Lancaster University

Physics Department

A test of Lepton Flavour Universality by comparing the decay widths of W bosons produced in top quark pair events with the ATLAS detector at the LHC

Submitted for examination 30/11/20

Under the Supervision of Guennadi Borissov

Acknowledgements

Declaration

This thesis is my own work and has not been submitted for award elsewhere.

The work contained in this thesis would not be possible without international teamwork and collaboration. I was fortunate to work both as part of the wider ATLAS collaboration, as well as in smaller analysis subgroups. As such this work should be considered joint research. My individual contributions to these research and findings are outlined in the following:

- Chapter 3 (Improvements to primary vertex reconstruction in ATLAS), Sections:
    - 3.7.1: vertex migration studies,
    - 3.8 (excluding 3.8.7): vertex reconstruction performance studies.
- Chapter 4 (Test of Lepton Flavour Universality in top quark decays in ATLAS), Sections:
    - 4.5: event selection optimisation,
    - 4.6 (excluding 4.6.6) (and Appendices B, C.1, C.2): muons from hadron decays background scale factor calculations,
    - 4.7 (and Appendices B) : $Z^0(\to \mu\mu) + (b\text{-})$jets background scale factor calculations.

The results of the vertex reconstruction studies have been published as a Public note Ref. [1]) and conference proceedings for Connecting the Dots Conference 2019 [2]. The results of the studies for the test of Lepton Flavour Universality in top quark decays in ATLAS analysis are included in Ref. [3], which has been accepted for publication by Nature Physics (as of 30/3/21). They are also published in proceedings from the Beauty 2020 conference [4].

Abstract

At this time, the Standard Model of particle physics represents our best understanding of physics at the smallest scales. A fundamental axiom of this theory is the universality of the couplings between the different lepton generations and the gauge bosons. This principle is known as Lepton Flavour Universality and can be tested by comparing the decay widths of (semi-)leptonic processes that differ only in lepton flavour. Such a comparison of the decay widths of $W^\pm$ bosons to taus and muons can be quantified by $R(\tau/\mu) = \mathrm{Br}(W^\pm \to \tau\nu)/\mathrm{Br}(W^\pm \to \mu\nu)$. A measurement of $R(\tau/\mu)$ is presented in this thesis. The measurement uses $W^\pm$ bosons produced in $t\bar{t}$ decays with a dileptonic decay mode. The measurement is based on 139 fb^{-1} of pp collision data recorded with the ATLAS detector at the Large Hadron Collider (LHC) at $\sqrt{s} = 13$ TeV. In this analysis, muons originating from $W^\pm$ bosons and those originating via an intermediate tau are distinguished by the lifetime of the tau, using the transverse impact parameter of the muons, as well as differences in the transverse momenta spectra of the muons. The observed best-fit value of $R(\tau/\mu)$ is $0.992 \pm 0.013 [\pm 0.007(\mathrm{stat}) \pm 0.011(\mathrm{syst})]$ and is therefore in agreement with the hypothesis of universal lepton couplings, as postulated in the Standard Model. This is the most precise measurement of $R(\tau/\mu)$, surpassing the previous LEP measurement by a factor 2. This result is consistent with the LEP measurement at the level of 2.7 standard deviations.

Such results would not be possible without the continued performance of the ATLAS detector. Primary vertices represent the locations of proton-proton collisions – the ultimate origins of all reconstructed objects used in physics analyses. They are therefore key to understanding the full kinematics of an interaction. However, increasing luminosity poses a challenge for primary vertex reconstruction in ATLAS. The recently finished Run-2 observed a rate of 60 or more proton-proton collisions per beam crossing, and an even higher vertex density is expected in future. As such, ATLAS has developed new tools: a Gaussian track density seed finder and an adaptive multi-vertex finder. The seed finder locates the position of candidate vertices using an analytic model of the track density along the beam axis, and the vertex finder applies a global approach to vertex finding and fitting that allows charged tracks to be assigned to their optimal vertex candidates. This thesis presents studies of the optimisation and exploitation of these tools for the vertex densities expected in the upcoming Run-3 and beyond.

List of Figures

viii

x

List of Tables

Contents

Introduction

We are living in a golden age of experimental particle physics. Through international cooperation, we are able to generate and analyse data at a rate that would have been inconceivable to the great minds who first opened this field of research. The Standard Model of particle physics is well validated by this experimental data, including the discovery of the Higgs boson by ATLAS [17] and CMS [18] in 2012, the top quark in 1995 at Fermilab [19, 20] and the $W^\pm$ [21, 22] and Z^0 [23, 24] bosons in 1983 at CERN. Whilst the Standard Model is well supported by all this data, it is not considered to be a complete description of the universe. The Standard Model does not provide candidates for a dark matter particle, nor does it describe gravity. It does not fully explain the observed matter-anti-matter asymmetry in the universe, nor does it justify the tiny masses of neutrinos. As such, searches for "new physics", that is to say physics not described in the Standard Model, are well motivated.

One such avenue of investigation is Lepton Flavour Universality. This is the *assumption* of the Standard Model that, in the massless limit, the vector gauge bosons couple equally to the three generations of leptons. They have individual couplings to the Higgs boson as indicated by their different masses.

Lepton Flavour Universality is an interesting line of research, since the theory does not *require* this to be the case. An observation that any of the three leptons generations experience individual couplings would therefore be new physics. Equally, confirmation of universal couplings raises the question, why?

Lepton Flavour Universality can be tested in leptonic and semileptonic processes that differ only in lepton flavour by comparing their decay widths. The most precise measurements of Lepton Flavour Universality in $W^\pm$ boson decays are the combined LEP measurements. These results show that the relative coupling of electrons and muons are the same (within the experimental uncertainty), in agreement with the Standard Model assumption. However, the relative coupling of taus to each of the light leptons showed tension with the Standard Model expectations. In particular, the test of universality between taus and muons is in agreement with the Standard Model at the level of 2.7 standard deviations. Furthermore, many recent tests of Lepton Flavour Universality from *b*-factory experiments also show tension with the Standard Model expectations.

Thus further tests of Lepton Flavour Universality are well motivated. Between 2015 to 2018, the ATLAS experiment amassed a dataset of pp collisions that includes over 100 million top quark pairs. Almost all of these top quarks decay to a bottom quark and a $W^{\pm}$ boson. By comparing the relative rates of the decays of $W^{\pm}$ bosons to taus and to muons, a precise test of Lepton Flavour Universality can be made.

This thesis will present a direct measurement of Lepton Flavour Universality using $W^{\pm}$ bosons produced in top-quark decays. The analysis is performed using 139 fb^{-1} of pp collision data collected between 2015-2018 by the ATLAS detector at the Large Hadron Collider (LHC). The thesis proceeds as follows: chapter 1 gives an overview of the relevant theory, before describing the current state of Lepton Flavour Universality measurements, then describing the relevant top quark production and decay results. Chapter 2 describes the LHC and the ATLAS detector, which together comprise experimental setup. Chapter 3 describes the work undertaken in improving primary vertex reconstruction software for the increasing luminosity in the ATLAS detector. Chapter 4 describes the analysis testing Lepton Flavour Universality in top quarks.

1

Background and motivation

This section covers the relevant theory to understand Lepton Flavour Universality and describes how it has been experimentally tested thus far. Section 1.1 provides a very brief overview of the Standard Model, followed by section 1.2, which describes the theory of weak charged current interactions in more detail. Section 1.3 covers the current status of Lepton Flavour Universality experimental tests. Section 1.4 provides a brief overview of the strong interaction. Finally, section 1.5 describes the production and decays of top quarks, an abundant source of on-shell $W^{\pm}$ bosons.

1.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) describes the fundamental particles and their interactions. It incorporates electromagnetism, the weak interaction and the strong interaction, but not gravity. It embodies our current best understanding of particle physics. In the Standard Model, the particle physics phenomena observed in experiments are described using quantum field theory.

The Standard Model describes two classes of fundamental particles, shown in Figure 1.1. The first are fermions, which are matter particles. The second are the bosons, which are subdivided into the force-carrying vector bosons and the scalar Higgs boson. The Standard Model describes how particle states change and how particles interact. It has proven to be a very successful model since its inception, both at describing the phenomena recorded in experimental data as well as being able to predict previously unseen phenomena. For every particle there is an antiparticle partner, though some particles are their own antiparticle. Antiparticles have the opposite electric charge to their particle counterparts, as well as some other quantum numbers such as colour, but appear identical in all other respects. For this reason it can be assumed in this section that the descriptions are general to both matter and antimatter, unless explicitly stated otherwise.
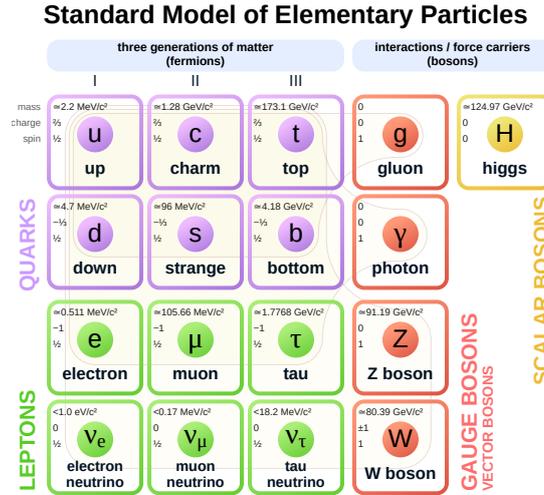
Figure 1.1: The fundamental matter particles. The fermions are shown on the left, with the quarks on top and leptons below. The gauge and Higgs bosons are shown on right. Outlines connect the gauge bosons with the particles they couple to. This image is taken from Wikipedia, and is public domain.

As shown in Figure 1.1, the Standard Model describes six quarks: up, down, charm, strange, top and bottom. The "up-type" quarks are the up quark, charm quark and top quark (which is sometimes called "truth" quark), whilst the "down-type" quarks are the down quark, the strange quark and the bottom quark (which is sometimes called "beauty" quark). The quarks are the only fermions to interact via the strong interaction, and they also interact via the weak interaction and electromagnetism. Due to colour confinement in the strong interaction they are never seen independently, only in bound states of two or more quarks [25]. These bound states are called hadrons. A hadron with an even number of quarks (usually a quark-antiquark pair) is known as a meson, whilst a hadron made up of an odd number of quarks (usually three) is known as a baryon. The up-type quarks have a fractional electric charge of $+2/3$ whilst the down-type quarks have a fractional electric charge of $-1/3$. In spite of the fractional charges of quarks, or rather because of them, hadrons always have integer electric charge. Heavier quarks decay principally via charged current weak interactions. Since this necessitates a change in electric charge of ± 1 this implies that "up-type" quarks decay to "down-type" and vice-versa. Flavour changing neutral currents are highly suppressed in the Standard Model [26]. The top-quark is the heaviest quark. Due to the high mass of the top quark it is very unstable and decays before it can hadronise, with a lifetime of just 5×10^{-25} s. It almost always decays to a bottom quark and $W^{\pm}$ boson. Due to its high mass, the $W^{\pm}$ bosons are produced "on-shell" which makes top-quark decays a useful probe for weak charged current analyses.

The other family of fermions are the leptons. These are divided into the charged leptons and neutrinos. The charged leptons are the electron, muon and tau. The neutrinos are

named according to the charged lepton they produce in charged current interactions. As their name suggests, the charged leptons interact via electromagnetism and the weak interaction, whilst neutrinos interact exclusively via the weak interaction. In individual interactions lepton number is preserved both overall and by generation. Oscillation between the three neutrino flavours has been observed in free neutrinos, which violates generational lepton number but not overall lepton number. In the Standard Model, the weak interaction is assumed to couple equally to each generation of charged lepton in the massless limit. This is known as Lepton Flavour Universality and is discussed in more depth in Section 1.3.

The fundamental interactions are communicated via the vector bosons. The vector bosons of the strong interaction are gluons. These are massless (and electrically neutral) spin-1 particles. Whilst it is often convenient to refer to just one gluon, there are in fact eight independent types of gluon in QCD. Their name comes from the fact they "glue" the quarks together. The photon is also a massless, electrically neutral spin-1 particle. Photons communicate electromagnetism. Unlike gluons there is thought to be only one type of photon. The weak interaction is communicated via three vector bosons: the electrically charged W^+ and W^- which are one another's antiparticle, and the electrically neutral Z^0. The $W^\pm$ and Z^0 bosons, in contrast to gluons and photons, have mass. This mass arises from symmetry breaking. Due to their large mass, on-shell weak bosons are observed in high energy collisions and decay very quickly. The Higgs boson communicates the scalar Higgs field, which is responsible for the electroweak symmetry breaking resulting in bosons with mass. The Higgs boson was discovered in 2012 at the LHC by ATLAS [17] and CMS [18].

As previously mentioned, the Standard Model has proven to be a very successful model at describing observed phenomena, as well as successfully predicting phenomena. However, it is not considered complete – there are some considerable absences. These include gravity – is there a quantum theory of this interaction communicated by a vector boson (the graviton) [27]? There is also the question of dark matter and dark energy [28]. There is an abundance of evidence demonstrating that dark matter exists (see, for example, [29]) but it is not known if or how it interacts with SM matter. The observed matter-anti-matter asymmetry, resulting in a matter dominated universe, is as-yet an open question. Requiring CP-violation to a greater extent than what is observed in the quark sector, ongoing and future long-baseline neutrino oscillation experiments hope to determine the extent to which CP-violation is observed in the lepton sector.

1.2   The weak interaction

The weak interaction is one of the fundamental interactions. It is the only one of the fundamental interactions confirmed to interact with all of the fermions. Mediating the weak interaction are three bosons: the neutral Z^0 boson, and the charged W^+ and W^- bosons. "Neutral current" (NC) interactions involve a Z^0, whilst "charged current"
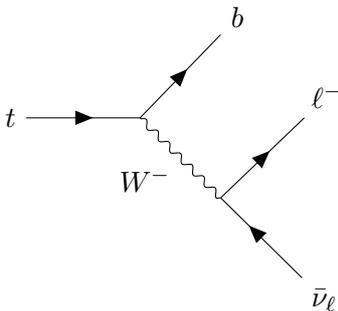
Figure 1.2: Lowest-order Feynman diagram of top quark decaying to a bottom quark and a W^- boson, with the W^- boson decaying to leptons. Here $\ell = e, \mu, \tau$.

(CC) interactions involve a $W^\pm$ boson. The substance of the physics analysis work in this thesis has a focus on the production of $W^\pm$ bosons in top quark decays, and subsequent decays of $W^\pm$ bosons into taus and muons. Therefore, a brief introduction to such weak interactions is required.

1.2.1   Structure of weak interactions of leptons

At low energies, charged leptons interact through either the electromagnetic or weak interactions, whilst at energies above 246 GeV, the so-called unification energy, they interact via the unified electroweak interaction. The electromagnetic interaction is well described by QED, as is the strong interaction by QCD. However, since the $W^\pm$ and Z^0 bosons of the weak interaction have mass, they require a different formulation to that of the photon or gluons. For brevity, only the charged current interaction will be described here.

A lowest order, charged current top-quark decay with a semileptonic final state is shown in Figure 1.2. The leptonic vertex factor associated with the $W^\pm$ boson is

$$\frac{-ig_W}{\sqrt{2}} \frac{1}{2} \gamma^\mu (1 - \gamma^5), \tag{1.1}$$

where g_W is the weak coupling constant, and γ^μ and γ^5 are gamma matrices.

This is a vector minus axial vector $(V - A)$ of the form $\gamma^\mu - \gamma^\mu \gamma^5$. If either the vector or axial vector component were zero, the weak interaction would conserve parity. It is observed that this is not the case [30]. As a result of this form, only left-handed chiral particle states and right-handed antiparticle states participate in the charged current interaction. In the limit $E \gg m$, chiral and helicity states are the same. The weak interaction coupling constant g_W is assumed to be the same for all generations of leptons. This is known as Lepton Flavour Universality, and is discussed in in more detail

in Section 1.3. The Feynman rule associated with the $W^\pm$ boson propagator is

$$\frac{-i}{q^2 - m_W^2}\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{m_W^2}\right),\tag{1.2}$$

where q is the 4-momentum of the $W^\pm$ boson, m_W is the $W^\pm$ boson mass, and $g_{\mu\nu}$ is the metric tensor.

1.2.2 Structure of weak interactions of quarks

The Cabbibo hypothesis is that the weak interaction of quarks have the same strength as the leptons, but the weak interaction eigenstates of quarks differ from their mass (flavour) eigenstates. The weak interaction eigenstates and mass eigenstates are related by a unitary matrix, the Cabibbo-Kobayashi-Masakawa (CKM) matrix [31]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.\tag{1.3}$$

The weak charged current vertices involving quarks are then given by:

$$-i\frac{g_W}{\sqrt{2}}(\bar{u},\bar{c},\bar{t})\gamma^\mu\frac{1}{2}(1-\gamma^5)\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix},$$

where, for example, b is a bottom-quark spinor and $\bar{t}$ is the adjoint spinor for a top-quark. The charge conjugate expression is

$$-i\frac{g_W}{\sqrt{2}}(\bar{d},\bar{s},\bar{b})\gamma^\mu\frac{1}{2}(1-\gamma^5)\begin{pmatrix} V_{ud}^* & V_{us}^* & V_{ub}^* \\ V_{cd}^* & V_{cs}^* & V_{cb}^* \\ V_{td}^* & V_{ts}^* & V_{tb}^* \end{pmatrix} \begin{pmatrix} u \\ c \\ t \end{pmatrix},$$

where, for example, $\bar{b} = b^\dagger\gamma^0$ is the adjoint spinor for a bottom quark and t is the spinor for a top quark.

The relative strength of a given quark-quark weak interaction is defined by the relevant element of the CKM matrix. For example, the weak charged current interaction associated with the tbW vertex in Figure 1.2 is

$$j_{bt}^\mu = -i\frac{g_W}{\sqrt{2}}V_{tb}^*\bar{b}\gamma^\mu\frac{1}{2}(1-\gamma^5)t.$$

The CKM matrix is analogous to the PMNS matrix for the weak interaction of leptons which describes neutrino flavour mixing, also known as neutrino oscillations. The CKM matrix is unitary and can be described by three rotations and a complex phase:

$$
V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta'} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta'} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.4)
$$

where $s_{ij} = \sin\phi_{ij}$ and $c_{ij} = \cos\phi_{ij}$.

Whilst the structure of the weak interactions of quarks is the same as that of leptons, the phenomenology is very different. Due to colour confinement, free quarks do not propagate, but instead hadronise on a length scale of 10^{-15} m [25]. Thus the final states of the weak interactions of quarks have to be described in terms of hadrons. It is the flavour eigenstates of the quarks that comprise the observable quantities in hadronic weak interactions. This means that the nine individual elements of the CKM matrix are measured separately. The measured values of the CKM matrix, as described in Ref. [32], are:

$$
V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.97446 \pm 0.0001 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}.
$$

$$(1.5)$$

The off diagonal elements of the CKM matrix are very small. This implies that the rotation angles between the quark flavour and weak interaction eigenstates are also small. This smallness leads to a near diagonal form.

The CKM matrix contains a CP-violating phase δ', which leads to a slight matter-anti-matter asymmetry. This was first observed in neutral kaon decays in 1964 by Cronin and Fitch [33] for which they were awarded a Nobel Prize in Physics in 1980.

In the SM, the branching fractions of leptonic $W^\pm$ decays are dependent on the six matrix elements $|V_{qq'}|$ of the CKM matrix excluding the top quark elements ($q \neq t$). The leptonic branching fraction of $W^\pm$, $\text{Br}(W \to \ell\bar{\nu}_\ell)$ in terms of these elements is

$$
\frac{1}{\text{Br}(W \to \ell\bar{\nu}_\ell)} = 3 \left\{ 1 + \left[1 + \frac{\alpha_s(\text{M}_\text{W}^2)}{\pi} \sum_{\substack{i=(u,c) \\ j=(d,s,b)}} |V_{ij}|^2 \right] \right\}
$$

where $\alpha_s(\text{M}_\text{W}^2)$ is the strong coupling constant and fermion mass effects are negligible.

1.3   Lepton Flavour Universality

Lepton Flavour Universality (LFU) is the name given to the assumption that, in the massless limit, the coupling strength of the SM lepton doublets to the $W^{\pm}$ boson is equal across the three generations. As described in Ref. [34], LFU is a consequence of the accidental flavour symmetry of the Standard Model Lagrangian. In the limit where the (small) lepton Yukawa couplings are neglected

$$\mathscr{L}_{\mathrm{SM}} = \mathscr{L}_{\mathrm{gauge}}(A_a, \psi_i) + \mathscr{L}_{\mathrm{Higgs}}(H, A_a, \psi_i),$$

there are three identical replica of the basic fermion family

$$[\psi = Q_L, u_R, d_R, L_L, e_R]$$

in the gauge sector, which means there is huge flavour degeneracy. There is no reason to assume this holds beyond the Standard Model. However, the principle of LFU has been verified with high precision in many measurements (see, for example, Ref. [11]). For this reason it is assumed a "sacred principle", with no deep reason behind it. No strong results from the third generation of quarks had been made before the b-factory results described below.

An analogy for why LFU has been observed at low energies (if it does not hold at high energies) is the following: if physicists could only probe atoms with long wavelength photons, it could be assumed that positrons and protons have identical properties except for their masses. This is the same argument used to infer LFU. The three families seem to be identical except for their mass. The SM quantum numbers could be an accidental low-energy property. The particles could well have different properties at higher energies, as indicated by their different masses. All flavour symmetries could well be only accidental low-energy properties, like isospin or $SU(3)$ in QCD.

The assumption of Lepton Flavour Universality in charged current interactions can be probed by comparing the measured decay widths of leptonic or semileptonic decays that differ only by lepton flavour. These measurements can be performed at both low momentum transfers (with virtual or "off-shell" $W^{\pm}$ bosons) and high momentum (with real or "on-shell" $W^{\pm}$ bosons). The measured quantities have the form $g_\ell/g_{\ell'}$ with ℓ and ℓ' being the three lepton generations. The current status of these measurements are shown in table 1.1 and described in the following sections. Whilst not described in detail here, it is also worth mentioning that LFU has been verified using the charged-current interactions of electron neutrinos and muon neutrinos, for example in the CHARM experiment [35].

