, *Optimisation of variables for studying dilepton transverse momentum distributions at hadron colliders*, Eur. Phys. J. C **71** (2011) 1600, arXiv:1009.1580.

[120] M. Needham and T. Ruf, *Estimation of the material budget of the LHCb detector*, LHCb-2007-025, CERN-LHCb-2007-025, 2007.

[121] T. W. Anderson and D. A. Darling, *Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes*, Ann. Math. Statist. **23** (1952) 193.

[122] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[123] R. Hagedorn, *Multiplicities, p_{T} Distributions and the Expected Hadron $\to$ Quark-Gluon Phase Transition*, Riv. Nuovo Cim. **6N10** (1983) 1.

[124] LHCb, R. Aaij *et al.*, *Measurement of the W boson mass*, JHEP **01** (2022) 036, arXiv:2109.01113.

[125] J. Van Tilburg, *Track simulation and reconstruction in LHCb*, PhD thesis, NIKHEF, Amsterdam, 2005, CERN-THESIS-2005-040.

[126] M. Awramik, M. Czakon, A. Freitas, and G. Weiglein, *Precise prediction for the W boson mass in the standard model*, Phys. Rev. D **69** (2004) 053006, arXiv:hep-ph/0311148.

[127] A. Sirlin, *Radiative corrections in the $\mathrm{SU}(2)_L \times \mathrm{U}(1)$ theory: A simple renormalization framework*, Phys. Rev. D **22** (1980) 971.

[128] J. Haller *et al.*, *Update of the global electroweak fit and constraints on two-Higgs-doublet models*, Eur. Phys. J. **C78** (2018) 675, arXiv:1803.01853.

[129] J. de Blas *et al.*, *Global analysis of electroweak data in the Standard Model*, Phys. Rev. D **106** (2022) 033003, arXiv:2112.07274.

[130] ATLAS, M. Aaboud *et al.*, *Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C78** (2018) 110, arXiv:1701.07240, [Erratum:Eur. Phys.J.C78,no.11,898(2018)].

[131] ALEPH, DELPHI, L3, OPAL, LEP Electroweak, S. Schael *et al.*, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, Phys. Rept. **532** (2013) 119, arXiv:1302.3415.

[132] CDF, D0, T. A. Aaltonen *et al.*, *Combination of CDF and D0 W-boson mass measurements*, Phys. Rev. D **88** (2013) 052018, arXiv:1307.7627.

[133] S. Farry, O. Lupton, M. Pili, and M. Vesterinen, *Understanding and constraining the PDF uncertainties in a W boson mass measurement with forward muons at the LHC*, Eur. Phys. J. C **79** (2019) 497, arXiv:1902.04323.

[134] G. Bozzi, L. Citelli, M. Vesterinen, and A. Vicini, *Prospects for improving the LHC W boson mass measurement with forward muons*, Eur. Phys. J. C **75** (2015) 601, arXiv:1508.06954.

[135] ATLAS, G. Aad *et al.*, *Measurement of the Transverse Momentum Distribution of W Bosons in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector*, Phys. Rev. D **85** (2012) 012005, arXiv:1108.6308.

[136] CMS, V. Khachatryan *et al.*, *Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV*, JHEP **02** (2017) 096, arXiv:1606.05864.

[137] O. Lupton and M. Vesterinen, *Simultaneously determining the $W^{\pm}$ boson mass and parton shower model parameters*, arXiv:1907.09958.

[138] LHCb collaboration, R. Aaij *et al.*, *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV*, JHEP **09** (2016) 136, arXiv:1607.06495.

[139] F. James, *MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1* CERN-D-506, CERN-D506, 1994.

[140] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219.

[141] CDF Collaboration and D0 Collaboration and Tevatron Electroweak Working Group, V. M. Abazov *et al.*, *Combination of CDF and D0 results on the W boson mass and width*, Phys. Rev. D **70** (2004) 092008.

[142] D0, V. M. Abazov *et al.*, *Measurement of the W Boson Mass with the D0 Detector*, Phys. Rev. Lett. **108** (2012) 151804, arXiv:1203.0293.

[143] CDF, T. Aaltonen *et al.*, *High-precision measurement of the W boson mass with the CDF II detector*, Science **376** (2022) 170.

[144] A. Denner, *Techniques for calculation of electroweak radiative corrections at the one loop level and results for W physics at LEP-200*, Fortsch. Phys. **41** (1993) 307, arXiv:0709.1075.

[145] B. A. Kniehl, F. Madricardo, and M. Steinhauser, *Gauge-independent W-boson partial decay widths*, Physical Review D **62** (2000) .

[146] D. d'Enterria and V. Jacobsen, *Improved strong coupling determinations from hadronic decays of electroweak bosons at N^3LO accuracy*, arXiv:2005.04545.

[147] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group, S. Schael *et al.*, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, arXiv:hep-ex/0509008.

[148] UA1, C. Albajar *et al.*, *Studies of Intermediate Vector Boson Production and Decay in UA1 at the CERN Proton - Antiproton Collider*, Z. Phys. C **44** (1989) 15.

[149] CDF, A. Abulencia *et al.*, *Measurements of inclusive W and Z cross sections in $p\overline{p}$ collisions at $\sqrt{s}$ = 1.96 TeV*, J. Phys. G **34** (2007) 2457, arXiv:hep-ex/0508029.

[150] D0 Collaboration, S. Abachi *et al.*, *W and Z Boson Production in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. Lett. **75** (1995) 1456.

[151] ATLAS, M. Aaboud *et al.*, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 367, arXiv:1612.03016.

[152] LHCb collaboration, R. Aaij *et al.*, *Measurement of forward $W \to e\nu$ production in pp collisions at $\sqrt{s}$ =8 TeV*, JHEP **10** (2016) 030, arXiv:1608.01484.

[153] Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, arXiv:2206.07501.

[154] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of the $\mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)$ and $\mathcal{B}(B^0 \to D^{*-}\mu^+\nu_\mu)$ branching fractions using three-prong τ-lepton decays*, Phys. Rev. Lett. **120** (2018) 171802, arXiv:1708.08856.

[155] Belle Collaboration, G. Caria *et al.*, *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a Semileptonic Tagging Method*, Phys. Rev. Lett. **124** (2020) 161803.

[156] The BABAR Collaboration, J. P. Lees *et al.*, *Measurement of an excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays and implications for charged higgs bosons*, Phys. Rev. D **88** (2013) 072012.

[157] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[158] LHCb, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, arXiv:2103.11769.

[159] LHCb Collaboration, R. Aaij *et al.*, *Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay*, Phys. Rev. Lett. **125** (2020) 011802.

[160] ATLAS, G. Aad *et al.*, *Test of the universality of τ and μ lepton couplings in W-boson decays with the ATLAS detector*, Nature Phys. **17** (2021) 813, arXiv:2007.14040.

[161] CMS, A. Tumasyan *et al.*, *Precision measurement of the W boson decay branching fractions in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Phys. Rev. D **105** (2022) 072008, arXiv:2201.07861.

[162] Particle Data Group, R. L. Workman and Others, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.