1.3.1   Low momentum transfer Lepton Flavour Universality measurements

The three $|g_\ell/g_{\ell'}|$ quantities have been measured at low momentum transfer as follows:

	$\Gamma_{\tau\to\mu}/\Gamma_{\tau\to e}$	$\Gamma_{\pi\to\mu}/\Gamma_{\pi\to e}$	$\Gamma_{K\to\mu}/\Gamma_{K\to e}$	$\Gamma_{K\to\pi\mu}/\Gamma_{K\to\pi e}$	$\Gamma_{W\to\mu}/\Gamma_{W\to e}$
$|g_\mu/g_e|$	1.0018 (14)	1.0021 (16)	0.9978 (20)	1.0010 (25)	0.996 (10)
	$\Gamma_{\tau\to e}/\Gamma_{\mu\to e}$	$\Gamma_{\tau\to\pi}/\Gamma_{\pi\to\mu}$	$\Gamma_{\tau\to K}/\Gamma_{K\to\mu}$		$\Gamma_{W\to\tau}/\Gamma_{W\to\mu}$
$|g_\tau/g_\mu|$	1.0011 (15)	0.9962 (27)	0.9858 (70)		1.034 (13)
	$\Gamma_{\tau\to\mu}/\Gamma_{\mu\to e}$				$\Gamma_{W\to\tau}/\Gamma_{W\to e}$
$|g_\tau/g_e|$	1.0030 (15)				1.031 (13)

Table 1.1: Previous experimental constraints on the universality of the couplings of leptons to the $W^\pm$ boson measured at various different momentum transfer [11]. Constraints from measurements of tau decays and light meson decays are measured to much higher precision than those using $W^\pm$ boson decays, and show good agreement with the Standard Model assumption of universal lepton couplings.

- $|g_\mu/g_e|$:
    - The ratio of partial widths of leptonic tau decays, $R_{(\tau\to\mu/e)}$,
    - The ratio of partial widths of leptonic pseudoscalar meson decays $R_{(M_{\mathrm{ps}}\to\mu/e)}$, where M_{ps} is a pseudoscalar meson $[\pi, K]$,
    - In the semileptonic decays of $K \to \pi\ell\nu$ ($K_{\ell 3}$), where the ratio is extracted by calculating the Cabbibo mixing angle of the first two generations, $|V_{us}|$, in both $K \to \pi e\nu$ and $K \to \pi\mu\nu$ then taking the ratio.

- $|g_\tau/g_\mu|$:
    - The $\mathrm{Br}(\tau \to \mu)/\tau_\tau$ relation described below,
    - The ratio of partial widths of tau to pseudoscalar meson, pseudoscalar meson to muon, $\Gamma(\tau \to M_{\mathrm{ps}})/\Gamma(M_{\mathrm{ps}} \to \mu)$.

- $|g_\tau/g_e|$:
    - The $\mathrm{Br}(\tau \to e)/\tau_\tau$ relation described below.

Many of these processes are sensitive to new physics that would be indicated by Lepton Flavour non-Universality (LFnU). In the following text these measurements are described in two categories: measurements in tau decays and measurements in light hadron decays.

Tau decays have one or two leptonic weak interaction vertices meaning they are useful probes for LFU measurements. The following will demonstrate how LFU measurements can be made in tau and muon decays, following the procedure described in Ref. [36]. To begin, first consider a muon decay $\mu^- \to e^- \bar\nu_e \nu_\mu$, which is an example of a process with an off-shell $W^\pm$ that contains two leptonic weak interaction vertices (see Figure 1.3). If it is not assumed that the strength of the weak interaction is the same at the two
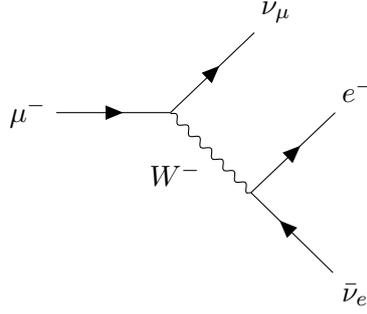
Figure 1.3: Lowest-order Feynman diagram of a muon decay. The muon decays to a muon neutrino and a virtual W^-, which subsequently decays to an electron and electron anti-neutrino.

| (a) Electronic | (b) Muonic | (c) Hadronic |
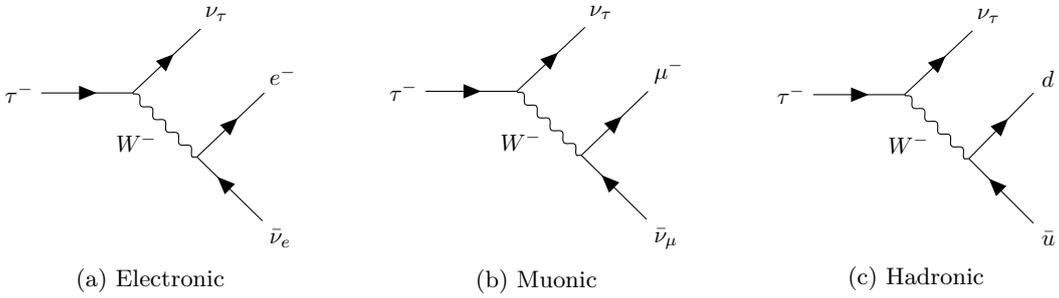
Figure 1.4: Lowest-order Feynman diagrams of a tau decay. The tau decays to a tau neutrino and virtual W^-, which subsequently decays to either light leptons or light quarks.

vertices, the muon decay rate can be written as:

$$\Gamma(\mu^- \to e^- \bar{\nu}_e \nu_\mu) \equiv \frac{1}{\tau_\mu} = \frac{G_F^{(e)} G_F^{(\mu)} m_\mu^5}{192\pi^3}, \tag{1.6}$$

where the mass of the electron is neglected. Taus can decay into electrons, muons or light mesons. The LO Feynman diagrams for these decays are shown in Figure 1.4.

Mirroring equation (1.6), the tau to electron decay rate can be written as:

$$\Gamma(\tau^- \to e^- \bar{\nu}_e \nu_\mu) = \frac{G_F^{(e)} G_F^{(\tau)} m_\tau^5}{192\pi^3}. \tag{1.7}$$

The tau lifetime is the inverse of the total decay rate, which is simply the sum of the partial decay rates:

$$\frac{1}{\tau_\tau} = \Gamma = \sum_i \Gamma_i.$$

The branching ratio of a process is simply the partial width divided by the total width. This means the $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ branching ratio is related to the τ lifetime as

$$\mathrm{Br}(\tau^- \to e^- \bar{\nu}_e \nu_\tau) = \frac{\Gamma(\tau^- \to e^- \bar{\nu}_e \nu_\tau)}{\Gamma} = \Gamma(\tau^- \to e^- \bar{\nu}_e \nu_\tau) \times \tau_\tau. \tag{1.8}$$

11

Then an expression for the tau lifetime can be found by substituting equation (1.7) into the right hand side of equation (1.8):

$$\tau_\tau = \frac{192\pi^3}{G_F^{(e)} G_F^{(\tau)} m_\tau^5} \mathrm{Br}(\tau^- \to e^- \bar{\nu}_e \nu_\tau). \tag{1.9}$$

Multiplying this by equation (1.6) gives

$$\frac{G_F^{(\tau)}}{G_F^{(\mu)}} = \frac{m_\mu^5 \tau_\mu}{m_\tau^5 \tau_\tau} \mathrm{Br}(\tau^- \to e^- \bar{\nu}_e \nu_\tau), \tag{1.10}$$

thus the relative weak interaction coupling strength of tau and muons, $|g_\tau/g_\mu|$, at low-momentum transfer can be determined from five experimental observables:

1. The muon mass

2. The muon lifetime

3. The tau mass

4. The tau lifetime

5. The tau to electron branching ratio.

$|g_\tau/g_e|$ can be calculated by using the tau to muon branching ratio in place of the tau to electron one. $|g_e/g_\mu|$ can be measured from the ratio of the tau to electron and tau to muon branching ratios.

The measurements of LFU in light pseudoscalar meson decays are of comparable precision to the tau decay measurements described above. The following section describes these measurements using the same outline as Ref. [11]. The measurement for light leptons takes the form

$$R_{M_{\mathrm{ps}} \to e/\mu} \equiv \frac{\Gamma(M_{\mathrm{ps}}^- \to e\bar{\nu}_e(\gamma))}{\Gamma(M_{\mathrm{ps}}^- \to \mu^- \bar{\nu}_\mu(\gamma))} = \left|\frac{g_e}{g_\mu}\right|^2 \frac{m_e^2}{m_\mu^2} \left(\frac{1 - m_e^2/m_{M_{\mathrm{ps}}}^2}{1 - m_\mu^2/m_{M_{\mathrm{ps}}}^2}\right)^2 (1 + \delta R_{M_{\mathrm{ps}} \to e/\mu}), \tag{1.11}$$

where $\delta R_{M_{\mathrm{ps}} \to e/\mu}$ are the known radiative corrections. The $V - A$ structure of the charged current weak interaction helicity suppresses the leptonic decay rate in the Standard Model. This is why pions favour decays to muons over decays to electrons. This feature also makes these measurements sensitive to new physics. The radiative corrections include a summation of leading QED logarithms $\alpha^n \log^n(m_\mu/m_e)$ and a systematic two-loop calculation of $\mathcal{O}(e^2 p^4)$ effects within Chiral perturbation theory. Measurements in $K \to \pi e \nu$ and $K \to \pi \mu \nu$ are not helicity suppressed. Known differences between the muon and electron arise from isospin and phase space. To obtain the value for $|g_\mu/g_e|$ in these decays, the values for the Cabibbo mixing angle V_{us} are determined in both decays and the ratio is taken.
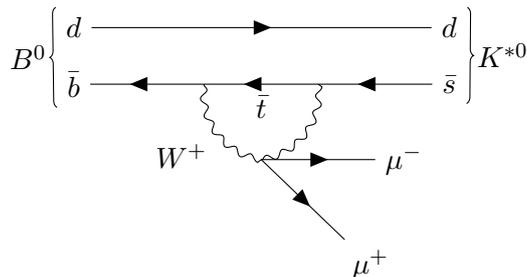
Figure 1.5: Feynman diagram of a Standard Model B^0, consisting of d and $\bar{b}$ quarks, decaying to K^{*0}, which consists of d and $\bar{S}$ quarks, as well as a dimuon pair.

The other set of light hadronic measurements are from the decays of $\tau^- \to \nu_\tau M_{\text{ps}}^-$ and $M_{\text{ps}}^- \to \mu^- \bar{\nu}_\mu$ through the ratios

$$R_{\tau/M_{\text{ps}}} \equiv \frac{\Gamma(\tau^- \to \nu_\tau M_{\text{ps}}^-)}{\Gamma(M_{\text{ps}}^- \to \mu^- \bar{\nu}_\mu)} = \left|\frac{g_\tau}{g_\mu}\right|^2 \frac{m_\tau^3}{2m_{M_{\text{ps}}} m_\mu^2} \left(\frac{1 - m_{M_{\text{ps}}}^2/m_\tau^2}{1 - m_\mu^2/m_{M_{\text{ps}}}^2}\right)^2 (1 + \delta R_{\tau/M_{\text{ps}}}), \quad (1.12)$$

where $\delta R_{\tau/M_{\text{ps}}}$ are the known radiative corrections, and the hadronic matrix elements cancel out.

1.3.2 Measurements of Lepton Flavour Universality in b-hadrons

In recent years, measurements from b-factory experiments have shown tension with the Standard Model, in so-called flavour anomalies that are possible observations of Lepton Flavour non-Universality. There are two key sets of measurements: $R_{K^{(*)}}$ and $R_{D^{(*)}}$.

$R_{K^{(*)}}$ measures the ratio of branching ratios

$$R_{K^{(*)}} = \frac{\text{Br}(B \to K^{(*)}\mu^+\mu^-)}{\text{Br}(B \to K^{(*)}e^+e^-)}. \tag{1.13}$$

As presented in Ref. [37], an angular analysis measuring the ratio of branching ratios of $B^0 \to K^{*0}\ell^+\ell^-$ has been performed in two regions of the dilepton invariant mass, q^2, which are $0.045 < q^2 < 1.1$ GeV^2 and $1.1 < q^2 < 6.0$ GeV^2, with the ratio measured to be

$$R_{K^{*0}} = \begin{cases} 0.66^{+0.11}_{-0.07}(\text{stat}) \pm 0.03(\text{syst}) & \text{for} \quad 0.045 < q^2 < 0.11 \text{ GeV}^2/c^4, \\ 0.69^{+0.11}_{-0.07}(\text{stat}) \pm 0.05(\text{syst}) & \text{for} \quad 1.1 < q^2 < 6.0 \text{ GeV}^2/c^4. \end{cases}$$

These results are compatible with the Standard Model expectations of unity (within an uncertainty of $\mathcal{O}(10^{-3})$ [38] at the level of 2.1-2.3 and 2.4-2.5 standard deviations in the two q^2 regions respectively. The $B^+ \to K^+\ell^+\ell^-$ measurement has been performed in the higher region only [39], and measures the ratio to be

$$R_K = 0.745^{+0.090}_{-0.074}(\text{stat}) \pm 0.036(\text{syst}),$$

Experiment	R_{D^*}	R_D	Citation
$B\textsc{a}B\textsc{ar}$	$0.332 \pm 0.024 \pm 0.018$	$0.440 \pm 0.058 \pm 0.042$	[40] [41]
BELLE	$0.293 \pm 0.038 \pm 0.015$	$0.375 \pm 0.064 \pm 0.026$	[42]
LHCb	$0.336 \pm 0.027 \pm 0.030$	-	[43]
BELLE	$0.270 \pm 0.035^{+0.028}_{-0.025}$	-	[44] [45]
LHCb	$0.280 \pm 0.018 \pm 0.029$	-	[46] [47]
BELLE	$0.283 \pm 0.018 \pm 0.014$	$0.307 \pm 0.037 \pm 0.016$	[48]
Average	$\mathbf{0.295 \pm 0.011 \pm 0.008}$	$\mathbf{0.340 \pm 0.027 \pm 0.013}$	

Table 1.2: Table of $R_{D^{(*)}}$ results, as described in Ref. [5]. The average Standard Model predictions are $R_D = 0.299 \pm 0.003$, and $R_{D^*} = 0.258 \pm 0.005$ [5]. Comparing with the bottom row of the table, agreement with the Standard Model is at the level of 1.4σ and 2.5σ respectively. When taken together, including correlations, the $R_{D^{(*)}}$ results agree with the Standard Model at the level of 3.08σ.

which is compatible with the Standard Model within 2.6 standard deviations.

These processes are electroweak penguin decays, meaning their leading-order amplitudes proceed at loop level (see Figure 1.5), therefore they are very sensitive to contributions from new physics. The tension with SM is seen in both invariant mass regions. There are large theoretical uncertainties on the branching ratio of $\mathcal{O}(30\%)$, however these largely cancel in the ratio of the observables. The SM predicts the ratio to be unity within an uncertainty $\mathcal{O}(10^{-3})$.

The second set of measurements measure the ratios of branching ratios in $B \to D^{(*)}\ell\nu$:

$$R_{D^{(*)}} = \frac{\mathrm{Br}(B \to D^{(*)}\tau^-\overline{\nu}_\tau)}{\mathrm{Br}(B \to D^{(*)}\ell^-\overline{\nu}_\ell)}, \ \ell = e, \mu. \tag{1.14}$$

These measurements proceed at tree-level, therefore the new physics effects must be apparent at the electroweak scale $\mathcal{O}(1 \text{ TeV})$. The various results along with the average result are shown in Table 1.2 and Figure 1.6. When considering the combined fit of the averages of R_D and R_{D^*}, the SM predictions are exceeded by 3.08 standard deviations [5].

The $R_{K^{(*)}}$ and $R_{D^{(*)}}$ measurements represent the first measurements of LFU using the third generation of quarks. Their tension with the SM indicates potential Lepton Flavour non-Universality, hence they are described as flavour anomalies. These flavour anomalies point unambiguously to low energy scales for new physics (NP). They appear in a consistent and correlated way in various observables, most of which can be computed to high precision within the SM. This motivates further measurements of LFU in the third generation of quarks.

In addition to these flavour anomalies is the recent test of LFU in $\Upsilon(3S)$ meson decays
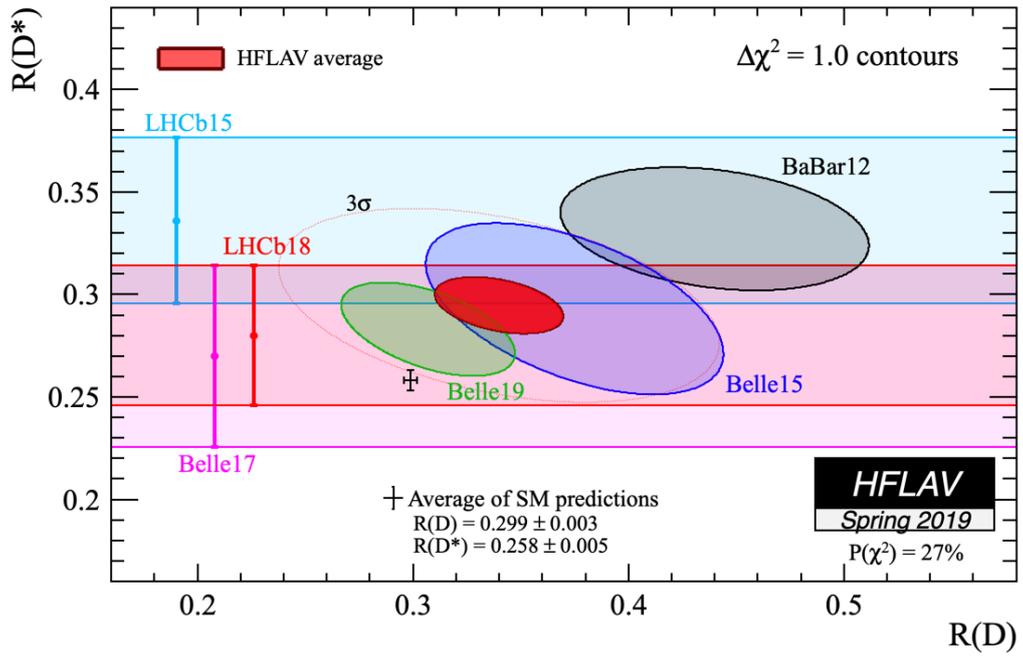
Figure 1.6: HFLAV $R(D^{(*)})$ averages [5]. The average of the Standard Model predictions are $R_D = 0.299 \pm 0.003$, and $R_{D^*} = 0.258 \pm 0.005$ [5]. When taken together, including correlations, the $R_{D^{(*)}}$ results agree with the Standard Model at the level of 3.08σ.

at *BaBar*, measuring

$$R_{\Upsilon(3S)} = \frac{\mathrm{Br}(\Upsilon(3S) \to \tau^+\tau^-)}{\mathrm{Br}(\Upsilon(3S) \to \mu^+\mu^-)}.$$

The precision measurement is based on a 28 fb^{-1} data sample collected at 10.355 GeV collected with the *BaBar* detector at the SLAC PEP-II electron-positron collider. The ratio is measured to be $R_{\Upsilon(3S)} = 0.966 \pm 0.008(\mathrm{stat}) \pm 0.014(\mathrm{syst})$ [49]. This in agreement with the Standard Model prediction of 0.9948 at the level of 1.8 standard deviations.

1.3.3 High momentum transfer Lepton Flavour Universality measurements

So far, the most precise tests of Lepton Flavour Universality between taus and muons in high-momentum transfer processes come from the LEP averages [50]. In these averages there is an excess in the branching fraction of $W^\pm$ to tau compared to the light leptons. ATLAS has also previously tested Lepton Flavour Universality between the light leptons to high precision [51].

The PDG averages [32] of the ratio of branching ratios are as follows:

$$\mathrm{Br}(W \to \mu\bar{\nu}_\mu)/\mathrm{Br}(W \to e\bar{\nu}_e) = 0.996 \pm 0.008 \tag{1.15}$$

$$\mathrm{Br}(W \to \tau\bar{\nu}_\tau)/\mathrm{Br}(W \to e\bar{\nu}_e) = 1.043 \pm 0.024 \tag{1.16}$$

$$\mathrm{Br}(W \to \tau\bar{\nu}_\tau)/\mathrm{Br}(W \to \mu\bar{\nu}_\mu) = 1.070 \pm 0.026. \tag{1.17}$$

Therefore, comparing electrons and muons, agreement is seen with the SM prediction of unity within the experimental uncertainty. However, comparing taus and electrons, agreement with the SM prediction of unity is at the level of 1.8 standard deviations, and comparing taus and muons, agreement with the SM prediction of unity is at the level of 2.7 standard deviations.

LFU is, with few exceptions, well constrained at both at low momentum transfer and between the first two generations of leptons. At high momentum transfer and between taus and light leptons the picture is less clear, thus further measurements are well motivated.

One source of on-shell $W^\pm$ bosons are those produced in top decays. Flavour changing neutral currents (FCNC) are heavily suppressed in the SM so top quarks decay almost exclusively via charged current, resulting in a bottom quark and $W^\pm$ boson. By measuring the leptonic branching fractions of the $W^\pm$ boson and taking their ratio, a test of LFU can be made.

1.4 The strong interaction

This section provides a very brief and qualitative overview of the relevant QCD for understanding *pp* collisions in ATLAS. Some more detail is provided in the description

of *pp* event simulations in Section 2.11.

The strong interaction is described by Quantum Chromodynamics (QCD) [52–54]. QCD is a non-abelian gauge theory, with symmetry group $SU(3)$. It provides a description of colour-charged particles. Colour is the QCD analogue of electric charge in QED. There are three colours: red (r), blue (b), and green (g), with corresponding anti-colours. Gluons (g) are the vector gauge bosons that communicate the strong force. Gluons can be thought of as carrying both colour and anti-colour, meaning they come in eight colour configurations. Carrying colour means that gluons can self-interact. The strong coupling constant, α_S, is not actually constant, but rather it is highly dependent on the energy scale. The behaviour of α_S can be summarised as follows:

- It increases for large distances between colour-charged objects and is smaller for short distances.

- It becomes smaller for increasing energy scales, whilst it increases for smaller energy scales.

It is these features that give rise to confinement and asymptotic freedom, two key properties of QCD. These are described below.

The strong coupling constant can be expressed as:

$$\alpha_s \approx \frac{1}{(11 - 2n_f/3)\log(Q^2/\Lambda_{\mathrm{QCD}}^2)}, \qquad (1.18)$$

where Q is the energy scale, n_f is the number of quark flavours with masses above the energy scale and Λ_{QCD} is the QCD cut-off energy. Λ_{QCD} is determined experimentally. At energies higher than the cut-off (~ 200 MeV), the denominator decreases the coupling constant. This causes asymptotic freedom, and is the domain in which a perturbative treatment of QCD can be used. When Q^2 is approaching the cut-off, the strong coupling constant grows quickly, and the perturbative approach cannot be used.

At large distances, the coupling constant becomes so large that additional colour-charged particles are produced through vacuum fluctuations, such that the system reaches a lower-energy state. This is known as the hypothesis of confinement, the phenomenon which results in the non-observation of free gluons and quarks. Instead it is the bound states of quarks – mesons and baryons – that are the experimental observables. This leads to, for example, parton showers that form jets. However, at very short distances, the coupling constant is very small, and colour-charged particles can be considered asymptotically free. The colliding partons in high-energy *pp* collisions fall into this latter treatment.
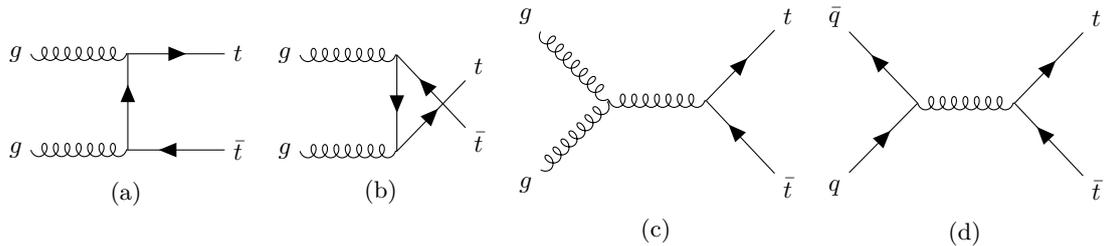
Figure 1.7: LO Feynman diagrams for top quark pair-production through (a), (b), (c) gluon-gluon fusion and (d) quark-antiquark annihilation.

1.5 Top quark production and decay

Top quarks have large mass and are highly unstable. They are produced in high energy collisions. They were first discovered at the Tevatron in the CDF and DØ experiments [19,20]. The high mass and short lifetime of top quarks make them interesting objects of study for new physics searches, as well as validation of the Standard Model. Since they decay before hadronising, they allow analysis of "bare-quark" behaviour. This section covers some of the production and decay modes of top quarks.

1.5.1 Production

The LHC operates with a centre-of-mass energy that is high enough to produce top quarks, which have a mass of $m_t = 172.9 \pm 0.4$ GeV [32]. There are two categories of top-quark production: pair production and single-top production, both of which have several distinct channels.

Top-quark pair production tends to occur via strong interactions. A high energy gluon is created before decaying to a top-antitop pair. Fusion of two gluons (gg fusion) is also possible. They can also occur via intermediate photon or Z^0-boson [55]. The production cross-section for $t\bar{t}$ is dependent on the top mass and the centre-of-mass-squared energy s. Leading order Feynman diagrams for $t\bar{t}$ pair-production are shown in Figure 1.7.

The total inclusive $t\bar{t}$ production cross-section in proton-proton collisions, $\sigma_{t\bar{t}}$, has been measured at $\sqrt{s} = 13$ TeV at ATLAS [56] and CMS [57] using dileptonic final states, and extrapolating from the fiducial cross section to the full phase space. Predictions are available at next-to-next-to-leading-order (NNLO) accuracy in the strong coupling constant α_S, including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [58–63]. At $\sqrt{s} = 13$ TeV and assuming a fixed top mass of 172.5 GeV, the NNLO+NNLL prediction is $832 \pm 35^{+20}_{-29}$ pb as calculated using the TOP++2.0 programme [64]. The measured value by ATLAS, using 31.6 fb^{-1} of data [56], is

$$\sigma_{t\bar{t}} = 826 \pm 3.6(\text{stat}) \pm 11.5(\text{syst}) \pm 15.7(\text{lumi}) \pm 1.9(\text{beam}) \text{ pb},$$
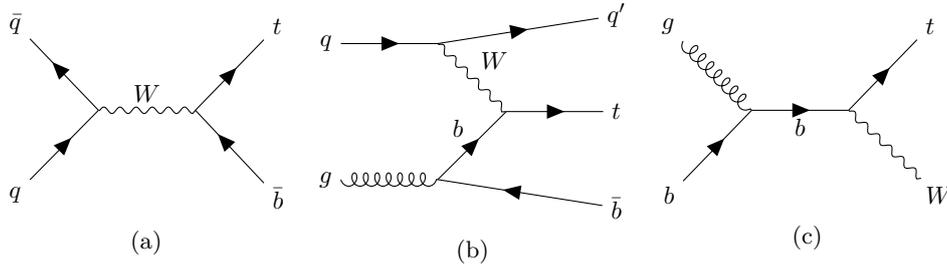
Figure 1.8: LO Feynman diagrams for single-top production, showing (a) s-channel, (b) t-channel and (c) Wt-channel.

s-channel		
σ_t (pb)	$\sigma_{\bar{t}}$ (pb)	total (pb)
$7.07 \pm 0.13^{+0.24}_{-0.22}$	$4.10 \pm 0.05^{+0.14}_{-0.16}$	$11.17 \pm 0.18 \pm 0.38$

t-channel		
σ_t (pb)	$\sigma_{\bar{t}}$ (pb)	total (pb)
$136^{+3}_{-1} \pm 1.1$	$82^{+2}_{-1} \pm 2$	$218^{+5}_{-2} \pm 5$

Wt-channel	
σ_{W^-t} (pb)	$\sigma_{W^-t+W^+\bar{t}}$ (pb)
$35.20 \pm 0.9^{+1.6}_{-1.7}$	$70.40 \pm 1.8^{+3.2}_{-3.4}$

Table 1.3: Comparison of predicted single top production cross sections in pp collisions at $\sqrt{s} = 13$ TeV. The theory predictions are at approximate next-to-next-to-leading-order (aNNLO) and assume a fixed top quark mass $m_t = 173.3$ GeV, as described in Ref. [12].

and by CMS, using 35.9 fb^{-1} [57], is

$$\sigma_{t\bar{t}} = 803 \pm 2(\text{stat}) \pm 25(\text{syst}) \pm 20(\text{lumi}) \text{ pb}.$$

Thus there is good agreement between predicted and measured values.

Single top production occurs via the weak interaction. The three main channels, shown in Figure 1.8, are s, t and Wt. In the s-channel (Figure 1.8a), an intermediate $W^\pm$ boson decays into a top and an antibottom. In the t-channel (Figure 1.8b), a bottom quark produces a top quark via a charged current interaction. In the Wt-channel, a $W^\pm$ boson is produced in association with a top quark. The frequency of these production processes is directly proportional to the $|V_{tb}|^2$ component of the CKM matrix.

The theory predictions for single top production cross sections in pp collisions at $\sqrt{s} = 13$ TeV assuming a fixed top mass $m_t = 173.3$ GeV are shown in Table 1.3. The dominant

channel for single top production is the t-channel, accounting for about 73% of the total single top cross section, followed by the Wt-channel, accounting for about 23% of the total, with s-channel being the rarest process, accounting for less than 4% of the total. The predicted pair production cross section described above is nearly three times the total combined single top production cross section, and nearly four times the dominant t-channel cross section, thus pair production is by far the dominant source of top quark production. The inclusive t-channel t and $\bar{t}$ production cross sections have been measured at $\sqrt{s} = 13$ TeV at ATLAS using 3.2 fb^{-1} of data [65] to be

$$\sigma(tq) = 156 \pm 5(\text{stat.}) \pm 27(\text{syst.}) \pm 3(\text{lumi.}) \text{ pb}$$

$$\sigma(\bar{t}q) = 91 \pm 4(\text{stat.}) \pm 18(\text{syst.}) \pm 2(\text{lumi.}) \text{ pb}$$

and at CMS using 35.9 fb^{-1} of data [66] to be

$$\sigma(tq) = 130 \pm 1(\text{stat}) \pm 19(\text{syst}) \text{ pb}$$

$$\sigma(\bar{t}q) = 77 \pm 1(\text{stat}) \pm 12(\text{syst}) \text{ pb}$$

showing agreement with Standard Model predictions. The inclusive Wt-channel production cross sections have been measured at ATLAS using 3.2 fb^{-1} of data [67] to be

$$\sigma_{Wt} = 94 \pm 10(\text{stat.})^{+28}_{-22}(\text{syst.}) \pm 2(\text{lumi.}) \text{ pb},$$

and at CMS using 35.9 fb^{-1} of data [68] to be

$$\sigma_{Wt} = 63.1 \pm 1.8(\text{stat}) \pm 6.4(\text{syst}) \pm 2.1(\text{lumi}) \text{ pb}.$$

Searches for s-channel single top production have been performed at $\sqrt{s} = 8$ TeV at ATLAS with an observed signal significance of 3.2 standard deviations. Previously, s-channel single top production has been observed at the Tevatron [69].

1.5.2 Decay

The high mass of the top quark means it is extremely short lived. It has a lifetime of just 5×10^{-25} s. The top quark has only been observed decaying via charged current weak interaction, producing a $W^{\pm}$ boson and a b quark. The ratio $\Gamma(Wb)/\Gamma(Wq(q = b, s, d))$ is equal to $|V_{tb}|^2$, and has the average measured value 0.957 ± 0.034 [32].

As shown in Table 1.4 the final state of a top decay is dependent on the $W^{\pm}$ decay. Thus the top decay final state can be either semileptonic or fully hadronic, with hadronic final states favoured. Since pair-production is the dominant top quark production mechanism, and the top and antitop are independent of one another, the final state for $t\bar{t}$ is a combination of two of these decay modes. In cases where both $W^{\pm}$ bosons decay leptonically, the final state charged leptons will have opposite electric charge.

t decay mode	Fraction (Γ_i/Γ)
$t \to e\nu_e b$	$(11.10 \pm 0.30)\%$
$t \to \mu\nu_\mu b$	$(11.40 \pm 0.20)\%$
$t \to \tau\nu_\tau b$	$(11.1 \pm 0.9)\%$
$t \to q\bar{q}b$	$(66.5 \pm 1.4)\%$

Table 1.4: t decay final state branching fractions, as described in Ref. [13]. Decays to hadrons dominate, occurring approximately two-thirds of the time. Decays to leptons occur approximately one-third of the time, with roughly equal fractions of electrons, muons and taus.

More exotic decays are possible in the Standard Model, such as flavour-changing-neutral-currents involving a Z^0 boson, photon or gluon. However these processes proceed at loop-level and are therefore highly suppressed. Predictions are of the order 10^{-12} (for $t \to gc$) or smaller [70].

1.6 Summary of chapter 1

This section has covered the theory relevant to LFU and an overview of the current experimental landscape. The experimental results from *b*-factory experiments indicate hints of new physics through tests of LFU. The combined direct test of LFU at LEP does not conclusively agree nor disagree with the Standard Model prediction. Thus further research is well motivated. If the central value of the LEP result is replicable, a precision of at least 1-2% is required to give an unambiguous discovery of new physics. Such precision is now possible using the data collected by ATLAS. In Run-2, ATLAS collected 139 fb^{-1} of *pp* collision data at $\sqrt{s} = 13$ TeV. Within this dataset are over 100 million top-anti-top-pairs which, as shown in this chapter, decay almost exclusively via the weak charged current interaction. By using these $W^{\pm}$ bosons to measure the ratio of branching fractions Br($W \to \tau\nu$)/Br($W \to \mu\nu$), a precise test of LFU can be made. The following chapters will cover the ATLAS detector in detail, and describe an analysis precisely measuring Br($W \to \tau\nu$)/Br($W \to \mu\nu$) in top quark decays.

2

ATLAS

ATLAS [7] is one of the two general purpose detectors at the Large Hadron Collider (LHC) at CERN, the Organisation for Nuclear Research. The purpose of the LHC is to accelerate protons to high energies, before colliding them in detectors such as ATLAS, where the resulting particles can be identified and their kinematic properties measured. This in turn allows analysis of Standard Model measurements and tests of new physics. This section briefly explains the key details of the LHC before describing the ATLAS detector and its various subdetectors.

2.1   Large Hadron Collider (LHC)

The LHC [71–73] is a circular proton collider with a circumference of 26.7 km. At the time of writing, it is both the highest energy particle accelerator and the largest machine in the world. It is housed in a tunnel that is up to 175 m deep, and runs under the Franco-Swiss border near Geneva, Switzerland. The tunnel was originally constructed to house the Large Electron Positron Collider (LEP), which had a centre-of-mass energy of up to 209 GeV. By colliding protons instead of electrons and positrons, the LHC has a significantly larger design peak centre-of-mass energy of $\sqrt{s} = 14$ TeV.

The LHC is built to collide beams of protons, with additional special runs of heavy ions. The focus throughout this section will be on the collision of protons, with these being the focus of this thesis. The protons enter the LHC through an injection chain comprising several smaller accelerators that successively increase their energy. This is shown in Figure 2.1. The protons are produced by ionising hydrogen gas. They are initially accelerated to 50 MeV using the linear accelerator (Linac2 up to Run-2, Linac4 from Run-3 [74]) from which they enter the Proton Synchrotron Booster (PSB). They are accelerated to 1.4 GeV before entering the Proton Synchrotron (PS), which accelerates them to 26 GeV. The penultimate link in the chain is the Super Proton Synchrotron which accelerates them to 450 GeV before they enter the final ring, the LHC, which is

The CERN accelerator complex
Complexe des accélérateurs du CERN
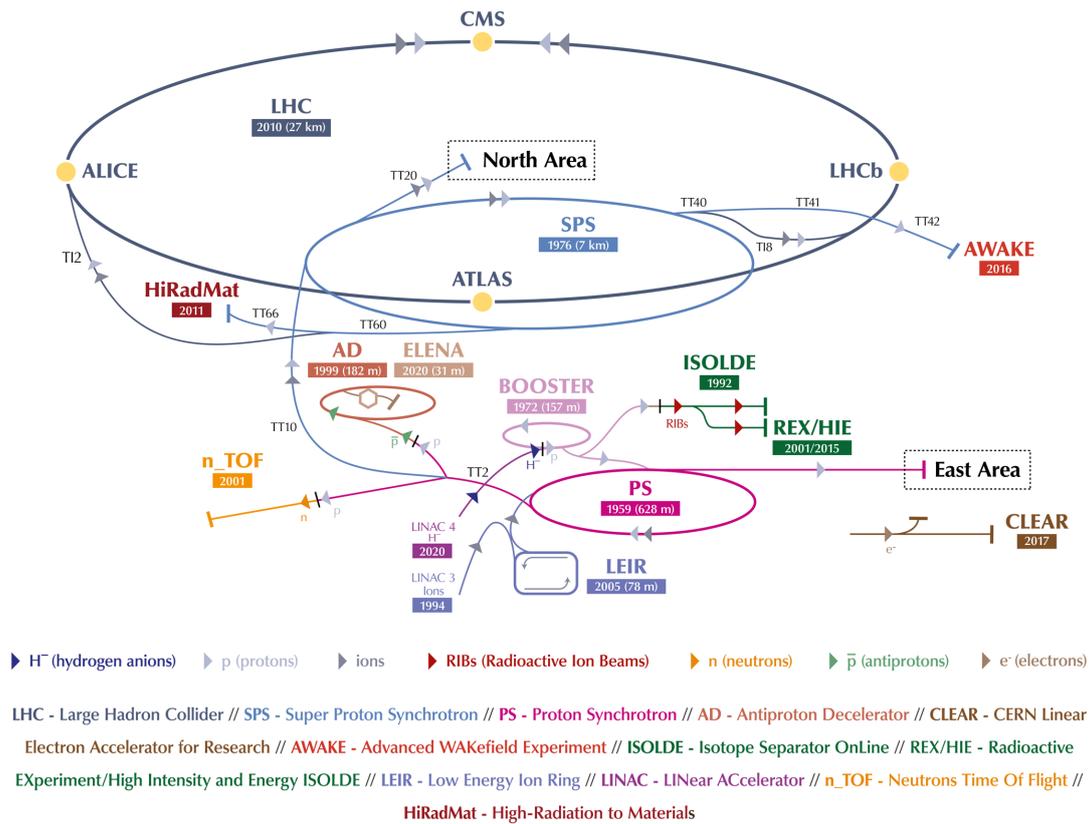
Figure 2.1: The LHC accelerator complex [6] including the Linac-4 replacement of Linac-2. Protons are accelerated through the complex starting in the Linac and moving through the chain before entering the LHC.

designed to accelerate two beams of protons up to 7 TeV in opposite directions. So far the LHC has operated with a maximum beam energy of 6.5 TeV, or $\sqrt{s} = 13$ TeV.

Since both proton beams have the same charge, they must be accelerated in separate but parallel pipes[1]. The pressure in the pipes is maintained at $\sim 10^{-13}$ atm, hence they are considered ultra-high vacuums. The collider is arranged into eight arc and eight straight sections. These include four interaction points, as well as collimation, radio frequency (RF) cavity and beam dump systems, amongst others. The protons are accelerated by passing through eight RF cavities per beam. The particle trajectories are maintained by a large number of superconducting magnets. It is not possible to have two separate magnet systems in the tunnel, which has an internal diameter of 3.7 m. There are several different types of magnets, including the 8 T dipole magnets that are used to bend the beams on the circuit. Quadrupole, sextupole, octupole and decapole magnets are used to control the beam optics. The magnets are cooled to 1.9 K using liquid helium. The beamlines cross in four intersection points where the main experiments are located.

The particle flux in the beam is not continuous. The protons travel in bunches. Each bunch contains up to 10^{11} protons. The bunch spacing is 25 ns. These bunches are collided at a rate of 40 MHz.

There are four collision points along the LHC, around which the experiments are located:

ATLAS [7] is a general purpose experiment. It has sensitivity to final state objects and topologies of Standard Model electroweak interactions.

CMS [75] is a general purpose detector, with a single magnetic field provided by a superconducting solenoid. It has a similar acceptance and resolution as well as similar physics goals to ATLAS. Physics measurements are expected to be compatible between ATLAS and CMS to ensure reproducibility.

LHCb [76] is a single arm forward spectrometer. Design considerations include a focus on particle identification and vertex resolution. With this it has good sensitivity to flavour physics including b-meson decays, mixing and the formation of (possible exotic) bound states. Key goals include constraining the CKM matrix parameters and unitary triangles in which new physics can be observed.

ALICE [77] is a general purpose detector with a focus on QCD interactions at high energy densities. A key physics focus is the formation of quark-gluon plasma. It has a focus on heavy ion collisions which produce around 8,000 charged particles per event.

---

[1]This stands in contrast to particle-anti-particle colliders such as LEP and Tevatron which can accelerate both beams in the same pipe.

2.2 ATLAS overview

The ATLAS detector is a multipurpose detector at the LHC in CERN. It is located in an underground cavern at Point-1 (P1) along the LHC beamline. The ATLAS control centre is located above ground in the same location. It is cylindrical in shape, with a length of 46 m and a diameter of 25 m. It weighs some 7000 tons. It is comprised of multiple systems:

- The Inner Detector (ID): tracks charged particles produced by pp collisions. The ID is comprised of the following subdetectors:

    - Pixel detector (PXT),

    - Transition radiation tracker (TRT),

    - Semiconductor tracker (SCT).

- Calorimetry: measures the energy of electrons, photons and hadrons produced in pp collisions. The calorimetry system is comprised of the following subdetectors:

    - Electromagnetic calorimeter (ECal),

    - Hadronic calorimeter (HCal),

    - Forward calorimeter (FCal).

- Muon spectrometers (MS): measures muon hits and trajectories.

- Trigger and Data Acquisition System (TDAQ): reduces data collection to a manageable rate by deciding which events to save.

- Magnet systems: bends the trajectories of charged particles to measure charge and momentum in the ID and MS.

These subdetectors are arranged symmetrically around the interaction point, meaning that the ATLAS detector covers almost the entire solid angle.

A schematic of the ATLAS detector is found in Figure 2.2. The following section describes each subdetector in detail.
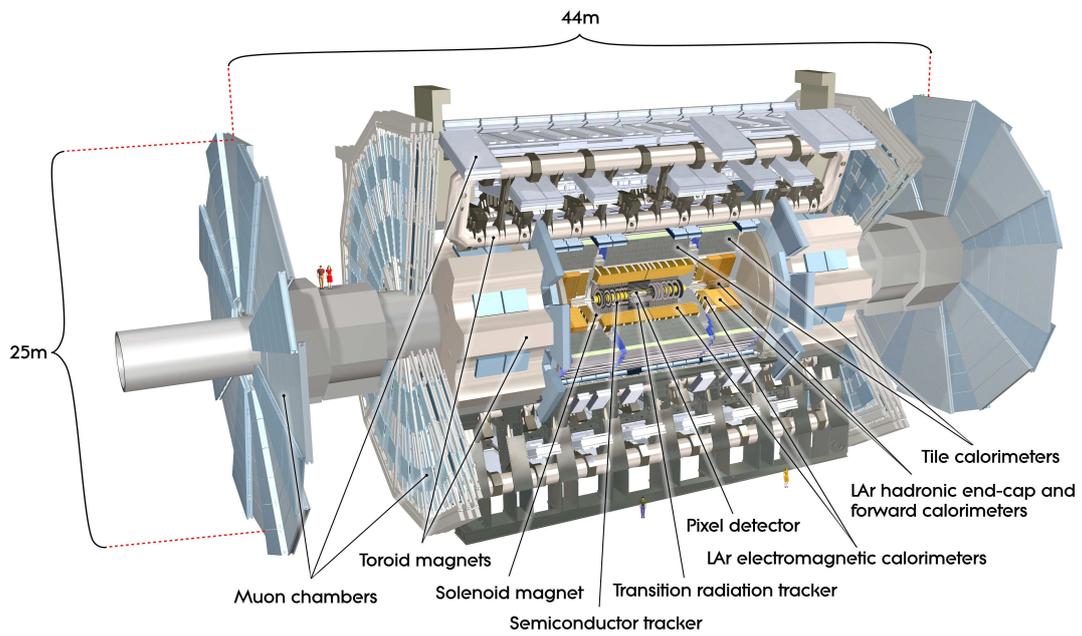
Figure 2.2: Schematic of the ATLAS detector [7]. The various subdetectors are arranged symmetrically around the collision point in the centre, beginning with the Inner Detector (pixel detector, silicon tracker and transition radiation tracker), the Calorimetry system (ECal, HCal and FCal) and the muon spectrometers. Toroid magnets are used to bend the trajectories of charged particles to allow their charges and momenta to be determined.

2.3 Coordinate system

With ATLAS being comprised of several systems, which are themselves comprised of subdetectors, a common coordinate system must be used throughout the detector. The ATLAS experiment uses a right-handed coordinate system. The origin is at the nominal interaction point (IP) in the centre of the detector such that $(x, y, z) = (0, 0, 0)$. The z-axis points is parallel to the beamline. The x-axis points from the interaction point (IP) to the centre of the LHC ring, and the y-axis points upwards to the sky. The x-y plane describes a plane transverse to the beam direction. ATLAS can also be described using cylindrical coordinates (θ, ϕ, z). The azimuthal angle, ϕ, covers the range $[-\pi, \pi]$ and the polar angle, θ, covers the range $[0, \pi]$. Pseudorapidity is defined as

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right),$$

which is often used as a measure for the polar angle in place of θ. This is due to the fact that separation in η between two massless particles is an invariant quantity under a boost in the z-direction. A pseudorapidity $\eta = 0$ corresponds to an angle of 90° and $\eta = \infty$ corresponds to the direction of the beamline. Angular distances in the transverse plane are defined using

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}.$$

Other common quantities include:

- p_T, the magnitude of the momentum in the transverse plane.

- d_0, the transverse impact parameter. This is usually defined as the closest approach of a track with respect to the beamspot in the transverse plane, though it can also be defined with respect to the primary pp interaction.

- z_0, the longitudinal impact parameter. This is defined as the closest approach of a track with respect to the primary pp interaction in the longitudinal plane.

- $E_\mathrm{T}^\mathrm{miss}$, the missing transverse energy.

2.4 Inner Detector

The Inner Detector (ID) is located closest to the interaction point. A schematic is shown in Figure 2.3. It has the highest granularity in ATLAS, necessitated by the frequency of collisions and density of particles around the interaction point. It measures tracks of charged particles in the region $|\eta| < 2.5$. Measurements from the ID allow for the reconstruction of pp interaction vertices (also called primary vertices). A central solenoid magnet maintains a 2 T magnetic field inside the ID, which curves the trajectories of the tracks. This allows measurements of charge using the direction of curvature and of

28
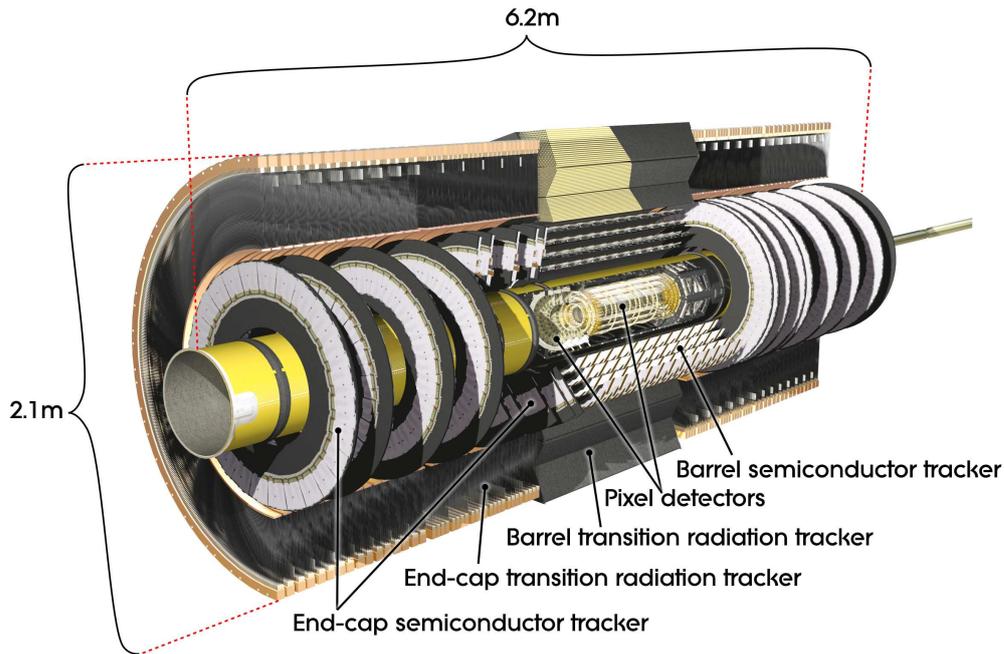
Figure 2.3: The ATLAS Inner Detector [7]. The pixel detector is found the in region closest to the beamline, and is surrounded by the silicon tracker and the transition radiation tracker.

momentum using the radius of the curvature. The ID has a lower resolution for higher p_T tracks. The ID is comprised of three trackers. There are two precision trackers: the innermost pixel detector (PIX) and the semiconductor tracker (SCT), as well as the outermost transition radiation tracker (TRT).

2.4.1 Pixel Detector

The pixel detector is the innermost tracker of the ID. It has the highest granularity. It sits at a distance of 33.25 to 122.5 mm from the beamline. It is comprised of a cylindrical barrel in the most central region, with two endcaps. The barrel consists of four layers and the endcaps consist of three disks. The pixel detector covers a solid angle in $|\eta| < 2.5$. It is composed of silicon detectors of size $50 \times 400~\mu\text{m}^2$. It has some 80 million readout channels. Silicon is a semiconductor, meaning its properties can be modified by doping. The pixels in the pixel detector have a p-doped region and an n-doped region. This generates a depletion zone between the two regions, producing similar behaviour to a diode.

The innermost layer of the pixel detector is the Insertable B-Layer (IBL) [78]. This is a relatively new component installed for Run-2. The IBL aims to ensure tracking robustness in the event of module failures in the adjacent layer (the B-Layer), to mitigate effects of increasing luminosity beyond the designed peak, to improve tracking precision

(especially for vertexing and *b*-tagging), as well as further radiation damage considerations. It allows for precise determination of primary *pp* interaction vertices and the impact parameters z_0 and d_0. The pixel size of the IBL is 50×250 μm^2. This small size is required to guarantee a low hit occupancy, allowing the reconstruction of the trajectories of the many particles close to the interaction point, as well as the displaced vertices from, for example, *b*-mesons.

2.4.2   Semiconductor Tracker

The Semiconductor Tracker (SCT) lies around the pixel detector at a distance of 299 to 514 mm from the beamline in the barrel region. It is composed of silicon microstrips, as opposed to the pixels of the PIX. The SCT is composed of four layers in the barrel region, with two endcaps each composed of nine disks. 768 strips join together with a pitch of 80 μm to form modules that are 6.36×6.4 cm^2. These layers are used to build 12.8 cm strips by joining two back-to-back. They are joined at an angle of 40 mrad such that they are tilted. This allows for measurements in R–ϕ and z directions. The SCT modules have a surface area of about 61 m^2 and some 6.36 million readout channels. Each track crosses eight strip layers (four space points) in the SCT.

2.4.3   Transition radiation tracker

The Transition Radiation Tracker (TRT) is the outermost part of the ID. It is located around the SCT at a distance of 554 to 1082 mm in the barrel region. It is composed of gas-filled polypropylene straw drift tubes each of which has a 4 mm diameter. The TRT is therefore a gaseous detector. In the barrel region, the TRT has 73 straw planes that are arranged parallel to the beamline, forming a wheel. It has some 351,000 readout channels, and covers the region $|\eta| \leqslant 2$. It averages some 36 hits per track, which is large compared to the precision trackers.

The gaseous mixture in the tubes is primarily xenon-based, however, channels that are affected by irreparable gas leaks now use a much cheaper argon-based mixture [79]. The anodes inside the tubes are tungsten wires. Tungsten allows for measurement of electrons produced through ionisation. When a charged particle moves in the material between the tubes, photons are produced. When a charged particle crosses the TRT straws, the active gas mixture is ionised and electrons are liberated. The straw wall is kept at high negative voltage such that these primary electrons are accelerated towards the central anode, liberating more electrons as they go. This produces a detectable signal.
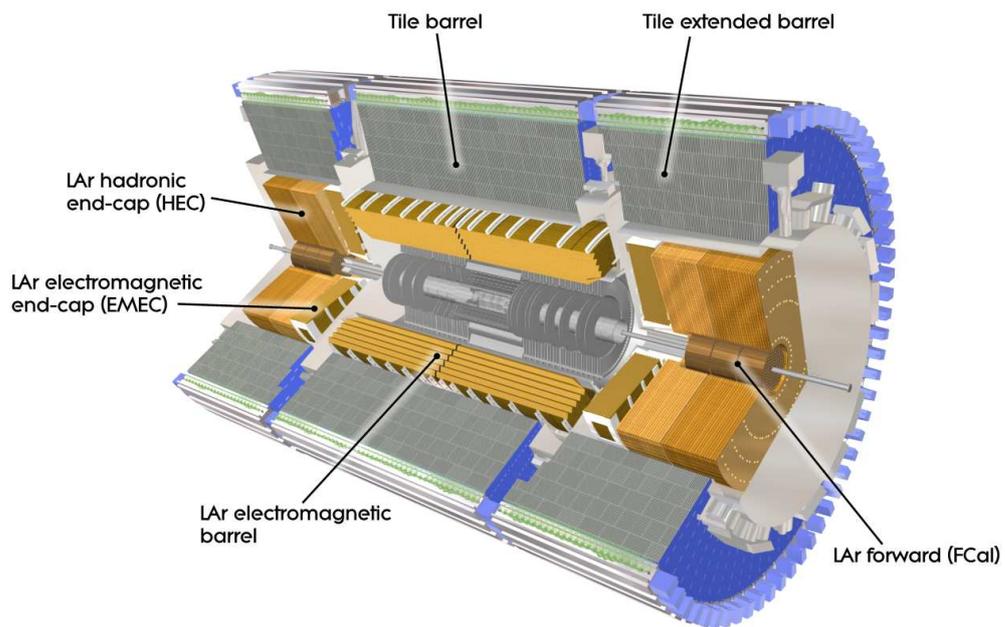
Figure 2.4: The ATLAS Calorimetry System [7]. The ECal is located around the Inner Detector. The HCal (tile) surrounds the ECal. The ECal and HCal endcaps and the FCal are located in the forward region.

2.5 Calorimetry

The calorimetry system is chiefly comprised of the electromagnetic calorimeter (ECal) and the hadronic calorimeter (HCal), as well as the forward calorimeters (FCal). A schematic is shown in Figure 2.4. The ECal is located around the ID, and the HCal is located around the ECal. The FCal is located in the forward region and allows the calorimetry system pseudorapidity coverage to extend up to $|\eta| = 4.9$. The system uses a variety of techniques suited to the widely varying physics and radiation requirements over the large η range. The aim of calorimetry is to measure the energy of particles by inducing particle showers in a high density material. Electromagnetic showers are caused by bremsstrahlung and pair production. There are several processes for hadronic showers. Both the ECal and HCal are sampling calorimeters, meaning they have alternating layers of active detector material and absorber material. The absorber layers absorb energy from the original particle, whose energy is then measured in the active medium. Incoming particles should deposit all energy inside the calorimeters. The ECal is designed to identify electrons and photons. Hadrons should pass through the ECal and into the HCal. The shower properties are used to identify the incoming particles. The calorimetry system covers a pseudorapidity range $|\eta| < 4.9$ and extends up to 2.5 m in radius. The individual calorimeters overlap to provide full coverage.

For electromagnetic calorimetry, radiation length, X_0 is defined as the length after which the energy of a relativistic electron would be reduced by a factor $1/e$ due to bremsstrahlung. For hadronic calorimetry, the radiation length λ is defined as the mean free path of the particle before undergoing inelastic scattering.

2.5.1 Electromagnetic calorimeter

The Electromagnetic calorimeter (ECal) is a sampling calorimeter. Lead is used as the absorber medium to induce showers, and liquid argon (LAr) is used as the detector medium to detect them. Showers are usually produced by bremsstrahlung and pair production. The ECal provides measurement in $|\eta| < 3.2$, with the barrel covering $|\eta| < 1.475$ and the endcaps covering $1.375 < |\eta| < 3.2$. The lead absorber plates and kapton[2] electrodes have an accordion shape to ensure full coverage in ϕ. There are three different sections of the ECal, with differing granularity depending on radial depth. Good energy resolution is ensured over the range of coverage by varying the amount of absorber material as a function of $|\eta|$. The layers have the following properties:

Layer 1: 4.4 X_0, 0.003×0.0245 in $\Delta\eta \times \Delta\phi$

Layer 2: 16 X_0, 0.025×0.0245 in $\Delta\eta \times \Delta\phi$

Layer 3: 2 X_0, 0.05×0.0245 in $\Delta\eta \times \Delta\phi$.

---

[2]A polyimide film stable across a very wide range of temperatures.

This means the ECal totals around 22 X_0 in the barrel and 24 X_0 in the endcaps. The ECal is segmented in three sections in depth over the region devoted to precision physics $|\eta| < 2.5$. This is where ID measurements are available. Here, an accurate position measurement is obtained by finely segmenting the first layer in η.

The strip layer is the first, fine layer of the ECal. It allows for resolution of showers of photons from neutral hadron decays, which are important for suppressing hadronic signatures in photon reconstruction. Most energy is deposited in the second layer of the ECal. The third layer is to correct for leakage into the HCal. There is a presampler installed in front of the ECal in the region $|\eta| < 1.8$. It consists of a thin layer of LAr. It corrects for energy losses of photons and electrons upstream of the ECal from interactions with ID material.

2.5.2   Hadronic calorimeter

Hadronic showers develop differently compared to electronic showers. The hadronic calorimeter (HCal) measures energy deposits of hadrons. Hadrons usually deposit only a small amount of energy in the ECal. The main purpose of the HCal is to perform jet and $E_\mathrm{T}^\mathrm{miss}$ measurements. The central region of the HCal has a coarser granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ in the central region. This is enough for jet reconstruction and $E_\mathrm{T}^\mathrm{miss}$ measurements. The barrel of the HCal is divided into a central barrel for $|\eta| < 1.0$ and two extended barrels for $0.8 < |\eta| < 1.7$, which provides space for electronics and service pipes. The barrel of the HCal is composed of alternating steel absorber layers and scintillating tile layers in the region $|\eta| < 1.7$. The scintillating tiles are the detector medium and emit a flash of light when struck by radiation. Wavelength shifting fibres are used in combination with photomultiplier tubes to read out the signal from the scintillators. Both sections have three layers. The barrel has a total thickness of 9.2 λ at $\eta = 0$, including the ECal material (the tile calorimeter has a thickness of $\sim 7.4\ \lambda$). The endcap wheels of the HCal are composed of LAr as the active material with copper absorbers. This is because the forward region sees increased radiation. The endcaps each consist of four sampling layers and cover the region $1.5 < |\eta| < 3.2$. The endcaps are around 12 λ long.

2.5.3   Forward calorimeter

The forward calorimeter (FCal) is in the region closest to the beamline, covering $3.1 < |\eta| < 4.9$. It provides measurements of energy deposits of photons, electrons and hadrons. The FCal needs to be able to cope with high radiation in the forward region. It is about 9.5 λ long. It is required to ensure the calorimetry system provides full coverage. It is composed of three modules using LAr as the active material. The first module provides measurements of electromagnetically interacting particles and uses copper absorbers.

Calorimeter	Resolution
Electromagnetic	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$
Hadronic	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$
Forward	$\sigma_E/E = 28.5\%/\sqrt{E} \oplus 3.5\%$ (FCal1)
	$\sigma_E/E = 94\%/\sqrt{E} \oplus 7.5\%$ (FCal2 and FCal3)

Table 2.1: The energy resolutions of the different calorimetry components as described in Refs. [14–16]. The energy resolution is characterised by $\sigma_E/E = a/\sqrt{E} \oplus b$, where a is the stochastic term and b is the constant term.

The other modules are optimised for measurements of hadrons and use tungsten absorbers.

2.5.4 Energy resolutions

The energy resolutions of the different calorimeter components are shown in Table 2.1. The different calorimeter components have a range of energy resolutions suited to their physics and radiation requirements.

2.6 Muon spectrometer

The muon spectrometer (MS) is the outermost part of the ATLAS detector. A schematic is shown in Figure 2.5. It serves two purposes. First and foremost, it measures tracks of muons. The MS is also a key component of the trigger system described in Section 2.7. Superconducting toroid magnets generate a magnetic field which bends the muons' trajectories such that momentum and charge can be measured. Muons are minimum ionising particles, meaning they interact little with the detector material and do not form showers in the calorimeters. The muon spectrometer is covers a range of $|\eta| < 2.7$. In the barrel, the muon chambers are arranged cylindrically around the calorimetry system. The endcaps consist of large wheels that are orthogonal to the beamline. It has a radius of between 4.25 to 11 m from the beamline. It is the largest part of ATLAS. The magnets have a bending power of between 2-8 Tm. It is composed of precision muon tracking chambers and trigger chambers.

2.6.1 Precision muon tracking chambers

The precision muon tracking chambers are composed of Monitored Drift Tubes (MDTs) in the region $|\eta| < 2.7$, except for the innermost layer in the forward region $2.0 < |\eta| <$
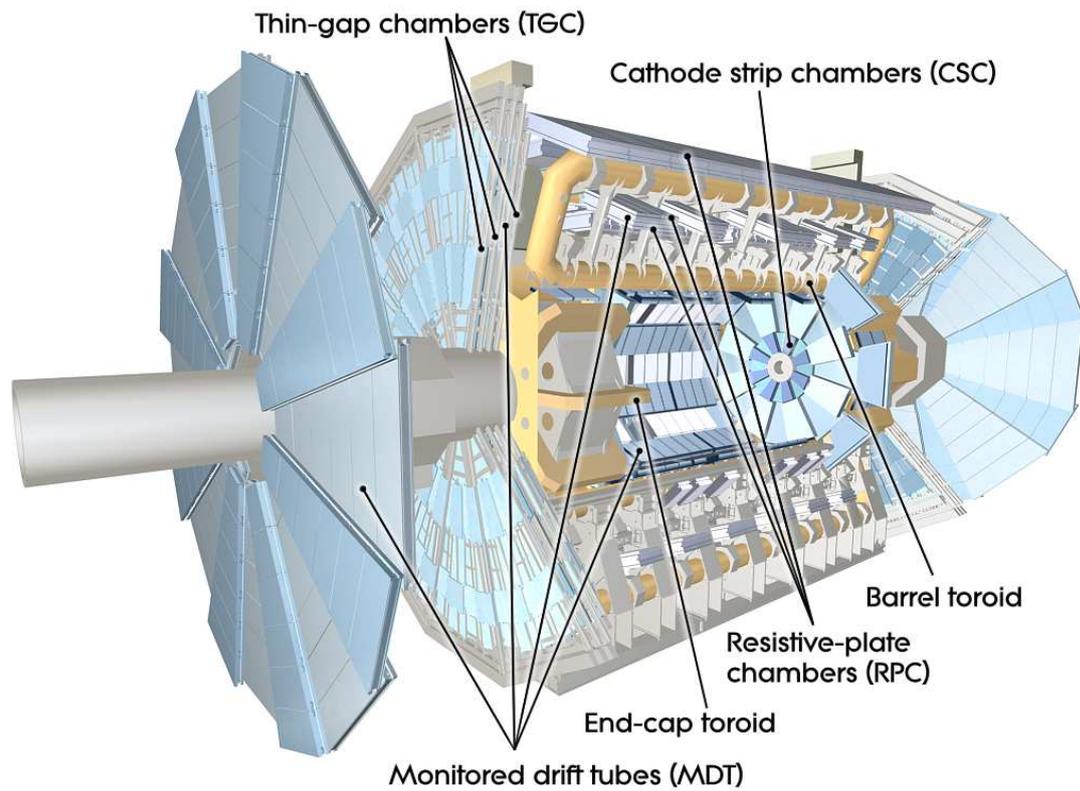
Figure 2.5: The ATLAS Muon System [7]. The components are arranged into large wheels covering the forward regions, as well as being arranged around the barrel. The muon detectors are placed around the outside of the detector as muons usually pass through the calorimeters.

2.7, where Cathode-Strip Chambers (CSCs) are used. MDTs consist of three to eight layers of drift tubes with a diameter of 30 mm. They are filled with Ar and CO_2 gas at a pressure of 3 bar. Ionising particles produce free electrons that are collected at the wires with a voltage of about 3000 V.

Tracking chambers in the forward region are exposed to high particle fluxes. This is why CSCs are used here – they have good time resolution and a higher granularity than MDTs. CSCs contain multi-wire proportional chambers filled with a gas mixture of Ar along with quenching agents CO_2 and CF_4. They contain planes of cathodes that are segmented into strips, and anode wires that are perpendicular to the cathodes. Muons ionise the gas, which produces avalanches of electrons that are collected at the anode wires. Positively charged gas ions induce charge at the cathodes. Due to the orthogonal layout of anodes and cathodes, a measurement of two coordinates is obtained for each traversing particle.

In order to produce a high precision muon trajectory, an optical alignment system is used to combine the MDT and CSC measurements. This system determines the positions of the MDTs and CSCs with respect to each other. Information from this alignment is used to apply corrections during offline reconstruction.

2.6.2 Trigger chambers

The trigger chambers are a specific part of the muon spectrometer system that provides fast tracking information. Resistive Plate Chambers (RPCs) are used in the barrel region covering $|\eta| < 1.05$. Thin gap chambers (TGCs) are used in the endcap region $1.05 < |\eta| < 2.4$.

RPCs are a sandwich of two parallel resistive plates with an electric field of 4.5 kV/mm and a tetrafluoroethane-based ($C_2H_2F_4$) gas mixture in between. Ionisation produces electrons that are accelerated towards the anode, producing electron avalanches en route. They have a fast response time of 60 μm and a 1 cm spatial resolution.

TGCs function in the same way as multi-wire proportional chambers. These provide good time resolution and high rate capability. They use a highly quenching gas mixture of CO_2 and n-pentane (C_5H_{12}). They have a response time of around 5 ns.

2.7   Trigger and data acquisition

The trigger and data acquisition system (TDAQ) is used to manage data collection. The amount of data produced within the ATLAS detector is far higher than the amount that can be read out or stored. In Run-2, ATLAS had a bunch crossing every 25 ns, corresponding to an event rate of 40 MHz [8]. This is equivalent to around 40 TB/s. TDAQ is used to select events with signatures that are interesting for analysis such
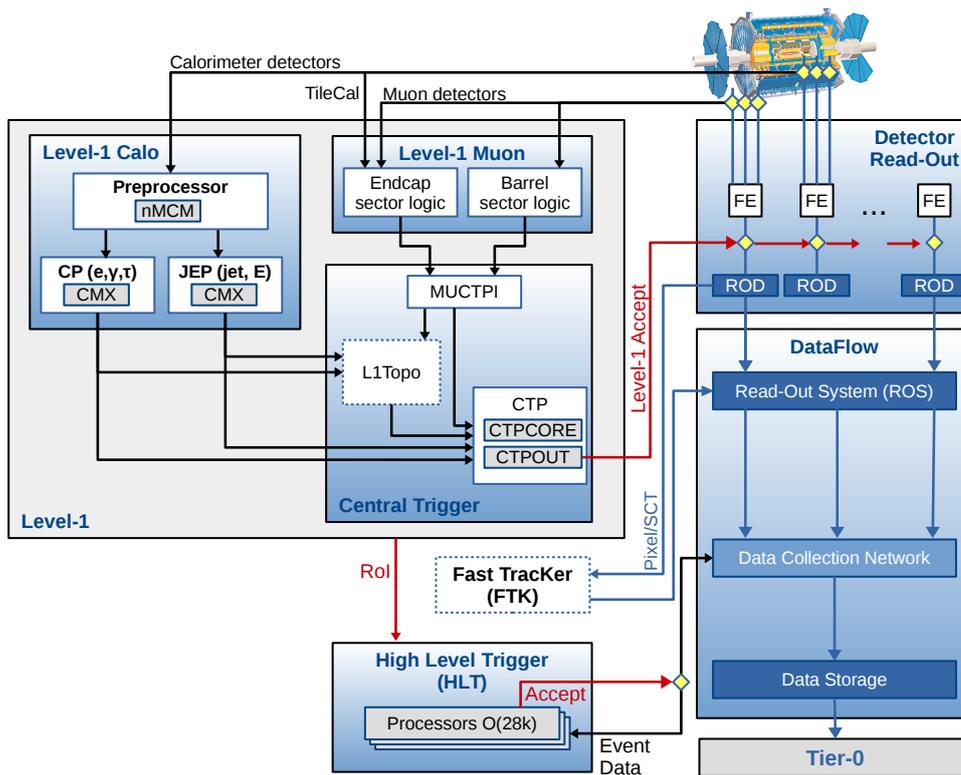
Figure 2.6: A flowchart showing the ATLAS TDAQ system [8]. Events that satisfy the hardware (L1) trigger requirements move onto the software-based High Level Trigger. Only events that pass both triggers are recorded and saved to disk.

that they can be permanently stored. The rest are rejected. TDAQ uses simplified reconstruction algorithms in order to select interesting events rapidly. Using the full offline reconstruction would take too long. To achieve the reduction in event rate, a two-level trigger system is used: the hardware-based Level 1 (L1) trigger and the software-based High Level Trigger (HLT).

The hardware-based L1 trigger rejects events based on information provided by the calorimeters and MS. The initial event rate of around 40 MHz is reduced to around 100 kHz by the L1 trigger in around 2.5 μs. After events have passed the L1 trigger, they are processed by the HLT using more detailed information that is not available at L1. This includes finer-granularity calorimeter information, precision MS measurements and ID tracking information. Depending on what is required, the HLT can reconstruct for the full detector, or in regions of interest defined by the L1 trigger.

2.8 Software

ATLAS uses a multi-tiered computing system to manage the vast amounts of data and CPU activity that arise from event reconstruction and analysis [80]. This is spread across the globe in the form of the Worldwide LHC Computing Grid (WLCG). From HLT to a published histogram the data moves through this system, entering from the event filter farm. This is located near ATLAS and is responsible for assembling experiment data into a stream to the Tier-0 centre. Tier-0 refers to the CERN data centre. It is responsible for the safekeeping of raw (RAW) experiment data, and performs the first pass at reconstructing raw data into a useable format. Tier-0 then distributes the raw data and reconstructed outputs to Tier-1 sites, in the form of Event Summary Data (ESD) and Analysis Object Data (AOD) files. Tier-1 sites are distributed around the world. These are responsible for additional reconstructions of raw data and producing new ESD and AOD files. Tier-2 centres are responsible for event simulation, in the same ESD and AOD format as real data. They are also responsible for physics analyses. Tier-3 is for physics analyses.

The Athena framework [81] is an enhanced version of the Gaudi framework [82]. Gaudi was originally developed by LHCb, but now a joint LHCb-ATLAS venture. All levels of data processing from the HLT to event simulation and reconstruction take place within the Athena framework. Key design principles in Athena are the separation between data and algorithms, as well as clear separation between transient and persistent data. Athena uses an Event Data Model (EDM) to ensure separation between data and algorithms. All physics objects are encapsulated in the EDM.

2.8.1 Object reconstruction

Physics objects that are reconstructed in ATLAS include:

Tracks are representations of the charged particle trajectories through the ID. Starting out as space-points called "hits" they are reconstructed in a multi-stage and iterative process with a Kalman Filter and Global χ^2 fitter [83]. Tracks form the input for vertex finding.

Vertices come in two types: primary and secondary. Primary vertices are the interaction points in the *pp* collisions. Secondary vertices are the points where particles produced in the primary *pp* interaction decay into secondary particles. Vertex reconstruction takes a full set of reconstructed tracks as input and returns a set of reconstructed vertices. The primary vertex is usually selected as the one with the highest $\sum p_{\mathrm{T}}^2$ [84].

Electrons are reconstructed from clusters in the ECal. They are matched with tracks from the ID and refit using a Gaussian Sum Filter [85].

Photons If no track is matched to an ECal cluster, the cluster is identified as an unconverted photon [85].

Muons are reconstructed using output from the muon spectrometer. Muon chamber hits are combined to ID track segments, which are then fitted together to determine the muon trajectory [86].

Jets are reconstructed from localised clusters in the HCal. The energy deposits in the calorimeter are reconstructed using the anti-k_t algorithm [87].

Transverse missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) quantifies the momentum imbalance in the transverse plane. It is computed as the negative vector sum of the transverse momentum of all calibrated jets, leptons and photons produced in the hard scatter, with a correction for any reconstructed tracks not matched to any calibrated objects applied [88].

The event data model (EDM) used in ATLAS Run-2 is based on the xAOD format [89]. In short, this format is an AOD that is directly readable from the ROOT software framework. The entire dataset collected by ATLAS is very large. As such, physics analysis and combined performance groups define derivations of the xAOD, called derived-xAOD or DxAOD. These are smaller subsets of the full xAOD collection, that can be replicated to more grid-sites and allow users to run analyses much more quickly. These may remove whole events based on pre-set criteria, thin the remaining events by removing non-required objects and slim the objects by removing non-required variables from the objects. Derivations can also be augmented – if using alternative setups to the standard ATLAS reconstruction chain, new objects can be added, and existing objects can be decorated with new variables.

2.8.2 Detector simulation

A key part of ATLAS results are tests of the hypothesis that the observed results agree or disagree with those predicted by the Standard Model (or any new physics model). This requires that *pp* collisions are fully simulated and reconstructed in the same software framework as the real data, resulting in a format identical to the output of the ATLAS data acquisition system. This section gives an overview of how this is achieved.

The ATLAS simulation programme [90] is integrated into Athena, and uses the Geant4 toolkit. The simulation of events generally runs in a three step chain:

1. Generation of the *pp* interaction and immediate decays (described in more detail in Section 2.11).

2. Simulation of the detector and physics interactions.

3. Digitisation of the energy deposits in the sensitive regions of the (emulated) detector into voltages and currents for comparison to the readout of the ATLAS detector.

The simulated data can be presented in any of the formats used in Athena chain, allowing simulated data to run through the same reconstruction software as real data. The generated information ("truth"), including the histories and origin of decayed particles, can be stored if needed. The reconstruction can then be assessed by comparing against the "truth" information. The ATLAS detector geometry is built from databases of physical construction and conditions data.

Simulation is very resource intensive, sometimes so-called "fast" simulations are used to generate the necessary statistics for an analysis in a reasonable time frame. These methods sacrifice some accuracy for speed. This is done, for example, by using a simplified geometry of the detector. Another approach involves bypassing the CPU intensive pattern recognition stage for track reconstruction by using MC generator based trajectory building [91]. In the analysis described in Chapter 4, fast simulations are used to estimate systematics from the choice of MC generator, where it would have been too resource intensive to produce full simulations for all generator configurations.

2.9 Luminosity

The *pp* collisions in ATLAS are stochastic processes, and the interaction probability comes in the form of a cross-section, σ. The number of *pp* collisions occurring per unit time, $\mathrm{d}N/\mathrm{d}t$, is then

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathcal{L}\sigma, \tag{2.1}$$

where $\mathcal{L}$ is the instantaneous luminosity, a proportionality factor that encapsulates the relevant beam parameters. It is calculated as:

$$\mathcal{L} = f_{\text{rev}} n_b \frac{N_1 N_2}{2\pi \Sigma_x \Sigma_y},$$ (2.2)

where f_{rev} is the beam revolution frequency, n_b is the number of bunches colliding per second, N_1 and N_2 are the number of protons in the two bunches, and Σ_x and Σ_y are the beam spread in the x and y directions respectively. This can also be written as:

$$\mathcal{L} = \frac{N_1 N_2 f_{\text{rev}} n_B}{4\pi \beta^* \epsilon_{xy}} F$$ (2.3)

where:

- ϵ_{xy} is the transverse beam emittance, an indication of the quality of the beam due to the bunch preparation all the way back to the start of the injection chain. If a particle beam has low ϵ_{xy}, then the particles have nearly the same momentum and are confined within a small distance. It is more common to give a value of the normalised transverse beam emittance, $\epsilon_n = \epsilon_{xy} \gamma_r \beta_r$, where γ_r and β_r are the relativistic gamma and beta factors. The design value for the collision point in ATLAS is $\epsilon_n = 3.75$ μm, and for typical pp collisions in Run-3 it had values ranging between 1.9 and 2.4 μm [92].

- β^* is the amplitude function at the interaction point, defined as the distance at which the beam is twice the width as at the interaction point. It has units of length. A low β^* means that beam is narrower and squeezed, whilst a high β^* means the beam is wide and straight. The design value for the collision point in ATLAS is 0.55 m, and for typical pp collisions in Run-2 it had values ranging from 0.3 to 0.8 m [92].

- F is the geometric reduction factor, which depends on the bunch length and crossing angle.

The instantaneous luminosity, L, is the delivered instantaneous luminosity over a given period of time. It is calculated as

$$L = \int \mathcal{L} \, \mathrm{d}t.$$ (2.4)

Comparing equations (2.2) and (2.1), it can be seen that the units of L are inverse area, and at the LHC and ATLAS they are given in fb^{-1}. The integrated luminosity for the analysis described in Chapter 4 is 139 fb^{-1}, which represents the dataset accumulated over Run-2 between 2015 and the end of 2018.

The number of pp collisions in a given bunch crossing is distributed according to Poisson statistics, meaning that there are many pp collisions in a single stored event. Typically it is the pp interaction with the highest Σp_{T}^2 that is of interest for analysis. The other

pp interactions are a background, referred to as pile-up. The amount of pile-up depends on the luminosity. If the pile-up occurs in the same bunch-crossing then it is known as in-time pile-up, whereas if it results from interactions with surrounding *pp* bunches then it is known as out-of-time pile-up. In-time pile-up results from a small bunch crossing (i.e. a more focussed beam), and out-of-time pile-up results from small bunch spacing.

The luminosity delivered to ATLAS is measured by two systems in the forward region. These are LUCID-2 (LUminosity measurement using Cerenkov Integrating Detector) [93], located at ± 17 m from the interaction point, and ALFA (Absolute Luminosity for ATLAS) [94], located at ± 240 m from the interaction point. LUCID detects inelastic *pp* scattering in the forward direction, and acts as the main online relative-luminosity monitor for ATLAS. ALFA consists of scintillating fibre trackers located in Roman pots[3] which are designed to approach as close as 1 mm to the beam. During Run-2, ATLAS recorded a *pp* dataset with a total integrated luminosity of 139 fb^{-1} $\pm 1.7\%$ [95], measured using the LUCID-2 detector [93].

2.10   Increasing pileup

The ATLAS detector was designed to operate with a constant instantaneous luminosity of $\mathcal{L} = 1.0 \times 10^{34}$ cm^{-2}s^{-1}, 14 TeV centre-of-mass energy, a bunch spacing of 25 ns and and average pile-up of 23 proton-proton interactions [7]. During Run-2, the LHC and ATLAS far exceeded the design parameters, running in 2018 with a peak luminosity of $\mathcal{L} = 1.9 \times 10^{34}$ cm^{-2}s^{-1} and a peak (average) pile-up of 55 (36) [95]. In spite of the much higher pile-up, ATLAS has performed adequately for physics analyses. This luminosity is being pushed up even further. In Run-3, expected to start in 2021 and end in 2023, ATLAS is expected to receive a luminosity of up to around $\mathcal{L} = 3 \times 10^{34}$ cm^{-2}s^{-1}. Run-3 will act as a stepping stone to the High-Luminosity LHC (HL-LHC) [96]. HL-LHC will maximally exploit the LHC infrastructure and requires significant upgrades. It is expected to start running in 2026, and will have a peak (ultimate) luminosity of $\mathcal{L} = 5 \times 10^{34}$ cm^{-2}s^{-1} ($\mathcal{L} = 7.5 \times 10^{34}$ cm^{-2}s^{-1}) and an average $\langle \mu \rangle = 140$ (200).

The increased luminosity and pile-up, as well as the increased data rate and accumulated radiation damage will require significant hardware and software upgrades after Run-3. Even for Run-3, Phase-I hardware upgrades and significant software upgrades are underway to cope with the new challenges presented by the harsh new conditions. One such area, presented in the next chapter, is primary vertex reconstruction software, which is particularly impacted by pile-up contamination.

The analysis described in Chapter 4 is impacted the modelling of pile-up, with an uncertainty estimated by reweighting the simulated events to change the amount of pile-up.

---

[3] A "Roman pot" (proton-on-target) is the name of the technique and hardware used to localise the trajectories of protons in a beam.

This is due to the fact that different pile-up conditions have different effects on the muon p_{T} modelling.

2.11   Event simulation

This section provides a brief outline of how Monte Carlo simulations of pp interactions are performed.

The simulation of pp interactions in ATLAS occurs in several stages, with the different steps corresponding to different stages of calculations of the interaction. These can be summarised as:

1. Hard scatter interaction;

2. Parton showering;

3. Hadronisation;

4. Decays of unstable hadrons;

5. Underlying event;

6. Detector simulation;

These different stages are illustrated in Figure 2.7. These different stages and calculations are usually handled by different algorithms. Protons are composite particles – they are composed of partons: quarks and gluons. It is the interaction between the partons of colliding protons that are observed in ATLAS. The properties of the partons are described by parton distribution functions (PDFs). These are functions of the fraction of longitudinal momentum carried by a parton, x, and the momentum of the parent proton, Q – i.e. $f(x, Q^2)$. They are defined as the probability density for finding a particle with a certain longitudinal momentum fraction x at resolution scale Q^2. Due to the fact that partons are not observed as free particles due to colour confinement, PDFs cannot be calculated by perturbative QCD and must be taken from data. Top quarks and other heavy particles can be produced when the momentum transfer between two colliding partons is large compared to the QCD scale, and are therefore often referred to as "hard scatter" interactions. A majority of parton interactions do not produce particles of interest in ATLAS, since they have much lower momentum transfers. Such interactions are therefore referred to as "soft". The cross sections of hard scatter events are significantly lower than for soft interactions. As such, MC samples of hard scatter events are generated by defining specific initial and final state particles. The modelling of pile-up pp collisions does not include a hard scatter, only soft interactions. Such interactions are known as "minimum-bias" interactions.
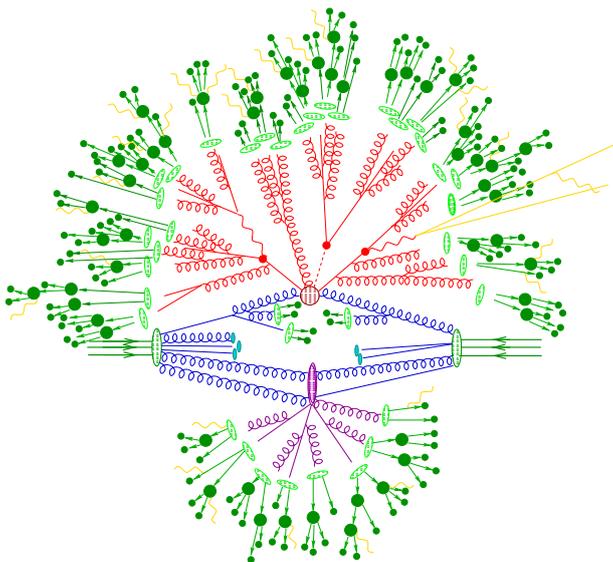
Figure 2.7: A diagram of a Higgs boson produced in association with $t\bar{t}$, as produced by an event generator. As shown in Ref. [9]. The hard scatter interaction is indicated by the large red circle, and the decays of the $t\bar{t}$ and Higgs boson are indicated by the small red circles. Additional hard QCD radiation is produced (indicated by red lines) and a secondary interaction takes place (purple), before the final-state partons hadronise (indicated by light green) and hadrons decay (indicated by dark green). QED radiation can occur at any stage (yellow).

Hard scatter interaction

MC event generators calculate the relevant transition matrix elements that describe the transition between the specified initial and final state particles in order to simulate the desired hard scatter process. The transition probability is the squared amplitude of the matrix element. The initial state particles are described by the PDFs – they can be either valence quarks, sea quarks or gluons. At the energy scales considered in *pp* collisions in ATLAS, the colliding partons in hard scatter interactions can be considered asymptotically free, meaning that the they can be considered as quasi-free particles and their interactions are well described by perturbative QCD. The PDFs are not described by perturbative QCD and must be taken from data. The cross section of a process is calculated by integrating the probability density function, which is given by the transition matrix element and the PDFs, over the full phase space. This integral is approximated by a random sampling of the phase space using MC methods. The transition matrix calculation is performed at a fixed-order in perturbation theory, e.g. leading order (LO), next-to-leading order (NLO) or next-to-next-to-leading order (NNLO). Bottom quarks are significantly heavier than the proton and require special treatment in the calculation. There are two ways in which *b*-quarks can be included in the initial state particles. The first is the four-flavour scheme (4FS), where the *b*-quarks are assumed to come from gluon splitting and not from the proton. The second is the five-flavour scheme (5FS), which assumes that there is a bottom quark contribution from the proton. Examples of different

generators include PoWHEG [97], MadGraph5_aMC@NLO [98] and Sherpa [9].

Parton showering

The initial state particles are able to radiate gluons, which in turn can radiate further gluons or quark-anti-quark pairs. Any colour-charged particles produced in hard scatter may also produce QCD radiation, creating cascades of partons. The basic principle of a parton shower algorithm, then, is to approximate in a probabilistic way the higher order corrections to the hard scatter. It is not feasible to calculate these directly, hence an approximation is used with the dominant contributions included in each order. The dominant contributions are associated with soft and collinear parton splittings, meaning the showering algorithms describe the decrease of momentum transfer scale of the partons, before hadronisation processes become relevant. Electrically charged particles may also produce QED radiation at any stage and this is treated by similar showering algorithms.

Shower algorithms are normally inaccurate for hard, wide angle emissions (such as additional well-resolved jets), due to the fact they are they use a combination of small-angle (collinear) and soft approximations. One technique to overcome this is Matrix Element + Parton Shower (ME+PS) matching, which uses tree-level matrix elements for hard, large-angle emissions. It was first formulated in the Catani-Krauss-Kuhn-Webber (CKKW) paper [99]. Another method of calculating jet cross sections includes the use of Catani-Seymour dipoles [100]. The NLO generators can be improved in such a way that the production of associated jets achieves NLO accuracy using, for example, the MEPS@NLO [101] method.

Hadronisation

As the parton showers evolve, the momentum decreases to the hadronisation scale. At this scale, the partons can no longer be treated as quasi-free particles, and their confinement must be considered. Hadronisation refers to the process by which the set of coloured partons produced by the showering algorithms are turned into a set of "primary hadrons", which may subsequently decay further. These processes cannot be determined from first principles and must be modelled instead. The most common techniques include the string model and cluster hadronisation model. The modelling of the parton shower, hadronisation and underlying event are usually handled by a single showering program, using parameter sets that are tuned to data. These parameter sets are referred to as "tunes". Examples of different showering and hadronisation programmes include Pythia [102, 103] and Herwig [104, 105].

Decays of unstable hadrons

The decays of the set of primary hadrons calculated in the hadronisation stage must also be simulated. These are usually not stable on the timescales considered in ATLAS, and their decays are modelled by specific algorithms. These algorithms use measurements of hadron decays and theoretical calculations of the decay amplitudes to simulate the decays to either "secondary hadrons" or stable particles. The decays of *b*- and *c*- hadrons is particularly important for the physics considered in ATLAS, and these are performed by EvtGen [106].

Underlying event

Any additional activity beyond the hard scatter process, the initial state radiation activity and final state radiation activity is referred to as the "underlying event" (UE). It is believed that the dominant contribution to the UE comes from additional colour exchanges between the colliding hadronic states.

The UE is only defined in the context of a hard scatter interaction, unlike minimum bias interactions which do not require a hard scatter process.

Detector simulation

The simulated events are passed into the simulation software chain described in Section 2.8.2, and the detector responses to the simulated *pp* interactions are emulated. These simulations pass through the same reconstruction chain as real data. Unlike real data, all of the information about simulated events, including the entire history of final state particles, can be saved. This is known as MC truth information. This is a vital tool in many analyses.

2.12   Summary of chapter 2

This chapter has given a brief overview the LHC and ATLAS detector, the amazing machines used to generate an enormous dataset of *pp* collisions. In addition, an overview of *pp* event simulation is provided. These are described here to provide a better understanding of the content in the following chapters. Within this dataset are many top-quark pairs with dileptonic final states, which form the basis of the test of Lepton Flavour Universality in Section 4. But first, in the next chapter, a study of improvements to primary vertex reconstruction is presented.

3

Improvements to primary vertex reconstruction in ATLAS

3.1 Introduction

A primary vertex is the interaction point of a *pp* collision. Vertex reconstruction takes a set of reconstructed tracks as input and returns a set of reconstructed vertices. Primary vertices are the ultimate origins of all the reconstructed objects used in physics analyses. Therefore, accurate and efficient vertex reconstruction is required in order to fully and accurately reconstruct the kinematic properties of an interaction. As described in Section 2.9, when bunches of protons collide, there is a Poisson-distributed number of *pp* collisions. This means that when an event satisfies the trigger requirements, a number of pile-up events are recorded in addition to the interaction of interest. Usually, the interaction of interest has the highest Σp_{T}^2. This is why it is also known as the "hard-scatter" (HS). The hard-scatter vertex (HSV) and its reaction products must be isolated from the pile-up that occurs simultaneously nearby.

The increasing luminosity and average pile-up (μ) pose new challenges for ATLAS. This includes degrading vertex reconstruction performance. In the course of Run-2, ATLAS saw more than 60 inelastic *pp* collisions per bunch crossing, and with the planned increases to luminosity, Run-3 is expected to see even more. Figure 3.1 shows particle tracks of $p_{\mathrm{T}} > 100$ MeV for a $Z \to \mu\mu$ candidate event with a pile-up of $\mu = 65$, recorded in 2017. This high track multiplicity is a good illustration of the challenges facing vertex reconstruction in ATLAS. Thus, new and better primary vertex reconstruction tools are required. Starting in Run-3, ATLAS will begin using the adaptive multi-vertex finder (AMVF) [107, 108] and Gaussian density seed finder (GSF) [1]. These tools have been tuned to achieve sizeable improvements in vertex reconstruction efficiency and other performance metrics in the new pile-up conditions.

This chapter describes the strategy, optimisation and preliminary performance of this

Figure 3.1: A display of $Z \rightarrow \mu\mu$ candidate event from pp collisions recorded by ATLAS with LHC stable beams at a collision energy of 13 TeV. The Z boson candidate is reconstructed in a beam crossing with a pile-up of $\mu = 65$. Tracks with $p_\mathrm{T} < 100$ MeV are shown, with the two muon tracks highlighted in yellow. The large number of pile-up create a more challenging environment for reconstructing the hard scatter vertex. ATLAS Experiment ©2020 CERN.

new reconstruction strategy, with a focus on improvements in vertex finding efficiency, purity and spatial resolution in expected Run-3 pile-up conditions. The bulk of this work has been published in the form of a public note (see Ref. [1]) and conference proceedings (see Ref. [2]).

Section 3.2 describes the simulated data samples, as well as the truth matching criteria and vertex classifications used in the studies for this chapter. Section 3.3 gives an overview of the Adaptive Vertex fitting technique. Section 3.4 describes the Run-2 primary vertex reconstruction strategy, for comparison. Sections 3.5, 3.6 and 3.7 describe the development of the GSF and AMVF for high pile-up conditions. Section 3.8 shows the results of this optimisation. The author's research and findings are detailed in Sections 3.7.1 and 3.8.

3.2 Simulated data samples and truth matching criteria

3.2.1 Monte Carlo generated samples

The studies in this chapter use Monte Carlo (MC) generated samples. These were generated at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Each MC event contains a single HS event, overlaid with a Poisson-distributed number of random inelastic pp interactions (pile-up). Two very different types of HS interaction were studied separately: top-quark pairs (denoted as $t\bar{t}$ in the rest of this chapter) and Higgs bosons produced via Vector Boson Fusion (VBF) and decaying to undetected particles (VBF $H \rightarrow 4\nu$ in the rest of this chapter). These two processes were chosen as they are representative of high- and low-multiplicity HS vertices respectively. $t\bar{t}$ decays are described in Section 1.5.2, where it is noted that they are characterised by a number of high p_T jets and/or leptons. VBF $H \rightarrow 4\nu$ is characterised by two energetic jets spaced widely in η, with $\mathcal{O}(100)$ GeV missing transverse momentum. Both of the samples are produced with an average number of interactions per bunch crossing $\langle \mu \rangle$=60.

The $t\bar{t}$ pairs were generated using POWHEG [97,109] interfaced to PYTHIA6 [102] with the Perugia 2012 tune [110]. The POWHEG h_{damp} factor (a model parameter that controls the matching between matrix element and the parton shower and effectively regulates the high p_T radiation) was set equal to the top mass. The $t\bar{t}$ pairs were generated with a lepton filter, requiring at least one lepton in the final state. The VBF $H \rightarrow 4\nu$ interactions were generated using POWHEG interfaced to PYTHIA8 with the AZNLO tune [111] and the CT10 parameterisation [112] of the parton density functions. All of the overlaid pile-up collisions were simulated using the soft QCD processes of PYTHIA8 with the A3 [113] set of tuned parameters and the NNPDF2.3LO parton density function set.

All generated events are processed with the ATLAS detector simulation framework, described in Section 2.8.2. After full detector simulation, the MC events are reconstructed and analysed using the same software and format as real data would be.

3.2.2 Truth matching and vertex classification

The benefit of using simulated events to tune vertex reconstruction software is that it is possible to quantify performance by comparing the reconstructed vertices and their associated tracks against the true simulated interactions. The truth matching procedure is similar to that described in Ref. [114]. The reconstructed tracks are associated to truth particles according to the correspondence of simulated hits in the Inner Detector (see section 2.4). The tracks are all matched back to the primary generated interaction using the stored truth particle histories. This means that tracks produced in secondary decays are also traced back to a HS or pile-up interaction. Reconstructed tracks that do not have conclusive matching information are classified as "fake". Tracks that are not

associated to any truth particle are ignored, since without any information they cannot be classified[1]. This means that each reconstructed track receives one of the following classifications:

- A track matched to the hard-scatter interaction.

- A track matched to a pile-up interaction.

- An unmatched track. These tracks may be random combinations of detector hits that are incorrectly identified as charged particle trajectories.

These track classifications, along with their fit weights (as described later in sections 3.4 and 3.6) and p_T are used to classify the vertices. This is done by normalising the sum of the weights assigned to all contributing tracks to unity. The fractional weights of the individual tracks are calculated. The vertices then receive a classification based on this weight as follows:

CLEAN: at least 70% of the total track weight in the reconstructed vertex originates from a single simulated pp interaction.

MERGED: less than 70% of the total track weight in the reconstructed vertex originates from a single simulated pp interaction. This means that two or more simulated interactions contribute significantly to the accumulated track weight.

SPLIT: a single simulated pp interaction contributes the largest track weight to two or more reconstructed vertices. The reconstruction with the largest track Σp_T^2 is classified as either CLEAN or MERGED, whilst the other(s) are classified as SPLIT.

FAKE: fake tracks contribute more weight to the reconstructed vertex than any simulated pp interaction.

As noted in Ref. [114], this classification scheme was designed to study the effects of vertex splitting and merging, and the influence of these effects on the primary vertex efficiency and resolution. The studies described in this section use the same classification scheme for consistency.

The quality of the reconstructed HS vertex is also used to classify the event as a whole:

CLEAN/MATCHED: the event contains a CLEAN reconstructed vertex originating from the true HS interaction, and the HS interaction does not contribute more than 50% of the accumulated track weight to any other vertex.

LOWPU (low pile-up): the event contains a MERGED vertex with at least 50% of the accumulated track weight coming from the simulated HS interaction.

HIGHPU (high pile-up): the event contains a MERGED vertex with its main contribution coming from a simulated pile-up interaction, and in which the simulated

---

[1]This was agreed after discussion with the ATLAS tracking group.

HS interaction contributes between 1% to 50% of the accumulated track weight.

PUREPU (pure pile-up): the event does not contain any vertex with at least 1% accumulated track weight from the HS interaction.

3.3 Adaptive vertex fitting

This section outlines the adaptive vertex fitting technique as described in Ref. [115]. Vertex fitting is the process of computing the position and covariance matrix of a vertex given a set of reconstructed tracks. In ATLAS, this proceeds using implementations of a Kalman filter. The Kalman filter is a least-squares estimator. The vertex is fitted by minimising the sum of the standardised distances ($\chi = d/\sigma(d)$) of the associated tracks to the vertex position v:

$$\hat{v}_{\text{LS}} = \arg\min_v L(v), \tag{3.1}$$

where:

$$L(v) = \frac{1}{2}\sum_{i=1}^{n}\chi_i^2(v) = \frac{1}{2}\sum_{i=1}^{n}d_i^2(v)/\sigma_i^2. \tag{3.2}$$

The equation for $\hat{v}$ can be obtained by differentiating equation (3.2) with respect to v:

$$\frac{\partial L(v)}{\partial v} = \sum_{i=1}^{n}\chi_i(v)\frac{\partial\chi_i}{\partial v} = 0. \tag{3.3}$$

A first order Taylor expansion is used to approximate the distance d_i by an affine function of v:

$$d_i(v) \approx c_i + a_i^T v. \tag{3.4}$$

Doing this means that equation (3.3) becomes a linear equation for $\hat{v}$ that can be solved explicitly with the Kalman filter.

In an ideal setting, only tracks belonging to the vertex of interest would be associated to it for the fit. In practice, due to pile-up and detector effects, mis-associated tracks and tracks with mis-measured errors are usually included in the fit. Least-squares estimators are sensitive to these effects. To ensure a robust vertex fit, a weight can be calculated for each track used in the fit, such that equation (3.3) becomes:

$$\sum_{i=1}^{n}\omega_i\chi_i(v)\frac{\partial\chi_i}{\partial v} = 0, \tag{3.5}$$

i.e. the influence of each track is reduced by a weight ω_i.

The adaptive vertex fitter implements equation (3.5) with the weights calculated by the adaptive estimator:

$$\omega(\chi^2, T) = \frac{1}{1 + e^{-\frac{1}{2}(\chi_{\text{cut}}^2 - \chi^2)/T}}, \tag{3.6}$$
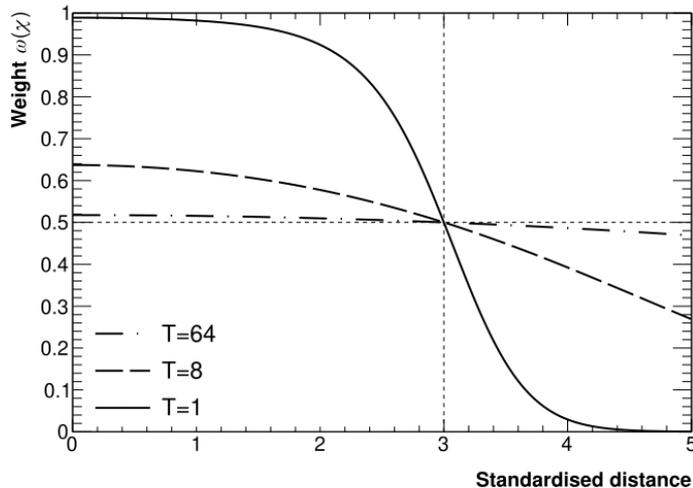
Figure 3.2: The adaptive weights (ω) calculated as a function of standardised distance (χ) for three different temperatures. At high temperatures, the weights have little dependence on χ. As the temperature is lowered, the function becomes more sensitive to χ.

where χ^2_{cut} represents a cutoff point, nominally set at 9 (corresponding to 3σ). T corresponds to "temperature" – a user defined set of variables to control the sensitivity of the vertex fit to track-to-vertex compatibility. The weights are calculated for an initial vertex position, and the vertex is estimated using these weights. This procedure is repeated until convergence. The temperature T is used to apply a deterministic annealing approach to avoid local minima. The estimation of the vertex position begins at a user-defined initial temperature $T_{\text{initial}} > 1$, and is lowered in a well-defined sequence that converges to 1. The iteration is stopped when either the temperature is equal to 1, or if the vertex candidate position does not change by more than 1 μm. Figure 3.2 shows ω as a function of χ for three different temperatures.

3.4 Vertex reconstruction in ATLAS Run-2

Prior to the upcoming Run-3, ATLAS used the "Iterative Vertex Finder" (IVF) to reconstruct primary vertices. This was designed for a much lower multiplicity than is expected in Run-3 and beyond [114]. This section summarises the IVF procedure [114] for comparison.

3.4.1 Procedure

Primary vertex reconstruction takes a set of reconstructed tracks as input. This also includes the transverse and longitudinal impact parameters of the tracks, d_0 and z_0

52
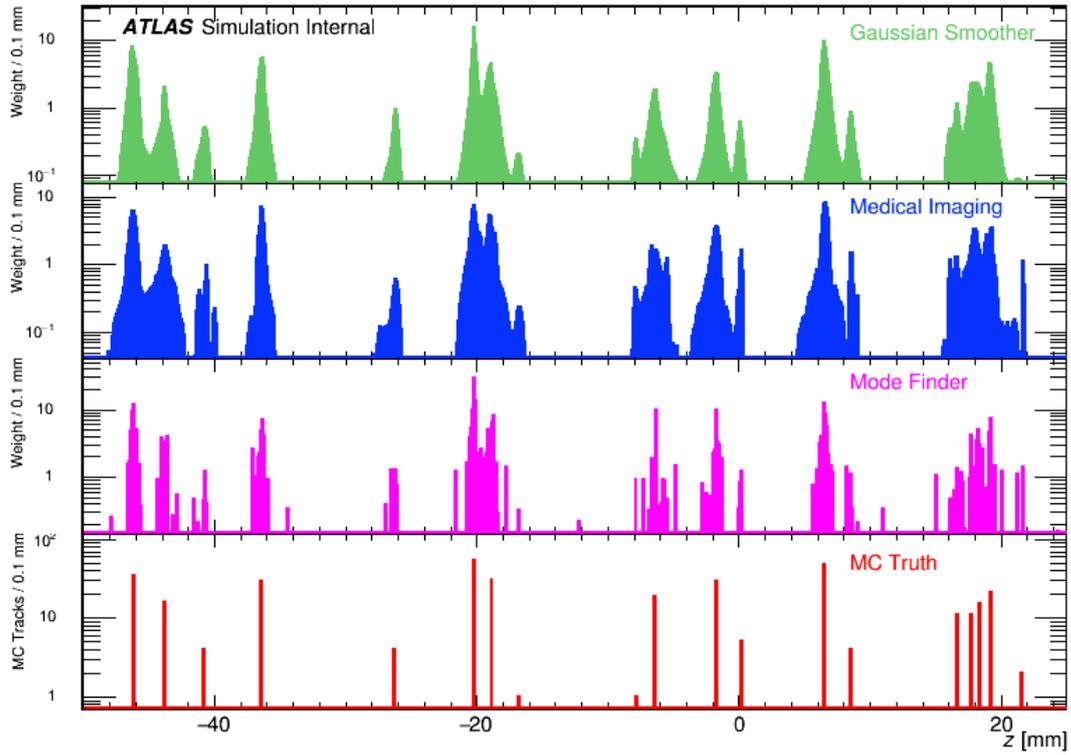
Figure 3.3: A single MC event analysed by three different primary vertex seed finders. The bottom plot indicates the locations of the *pp* interactions along the beamline as generated in MC, along with the number of tracks passing quality cuts. The top three plots show the track density estimators computed using the different seed finders. The approximate locations of the primary vertices, including pile-up interactions, are indicated by local maxima. The FSMW mode finder (labelled "Mode Finder") was used by ATLAS during Run-1 and Run-2. The medical imaging and Gaussian methods were both studied to improve high luminosities, with the Gaussian seed finder (labelled "Gaussian smoother") to be used from Run-3 onwards. As shown in Ref. [1].

respectively, as well as their associated uncertainties $\sigma(d_0)$ and $\sigma(z_0)$. These impact parameters are measured as the closest approach of the track with respect to the beamspot, which is the luminous region of the detector. In addition, the vertex reconstruction is constrained to the location and size of the beamspot [116].

The IVF reconstructs vertices one at a time in an iterative procedure involving four steps:

Seed-finding: A vertex seed is the most likely position of a new primary vertex. This is determined using tracks not yet assigned to any vertex candidate. The IVF uses a "Fraction of Sample Mode with Weights" (FSMW) [117] mode-finder. This estimates the point of maximum track density along the beam axis by recursively scanning a weighted histogram of z_0 (see Figure 3.3).

Track to seed assignment: After a vertex seed is found, tracks are assigned to it according to their impact parameter significance. Tracks are selected for fitting if they are not assigned to any previously fitted vertex and if their closest approach to the seed is within $12 \cdot \sqrt{\sigma^2(d_0) + \sigma^2(z_0)}$. This combination of seed and assigned tracks is then referred to as a vertex candidate.

Fitting: The helical parameters of the tracks in the vertex candidate are linearised at the seed position using an adaptive Kalman filter [115, 118]. During the fit, the transverse position is constrained by the beamspot, whilst the longitudinal position is left unconstrained. Outlying tracks are progressively de-weighted using a deterministic annealing schedule and the weight calculated using the adaptive estimator:

$$\omega(\chi^2, T) = \frac{1}{1 + e^{-\frac{1}{2}(\chi^2_{\text{cut}} - \chi^2)/T}}. \tag{3.7}$$

During annealing, the temperature, T, is decremented in stages, which increases the sensitivity of the track weights to χ^2, the square of the standardised distance to the vertex position. This dependence is shown in Figure 3.2. In the final stage of the annealing process, $T = 1$. At this point, compatible tracks should have weights close to unity, and incompatible outliers will have weights close to zero. This is achieved by setting $\chi^2_{\text{cut}} = 9$ such that compatible tracks are within three standard deviations. The set of temperatures is defined as $T = [64.0, 16.0, 4.0, 2.0, 1.5, 1.0]$.

Acceptance/rejection: The fitted vertex candidate is accepted if the number of degrees of freedom[2] (accounting for the final track weights) is greater than three and at least two tracks used in the fit have weights greater than 0.01. If it does not meet these requirements, then all tracks used in the fit are removed from the pool used to find new vertex seeds. If the fitted vertex candidate is accepted, associated tracks with a fit weight greater than 0.01 or an impact parameter significance less than 7σ (including both the track and fitted vertex errors) are removed from the pool used to find new vertex seeds. Tracks that fail both compatibility cuts are

---

[2]An individual track contributes two degrees of freedom to the vertex fit.

no longer associated to the fitted vertex, and remain in the pool to find new vertex seeds. This ensures that the pool of tracks to find new vertex seeds is always reduced in each iteration, whether the fitted vertex candidate is accepted or not.

The iterative procedure terminates when:

- less than two tracks remain in the pool to find new vertex seeds, or

- less than two tracks satisfy the seed proximity requirement for fitting, or

- the maximum number of allowed reconstructed vertices is reached (a failsafe to handle data quality issues).

3.5 Gaussian Density Seed Finder (GSF)

Seed finding is necessarily the first step in the vertex reconstruction process. It is an estimate of the most likely point of a vertex that is used as the starting point in the vertex fit. Therefore, it is unlikely that any vertex not found by the seed finder will be reconstructed.

The GSF was developed by ATLAS for Run-3 with higher levels of pile-up in mind [1]. Design considerations included:

- One-dimensional finding.

- To incorporate individual track measurement errors.

- To be CPU-efficient, including allowing calculations to be cached.

3.5.1 Gaussian density function

The Gaussian density seed finder finds the approximate location of pp interactions by analysing the track density along the beamline. The tracks are modelled as correlated radial and longitudinal Gaussian probability distributions, $P(r, z)$ centred at (z_0, d_0)[3] and normalised to one. Their shapes are characterised by the covariance matrix Σ of the corresponding helical parameters:

$$\Sigma = \begin{pmatrix} \sigma^2(d_0) & \sigma(d_0, z_0) \\ \sigma(d_0, z_0) & \sigma^2(z_0) \end{pmatrix}$$

[3]Where z_0 is the closest approach of the track in the longitudinal direction, d_0 is the closest approach in the transverse direction.

The density of a single track can then be evaluated at any longitudinal position z along the beamline. The seed finding density is defined as $\rho(z) = P(0, z)$ where

$$P(r, z) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}((r-d_0),(z-z_0))^T \Sigma^{-1} ((r-d_0),(z-z_0))}.$$

The total density as a function of z is then the sum of all nearby tracks:

$$W(z) = \sum_{i \in \text{tracks}} \rho_i(z) = \sum_{i \in \text{tracks}} P_i(0, z).$$

For the sake of illustration, one can simplify the track density $\rho(z)$ by ignoring correlations:

$$\lim_{\sigma(d_0, z_0) \to 0} \rho(z) = \left(\frac{1}{\sqrt{2\pi}\sigma(d_0)} e^{-\frac{1}{2}d_0^2/\sigma^2(d_0)} \right) \left(\frac{1}{\sqrt{2\pi}\sigma(z_0)} e^{-\frac{1}{2}(z-z_0)^2/\sigma^2(z_0)} \right)$$

3.5.2 Peak search

The global maximum of $W(z)$ can be interpreted as the most likely position of a vertex, i.e. the seed position. The density $W(z)$ is a sum of known analytic functions, meaning the first and second derivatives of $W(z)$ at any point can be calculated at the same time as $W(z)$. The search for the peak can be sped up by exploiting the fact that the maximum must lie near a track. Therefore $W(z)$ and its derivatives are evaluated at the value of z_0 for each track in the seed finding pool. The sign of the second derivative is used to determine whether a track is located near a maximum. If the sign is negative, indicating the track lies near a local maximum, the z step to that maximum ($W'(z) = 0$) is estimated. For a parabolic distribution, the step is simply a Newton step:

$$\Delta z = -\frac{W'(z)}{W''(z)}$$

whereas the z step of a locally Gaussian distribution is described by:

$$\Delta z = \frac{W(z)W'(z)}{W'^2(z) - W''(z)W(z)}.$$

The Gaussian approximation is used by default, since it gives better convergence. After the first step is computed, the procedure is repeated one time, using the adjusted trial position. This is to refine the position of the local maximum. Further repetitions are deemed unnecessary. If the second step produces a higher maximum than the first, the global maximum position and height are updated.

Multi-seed finding is possible with this technique, though there has been no need to implement it yet. Both the IVF and the AMVF reduce the seed pool as they find new vertices, which allows the GSF to seed low-multiplicity vertices, once the high-multiplicity ones have been removed from the seed pool. An example of the performance of the GSF is shown in Figure 3.3.

3.5.3   Peak width

One of the benefits of using an analytic function to find vertex seeds in place of a mode finder is that more information can be provided to the vertex finder. The Gaussian density functions allow an accurate estimate of the Gaussian width of the seed to be calculated. If it is assumed that the local shape of $W(z)$ at the global maximum z_{max} is Gaussian, then the width can be estimated as

$$\sigma(z) = \sqrt{-\frac{W(z_{\mathrm{max}})}{W''(z_{\mathrm{max}})}}.$$
(3.8)

As will be described in Section 3.7, this estimate is key to allowing multi-vertex fitting in high pile-up conditions by constraining the vertex fit to the approximate position of the seed.

3.6   Adaptive Multi-Vertex Finder (AMVF)

The AMVF is based on the same adaptive estimator as described in Section 3.4.1, however it is extended to allow tracks to have weight in the fit of multiple nearby vertices. With the deterministic annealing schedule, the weights generally evolve such that they are only left with a non-negligible weight to a single compatible vertex. During the fit, when tracks have weights to multiple vertices, the vertex candidates "compete" for tracks by adjusting their positions. The technique was originally developed pre-data-taking in ATLAS, however it has never been used in production. More recently, it was recommissioned for studies of the Phase-II Pixel Detector (ITk) upgrade, where it showed considerable promise [119].

3.6.1   Procedure

The procedure of vertex finding with the AMVF has an identical outline to that of the IVF (see Section 3.4.1), though there are significant differences in individual steps to accommodate multi-vertex finding. The AVMF proceeds as follows:

Seed Finding: Tracks not yet assigned to any vertex candidate are used as input to the GSF (see section 3.5) to estimate the most likely position of a primary vertex.

Track to seed assignment: After a seed is found, the set of nearby tracks to fit is chosen. This has a very different implementation in the AMVF compared to the IVF. Whereas the IVF only assigns tracks not yet compatible to a previously fitted vertex, the AMVF uses all tracks (that pass the quality selection), including tracks compatible to previously fitted vertices. This means that each track may, and probably will, be assigned to multiple vertices. The track assignment criteria and tuning are described in detail in Section 3.7.

Fitting: The linearised helical parameters of the assigned tracks are used in a weighted adaptive Kalman filter. The transverse and longitudinal positions are constrained in the fit by the beamspot and the seed width respectively. One key difference between the IVF and AMVF is that the AMVF performs a global fit where required. If a new vertex to be fitted shares a track with a previously fitted vertex, then all linked vertices are refitted simultaneously. The track weight ω_i to each vertex i is calculated with the multi-vertex implementation of the adaptive estimator:

$$\omega_i(\chi_i^2, T) = \frac{e^{-\frac{1}{2}\chi_i^2/T}}{\sum_j e^{-\frac{1}{2}\chi_j^2/T} + e^{-\frac{1}{2}\chi_0^2/T}} \tag{3.9}$$

where T is the annealing temperature and χ^2 is the standardised distance squared. The original set of temperatures used was defined as $T = [64.0, 16.0, 4.0, 2.0, 1.5, 1.0]$. The dependence of the weight on χ for three different temperatures is shown in Figure 3.2. The annealing takes place in six steps, with the weights recalculated each time. The total weight for a track across all assigned vertices is normalised to one. For the sole purposes of this normalisation, tracks are also given a notional three standard deviation (corresponding to $\chi_0^2 = 9$) compatibility with "unassigned".

Acceptance/rejection: The fitted vertex candidate is accepted if it satisfies three criteria:

1. It must have at least two compatible tracks from the pool of tracks not compatible to any previously fitted vertex. This ensures that a new vertex cannot be formed entirely from tracks compatible with some previous vertex. Track compatibility is defined as a χ_2^2 probability greater than 10^{-4}.

2. The position of the fitted vertex candidate must not be within 3σ of any previously fitted vertex, based on the calculated fit errors of both vertices.

3. The weighted average of track weights in the fit ($\sum \omega^2 / \sum \omega$) must be greater than two thirds.

If the fitted vertex candidate is accepted, all compatible tracks are removed from the pool of tracks not yet compatible to any vertex. If the vertex candidate is rejected, only the most compatible track is removed from the pool of tracks not yet compatible to any vertex. Thus, the pool of tracks to find vertex seeds is always reduced in each iteration, ensuring that vertex finding is guaranteed to terminate at some point.

The AMVF terminates when:

- fewer than two tracks remain in the pool of tracks not yet compatible to any vertex, or

- the seed finder is unable to return a seed, or

- the maximum number of allowed iterations is reached.
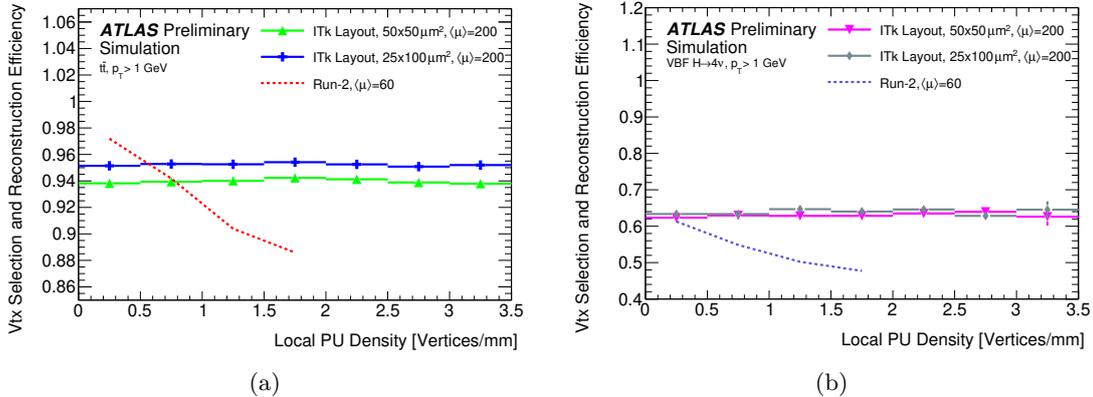
Figure 3.4: The primary vertex candidate reconstruction and selection efficiency for (a) $t\bar{t}$ and (b) VBF $H \to 4\nu$ as a function of local pile-up density in events with $\langle\mu\rangle = 200$. Results use analogue clustering and pixel sizes of $50 \times 50\mu\mathrm{m}^2$ or $25 \times 100\mu\mathrm{m}^2$ for the ITk layout. The primary vertex is selected as the vertex with the highest Σp_T^2 of associated tracks [10].

3.7 Multi-vertex fitting in high pile-up conditions

The AVMF was originally recommissioned for the Phase-II pixel tracker upgrade (ITk) [119]. With a simple tuning, it was found to offer better HS reconstruction compared to the IVF. Figure 3.4 shows an example of the expected HS selection efficiency using the AMVF for two different ITk geometries with the tight track assignment criteria [10] at $\langle\mu\rangle = 200$, compared to the Run-2 performance [114].

Preliminary studies showed that both the CPU performance and vertex reconstruction performance of the AMVF depended strongly on the track assignment criteria. Whilst one should expect loosening the track assignment criteria to degrade CPU performance, this also unexpectedly degraded vertex finding efficiency. However, the multi-vertex fitting aspect of the AMVF is intended to handle cases where tracks are assigned to multiple vertices. In fact, if not concerned about CPU time, a looser track assignment ought to improve performance by fitting globally where required. When the track assignment criteria are much tighter, the multi-vertex fitting of the AMVF is strongly suppressed, since it makes it unlikely that vertex candidates will share tracks. This is not desirable, since multi-vertex fitting has the potential to be more resistant to pile-up than single vertex fitting. This section describes the investigation, understanding and correction of this problem, along with optimisation of the track assignment and other parameters of the AMVF for data-taking around $\langle\mu\rangle = 60$.
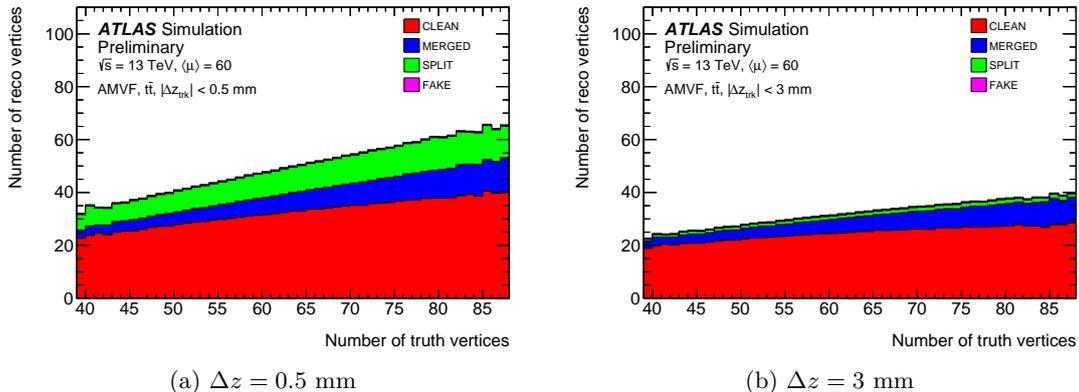
Figure 3.5: A comparison of the reconstructed vertex classification for the AMVF with difference track acceptance windows Δz, as a function of μ. It is seen that decreasing the track association window can increase the number of CLEAN and MERGED vertices, but also sees many more SPLIT vertices. The combined number of CLEAN and MERGED vertices is lower for the larger track association window on the right.

3.7.1 Vertex candidate migration

The original track assignment of the AMVF was very simple. Tracks would be assigned to a vertex seed if they were within a longitudinal window of $\Delta z = \pm 5$ mm. Since the AMVF fits vertices globally where required, a loose window therefore ought to improve reconstruction efficiency and accuracy, since vertex-dense regions will be fit as a whole rather than one-by-one. When the vertices are fit together, the optimal track assignment is decided by the fitter via the annealing procedure with the adaptive estimator (see equation (3.9)). If the multi-vertex fitting aspect is working as intended, the only drawback to a loose track assignment would be the added computational time. The more vertices linked through shared tracks, the more vertices get refitted. Surprisingly, large efficiency improvements were seen when tightening the track selection criteria to 0.5-1 mm. In effect, the multi-vertex finder was performing better when multi-vertex fitting was reduced. Valid vertex seeds, which ought to result in a reconstructed vertex, were somehow being lost at higher Δz. Figure 3.5 shows the AMVF performance in $t\bar{t}$ as a function of μ for two values of Δz. At lower values of Δz, more vertices are reconstructed, but many more interactions are split, with their tracks being divided between multiple reconstructed vertices.

In order to investigate how the AMVF behaved for vertex seeds with overlapping Δz, the initial and final fit positions (z_i and z_f, respectively) were compared to their nearest neighbour's position on either or both sides (z_+ and/or z_-). A simple ratio was used to quantify the extent to which a vertex candidate moved towards its nearest neighbour:

$$R_{\pm} = \frac{z_f - z_i}{z_{\pm} - z_i}.$$

60

Values of R near to 0 indicate that the position of the fitted vertex candidate is near to its seed position, whereas values near 1 indicate the position of the fitted vertex candidate has moved to one of its neighbours. As shown in Figure 3.6a, if a new vertex candidate lay within the track assignment window of an existing vertex, the multi-vertex fit would favour the higher track multiplicity vertex for both vertices, thus the lower multiplicity vertex would "migrate" into the other vertex. The frequency of this migration increased sharply as the Δz window was increased and the overlapping windows became more frequent. Such vertex candidate migration is not the intended behaviour of a multi-vertex finder.

The cause of this behaviour was revealed by careful examination of the fitter's logic. The annealing process begins by assigning nearly all tracks equal weights. For illustration, consider a true vertex A, with 40 tracks, and nearby vertex B, with 20 tracks. If the two vertices are separated by a distance less than the Δz window, then both fits will include all 60 tracks. Since the multi-vertex fit begins with all tracks having nearly equal weights to both vertices, both fits will prefer the location of the larger 40-track vertex. Furthermore, when using a large Δz window, this behaviour will occur even if none of the tracks from vertex A are compatible with vertex B and vice versa.

Several approaches were taken to mitigate this issue:

Significance-based track assignment: The fix-width acceptance window has the drawback of treating both well-measured and poorly-measured tracks the same. This approach can be refined by instead using the impact parameter significance. This is defined as:
$$s = \sqrt{(d_0/\sigma(d_0))^2 + (z_0/\sigma(z_0))^2}.$$
However, this approach alone does not stop poorly measured tracks from being assigned to large numbers of distant vertices. Therefore a loose absolute Δz cut is retained as a cut off.

Annealing schedule: Another way to reduce vertex candidate migration is to make the initial track weights more dependent on χ^2. This was studied by introducing a changeable "annealing temperature power" (ATP) parameter. Each of the six temperatures in the annealing schedule is raised to the ATP, meaning that ATP values less than one result in lower annealing temperatures, giving a greater sensitivity of weights to χ^2 in each stage except the last (in which $T = 1$). As an example, setting ATP $= 0.5$ would change the default temperatures from $T = [64.0, 16.0, 4.0, 2.0, 1.5, 1.0]$ to $T_{\text{new}} = [8.0, 4.0, 2.0, 1.41, 1.22, 1.0]$. Figure 3.2 shows the dependence of the weight on χ for $T = 64$, $T = 8$ and $T = 1$ as a comparison.

Longitudinal constraint: An (x, y) constraint from the beam spot has always been used in the vertex fit. A z constraint can be easily added to this, since the constraint is simply the initial Kalman filter state vector and its covariance. The GSF provides the required width, equation (3.8), of each seed. This therefore adjusts
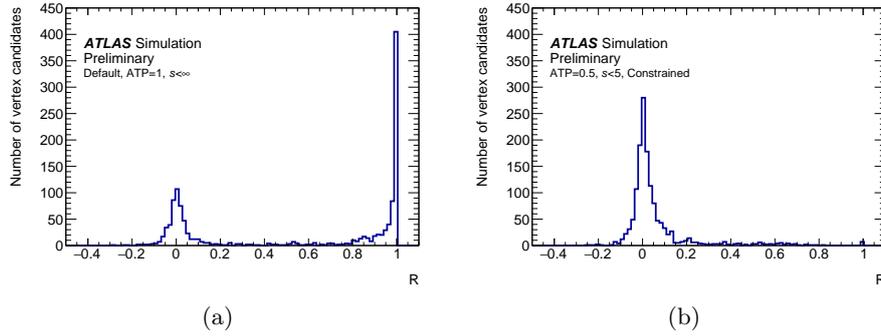
Figure 3.6: Vertex migration metric (a) before and (b) after introduction of a track assignment cut on impact parameter significance (s), stiffer annealing temperature power (ATP), and a longitudinal constraint. After fixing this issue, the vertex finder reconstructs 15% more vertices compared to the previous default, and does not suffer from multiple reconstructed vertices at the exact same location. Due to the "finding-through-fitting" approach, different numbers of vertex candidates will be found in the two scenarios.

> the strength of the constraint according to the quality of the tracks used to form the seeds. Seeds resulting from well-measured tracks will have a narrow constraint, whilst seeds resulting from poorly measured tracks will have a wider constraint.

All of these changes were successful at reducing the unwanted behaviour to some extent, but it was the longitudinal constraint that fully eliminated the migration issue. This is shown in Figure 3.6b, with the optimised significance cut ($s < 5$), stiffened annealing schedule (ATP=0.5) and longitudinal constraint.
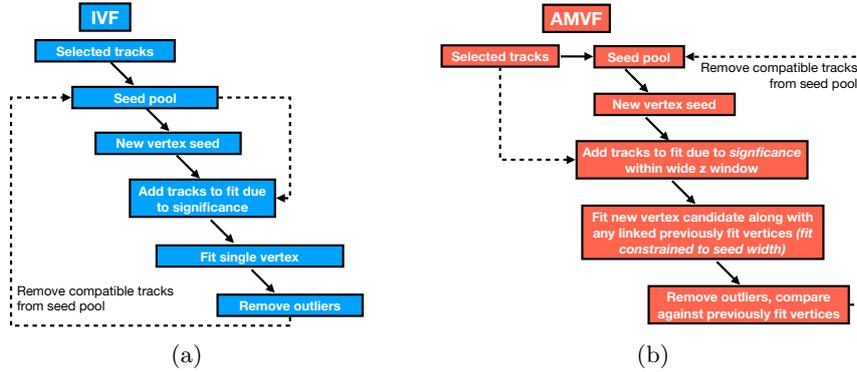
Figure 3.7: Comparison of the (a) IVF and (b) AMVF logic. The overall process is broadly similar between the two finders, however the individual steps differ in the AMVF to accommodate multi vertex fitting.

3.8 Performance

This section compares the vertex reconstruction quality, efficiency, spatial resolution and track association for the previous (IVF) and the new (AMVF) primary vertex reconstruction strategies, using the simulated $t\bar{t}$ and VBF $H \rightarrow 4\nu$ data described in Section 3.2. These studies use samples with $\langle\mu\rangle = 60$, but to make the results more generalised, a local pile-up density parameter is sometimes used. This is defined as the number of generated interactions within a symmetrical ± 2 mm longitudinal window of a given vertex (usually the hard-scatter). The vertex and event classifications as described in Section 3.2.2 are used to quantify the ability of the vertex finders to reconstruct and isolate individual pp interactions, and the HS interaction specifically.

3.8.1 Merge probability

Much of the performance improvements with the AMVF come from a reduced tendency to merge nearby vertices. The longitudinal separation for nearby reconstructed vertices is shown in Figure 3.8. At low values of Δz, individual vertices cannot be separated and are reconstructed as a single vertex. This is indicated by a steep decrease around Δz which appears as a trough in the distribution. A narrower trough is therefore indicative of improved performance. The study uses simulated $t\bar{t}$ events, however the generation of pile-up vertices is largely independent of the signal process, so this can be considered as general to all processes. The longitudinal separation highlights the superior ability of the AMVF to resolve vertices within a millimetre of each other. The depletion in the region around $\Delta z = 0$ mm for both vertex finders is due to merging. The AMVF being able to better resolve vertices separated by smaller distances results in fewer merged vertices. The IVF sees a small excess at $\Delta z = 0$ mm due to split vertices being reconstructed
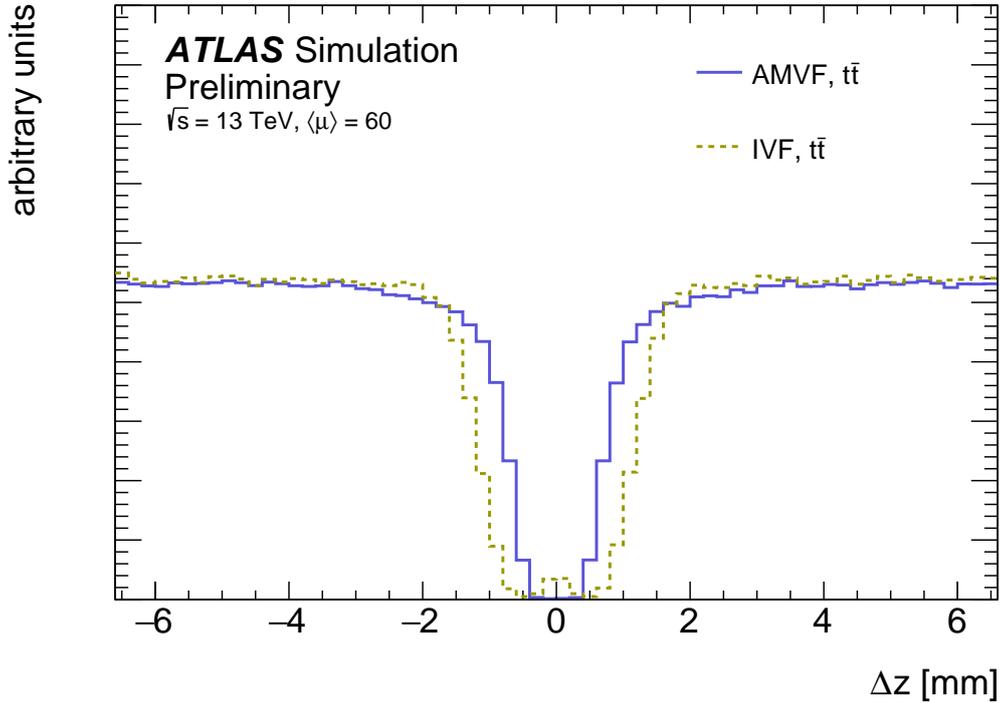
Figure 3.8: Distribution of the longitudinal separation between pairs of nearby reconstructed primary vertices in simulated $t\bar{t}$ events, for the IVF and AMVF. The AMVF is better able to resolve nearby vertices, as indicated by the smaller distance between reconstructed vertices. This can be seen as a narrower trough in the distribution.

at the same position as another vertex, which is generally prevented by the AMVF's acceptance/rejection criteria described in Section 3.6.1.

3.8.2 Hard-scatter vertex quality

Figure 3.9 compares the quality of HS vertex reconstruction and matching for the IVF and AMVF at $\langle\mu\rangle = 60$, showing the fraction of events in each classification. This is indicative of the amount of contamination in the reconstructed hard scatter vertex. Ideal performance would be indicated by 100% events reconstructed as CLEAN. The AMVF performs better, with a 5% higher fraction of events reconstructed as CLEAN/-MATCHED compared to the IVF. This corresponds to lower fractions of pile-up contaminated events.
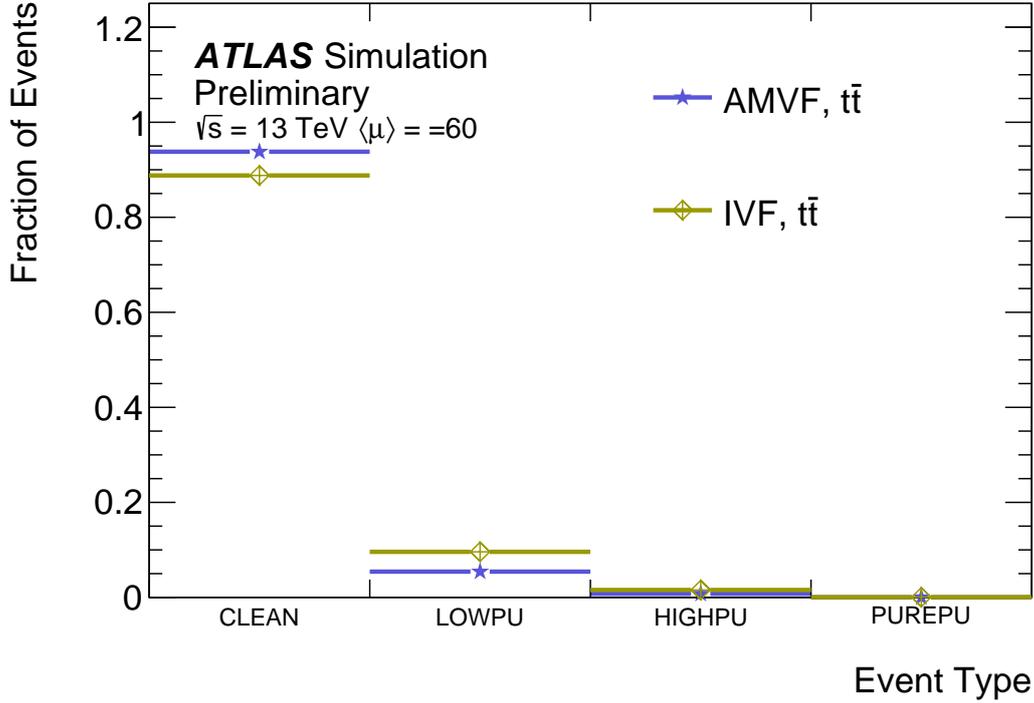
Figure 3.9: Fraction of events in each HS vertex reconstruction classification, for the IVF and AMVF. A higher fraction of CLEAN events are reconstructed with the AMVF, which indicates superior performance.

3.8.3 Hard-scatter efficiency

The HS reconstruction efficiency is defined as the fraction of events in which the HS vertex is both reconstructed and classified as either CLEAN/MATCHED, LOWPU or HIGHPU, which follows the method described in Ref. [114]. Figure 3.10 shows that for $t\bar{t}$ the AMVF is able to perform with a higher efficiency than the IVF, which already has a high efficiency. It also shows that for the more challenging VBF $H \rightarrow 4\nu$, the AMVF is able to recover about half the inefficiency seen with the IVF. The largest gains are seen in regions of highest vertex density, and the efficiency of the AMVF sees a reduced pile-up density dependency.

The HS selection efficiency is defined as the fraction of events in which the reconstructed vertex with the highest Σp_T^2 contains the largest total track weight from generated HS tracks. Figure 3.11 compares the selection efficiency as a function of local pile-up density for $t\bar{t}$ and VBF $H \rightarrow 4\nu$. The AMVF provides higher efficiency for both $t\bar{t}$ and VBF $H \rightarrow 4\nu$, eliminating pile-up density dependence.

(a) $t\bar{t}$ (b) VBF $H \to 4\nu$
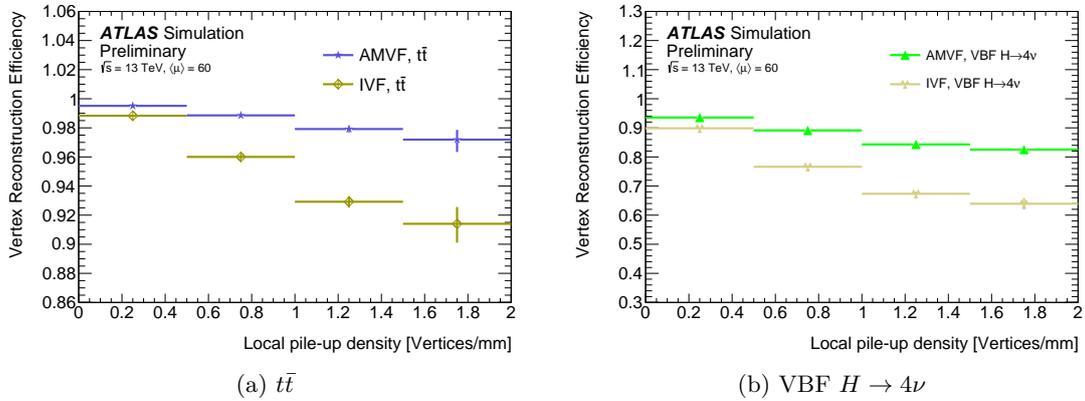
Figure 3.10: Comparison of IVF and AMVF HS vertex reconstruction efficiency as a function of local pile-up density, for (a) $t\bar{t}$ and (b) VBF $H \to 4\nu$. The two signal processes use different scales on the y-axis. The reconstruction efficiency is the fraction of events where the generated HS interaction is successfully reconstructed (classified as CLEAN/MATCHED, LOWPU or HIGHPU). The AMVF sees a higher reconstruction efficiency and reduced pile-up dependence in both cases. The efficiency gains are particularly prominent at high pile-up densities in for VBF $H \to 4\nu$, with around 50% of lost efficiency recovered.

(a) $t\bar{t}$ (b) VBF $H \to 4\nu$
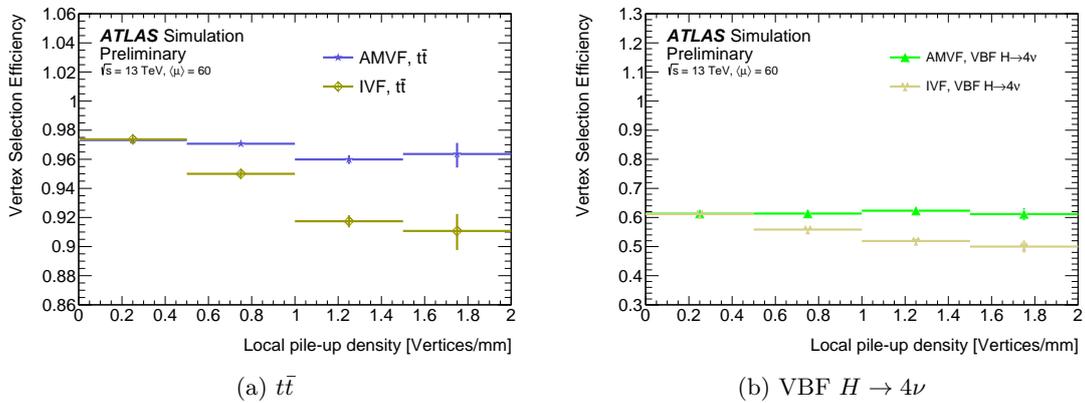
Figure 3.11: Comparison of IVF and AMVF HS vertex selection efficiency as a function of local pile-up density, for (a) $t\bar{t}$ and (b). The two signal processes use different scales on the y-axis. The selection efficiency is the fraction of events where the reconstructed vertex with highest Σp_{T}^2 contains the largest total track weight from the generated HS tracks. The AMVF significantly reduces the pile-up dependency.

3.8.4 Hard-scatter spatial resolution

Figures 3.12 and 3.13 show the radial (transverse) and longitudinal spatial resolutions for AMVF and IVF in $t\bar{t}$ and VBF $H \rightarrow 4\nu$. The resolution is calculated as the difference between the true and reconstructed vertex position in the MC simulation. The radial resolution is already well constrained by the beamspot, so little improvement is expected, though some is seen at the sub-micron level. In the z direction, the AMVF achieves 20% (10%) better resolution for $t\bar{t}$ (VBF $H \rightarrow 4\nu$).

3.8.5 Hard-scatter track efficiency and purity

Figure 3.14 shows the number of compatible ($\chi^2 \leqslant 9$) reconstructed tracks from the true HS interaction that have been correctly assigned to the identified HS vertex during the vertex fit, as a function of η. The AMVF and IVF have identical efficiency for $t\bar{t}$ in the central region ($|\eta| < 1.25$), while the AMVF correctly assigns 2% more HS tracks in VBF $H \rightarrow 4\nu$. In the forward region (large $|\eta|$), the IVF has a tendency to assign poorly-measured tracks to the first vertex to which they are compatible (usually the high-multiplicity HS). This is in contrast to the AMVF, which assigns them to the vertex they are *most* compatible with. As a result, the AMVF is less efficient in the forward region where tracking errors are larger. Figure 3.15 shows the amount of HS contamination from incorrectly associated pile-up tracks, as a function of $|\eta|$. For both $t\bar{t}$ and VBF $H \rightarrow 4\nu$, the AMVF reduces pile-up contamination by 25-70% over the entire range of $|\eta|$.

3.8.6 Pile-up vertex quality

Figure 3.16 shows a comparison of the average number of reconstructed vertices according to their quality classification, for the IVF and AMVF at $\langle \mu \rangle = 60$. The AMVF is able to reconstruct more CLEAN and MERGED vertices, but includes an average of one SPLIT vertex per event as well.

3.8.7 Pile-up vertex efficiency

Figure 3.17 shows a comparison of the average number of reconstructed vertices as a function of μ, for both the IVF and AMVF. The study uses simulated $t\bar{t}$ events, however since these are mostly pile-up vertices it can be considered as general to all processes. The upper dashed line shows the (unachievable) maximum limit of 100% efficiency, while the lower dashed line shows the theoretical maximum achievable efficiency, as calculated with the available reconstructed tracks after the quality selections. The number of vertices classed as CLEAN, MERGED, SPLIT and FAKE as reconstructed by the AMVF is also

shown. The AMVF is able to recover 35-50% of the reconstructable vertices that the IVF is unable to find at high μ.

(a) $t\bar{t}$ (b) VBF $H \to 4\nu$
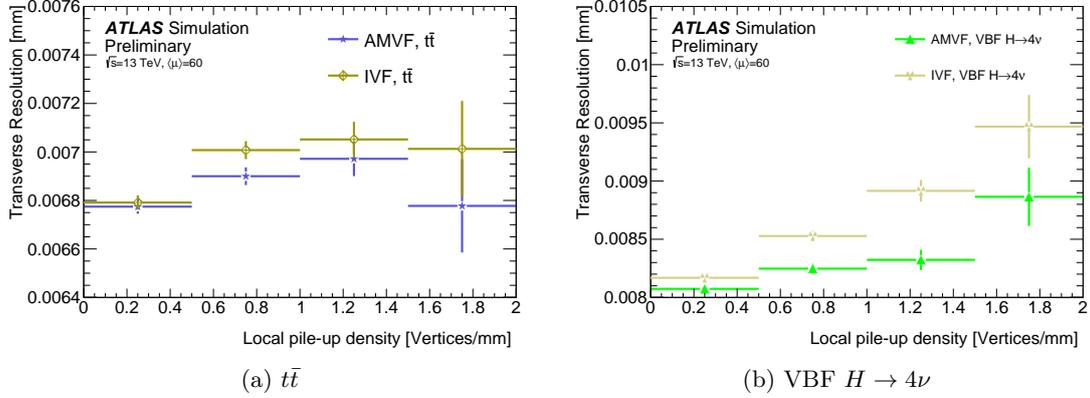
Figure 3.12: Comparison of IVF and AMVF HS vertex radial resolution as a function of local pile-up density for (a) $t\bar{t}$ and (b), obtained from the difference between the generator-level information and reconstructed primary vertex position in MC simulation. The scale is different for the two signal processes. The AMVF sees improved performance with reduced pile-up dependency compared to the IVF.

(a) $t\bar{t}$ (b) VBF $H \to 4\nu$
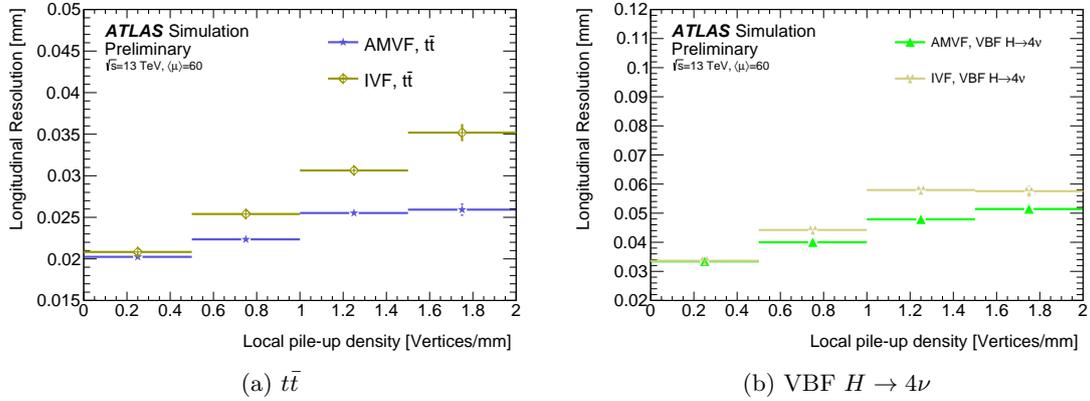
Figure 3.13: Comparison of IVF and AMVF HS vertex longitudinal resolution as a function of local pile-up density for (a) $t\bar{t}$ and (b) VBF $H \to 4\nu$, obtained from the difference between the generator-level information and reconstructed primary vertex position in MC simulation. The scale is different compared with Figure 3.12 due to the fact that the transverse resolution is narrowly constrained by the beamspot. The scale is also different for the two signal processes. The AMVF sees improved performance with reduced pile-up dependency compared to the IVF.

69

(a) $t\bar{t}$ (b) VBF $H \to 4\nu$
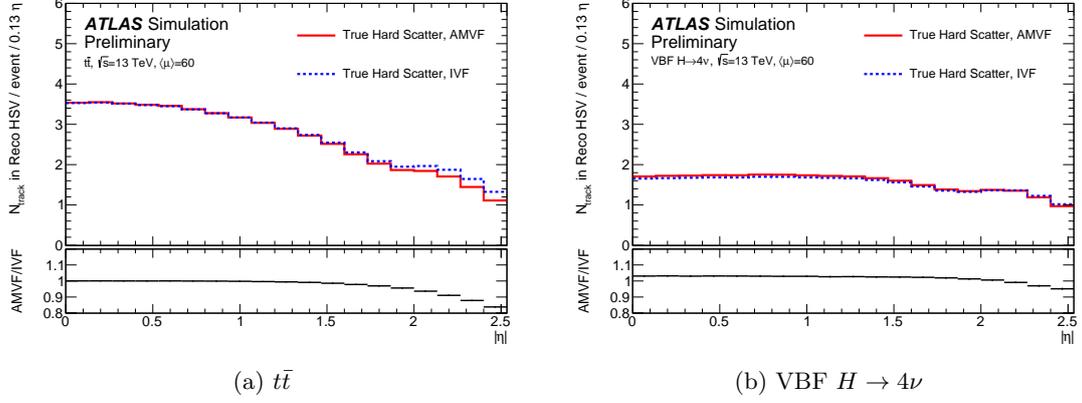
Figure 3.14: Comparison of IVF and AMVF HS track efficiency as a function of $|\eta|$ for (a) $t\bar{t}$ and (b) VBF $H \to 4\nu$ samples. Each plot shows the number of correctly associated compatible reconstructed tracks per $|\eta|$ bin originating from the generated HS vertex. The AMVF shows a higher efficiency by a few percent for VBF $H \to 4\nu$ across almost the entire $|\eta|$ range, and equal performance in $t\bar{t}$ up to around $|\eta| < 1.5$. The AMVF sees slightly lower efficiency in the forward region compared to the IVF.

(a) $t\bar{t}$ (b) VBF $H \to 4\nu$
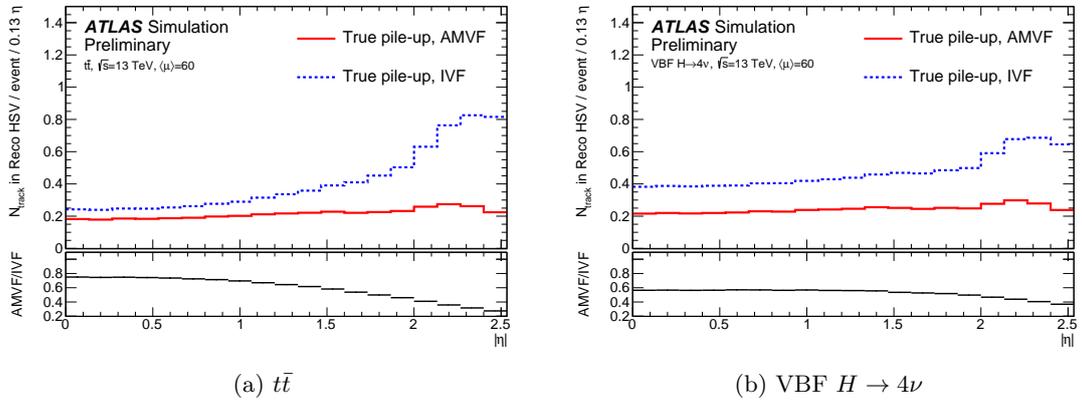
Figure 3.15: Comparison of IVF and AMVF HS track contamination as a function of $|\eta|$ for (a) $t\bar{t}$ and (b) VBF $H \to 4\nu$ samples. Each plot shows the number of compatible reconstructed tracks per $|\eta|$ bin originating from pile-up interactions that are incorrectly associated with the HS vertex. The AMVF sees reduced impurity across the entire $|\eta|$ range, especially in the forward regions.

70
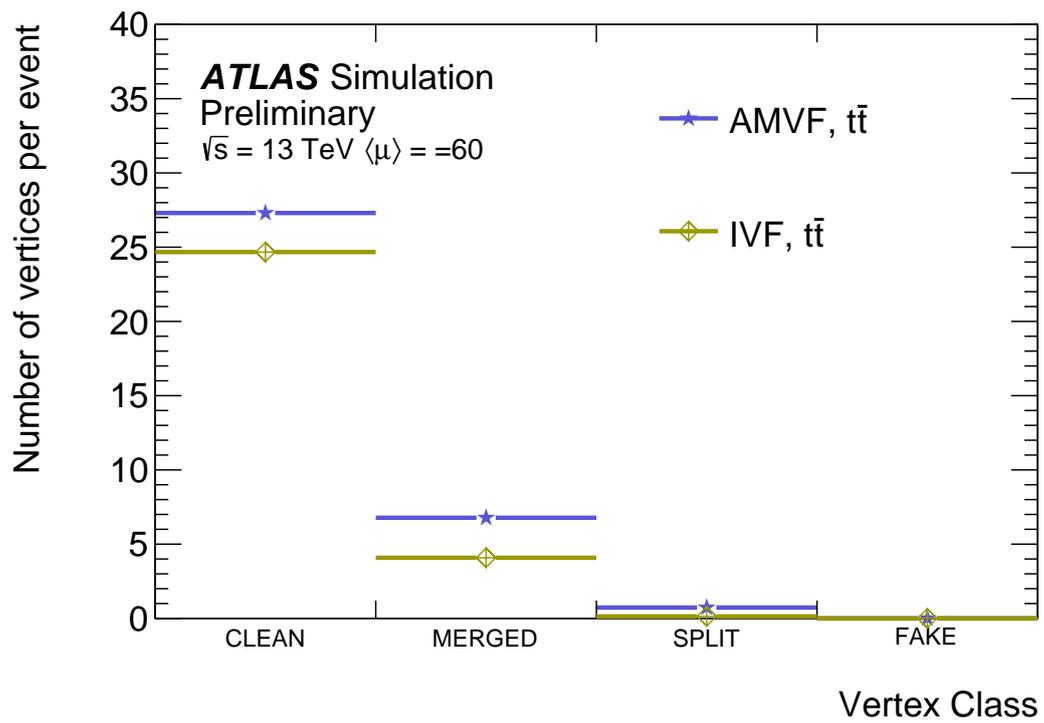
Figure 3.16: The number of reconstructed vertices with each quality classification, for the AMVF and AMVF at $\langle\mu\rangle = 60$. The AMVF demonstrates improved performance by reconstructing more CLEAN and MERGED vertices compared to the IVF. It also reconstructs around one more SPLIT vertex per event.
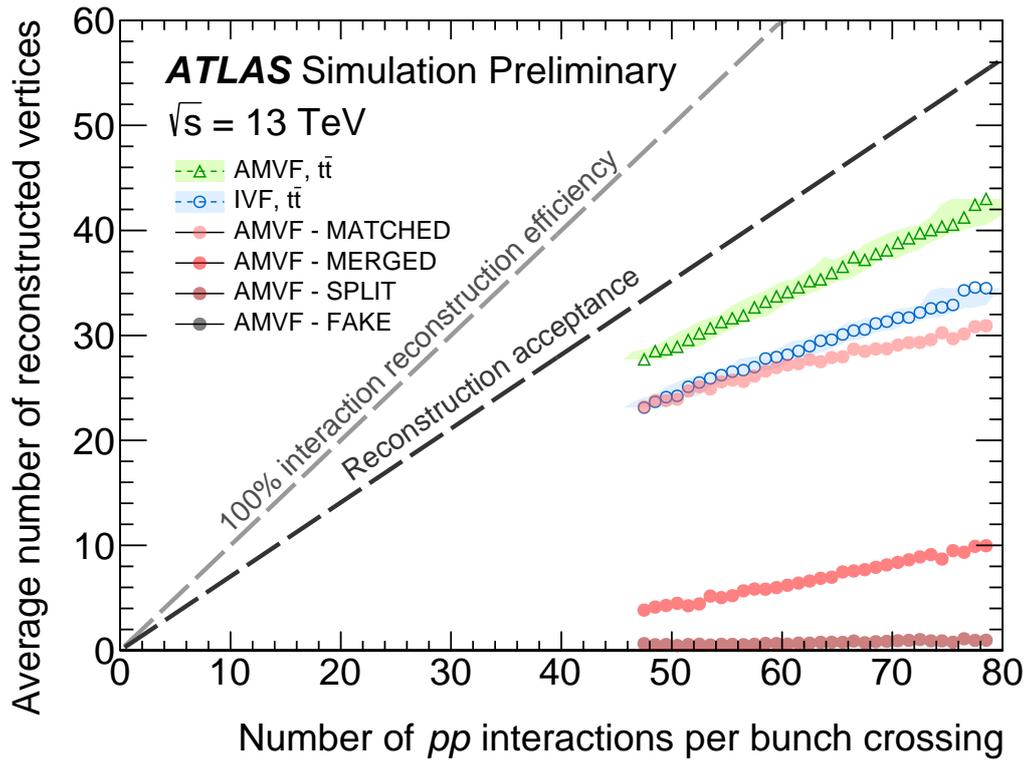
Figure 3.17: The average number of vertex reconstructed as a function of the number of *pp* interactions per bunch crossing (μ), in simulated $t\bar{t}$ events. For reference, the upper dashed line corresponds to perfect reconstruction efficiency, whilst the lower dashed line gives a more realistic idea of the maximum possible efficiency given the reconstructed tracks available to the vertex finder. Error bars on the data points are statistical uncertainties. The filled circles show the classification of AMVF vertices as CLEAN, MERGED, SPLIT and FAKE.

3.9 Summary of chapter 3

The Gaussian seed finder and the adaptive multi-vertex finder presented in this chapter are able to outperform the previous ATLAS vertex reconstruction strategy with respect to physics performance. This is especially true in high pile-up environments, where the improvement is required to maintain performance in ATLAS Run-3 and beyond. This includes the recovery of 35-40% of both the pile-up vertex and VBF $H \rightarrow 4\nu$ HS reconstruction efficiencies near $\langle \mu \rangle = 60$.

Future areas of study for this vertex reconstruction strategy include further ways to make use of the analytic seed-finding model, as well as diagnosing and eliminating remaining sources of inefficiency in the vertex finder. Computational efficiency did not yet form part of these studies, however this can be a focus in future. The project described in this section was defined by the qualification task. It did not include, for example, looking for potential biases, which has subsequently become a focus by the ATLAS vertexing group.

Test of Lepton Flavour Universality in top quark decays in ATLAS

4.1 Introduction

As described in chapter 1, a direct test of Lepton Flavour Universality (LFU) by comparing the decay rates of $W^{\pm}$ bosons to taus and muons is well motivated. This can be directly tested by measuring

$$R(\tau/\mu) = \mathrm{Br}(W \to \tau\nu_\tau)/\mathrm{Br}(W \to \mu\nu_\mu) \tag{4.1}$$

using the high number of $t\bar{t}$ pairs produced by the LHC in ATLAS. Assuming top quarks decay exclusively via charged current interactions as $t \to W^{\pm}b$, this leads to a very large sample of $W^{\pm}$ boson pairs to analyse. The leptonic decays of $t\bar{t}$ pairs can then be analysed using a *tag* and *probe* approach, using one lepton to identify the event (*tag*) and the other to perform the measurement (*probe*). The taus can be identified via decays to muons. The significant lifetime of the taus mean they travel some distance from the primary vertex and beamline, which typically results in a larger transverse impact parameter, $|d_0^\mu|$[1], and lower transverse momentum, p_T^μ[2], compared to muons produced via $W \to \mu\nu_\mu$. By using single lepton triggers on the *tag* lepton, it is possible to have a large, unbiased sample of *probe* muons that can have very low p_T^μ. This results in two final state channels, according to the *tag* and *probe* combinations: e-μ and μ-μ. Using these two channels, the measured value of $R(\tau/\mu)$ can be extracted using a 2-D profile likelihood fit in $|d_0^\mu|$ and p_T^μ comparing data to Monte Carlo. The main backgrounds for this analysis are $Z^0(\to \mu\mu) + (b\text{-})$jets in the μ-μ channel and, for both channels, non-prompt muons produced by decays other than $W \to \ell\nu_\ell$ ($\ell = \tau, \mu$) (typically b- and c-meson decays). Throughout this section, muons produced in the primary top decay are

---

[1]Defined as the closest approach of track to the beamline, introduced in Section 2.3.

[2]The magnitude of the momentum in the transverse plane, introduced in Section 2.3.

generally labelled as μ_{prompt}, muons produced via intermediate tau are labelled $\mu_{\tau(\to\mu)}$ and background muons produced in hadron decays are labelled μ_{had},

This chapter proceeds as follows: Section 4.2 gives details of the ATLAS Run-2 pp collision dataset and Section 4.3 describes the Monte Carlo simulated samples used in this analysis. Section 4.4 describes the event selection and MC calibration used in this analysis. Section 4.5 describes a study into optimising the event selection to reduce backgrounds. Sections 4.6 and 4.7 describe the data-driven background scale factor calculations. Section 4.8 describes the fit procedure and implementation of systematic uncertainties. Section 4.10 describes the results and finally Section 4.11 gives a summary of the conclusions. The author's main research and findings are detailed in Sections 4.5, 4.6 (excluding 4.6.6) and 4.7.

The results of this analysis have been accepted by Nature Physics as of 30/3/21 (see Ref. [3]). Figures that include the "ATLAS" label are taken from this publication. Figures without the ATLAS label in Sections 4.5, 4.6 (excluding 4.6.6) and 4.7 (and their associated appendices) are the author's own work. Figures without the "ATLAS" label in other sections are produced by the author's analysis colleagues, and included here for clarity and illustration. A summary of the analysis has also been published in proceedings from the Beauty 2020 conference [4].

4.2 Data sample

The analysis is performed using the data sample of pp collisions at $\sqrt{s} = 13$ TeV, collected by ATLAS between 2015-2018 (Run-2).

The analysis is performed on a derivation[3] of the large ATLAS dataset, specifically for top quark studies. The particular derivation is for standard top analyses with decay modes that include at least one lepton. These contain at least one electron or muon with $p_{\mathrm{T}} > 20$ GeV for data collected in 2015 and with $p_{\mathrm{T}} > 25$ GeV for data collected in 2016, 2017, 2018. They contain all other objects and variables required for analysing top quark decays. This analysis applies further selection criteria as described in Section 4.4.

4.3 Monte Carlo simulated samples

4.3.1 Overview of MC samples used

MC generated samples were used to develop the analysis, to compare to data and to evaluate the signal and background efficiencies. The full list of MC samples used can be found in Table 4.1, with more ATLAS-specific details provided in Appendix A. Both

---

[3]An overview of derivations is given in Section 2.8.1.

Process	Generator	Order of σ calculation	Showering	Tune
		Signal processes		
$t\bar{t}$	Powheg	NNLO+NNLL	Pythia8	A14
Wt	Powheg	(5FS) NLO+NNLL	Pythia8	A14
$t\bar{t} + V$	MadGraph5_aMC@NLO	NLO	Pythia8	A14
		Background processes		
single top s & t	Powheg	(5FS) NLO+NNLL	Pythia8	A14
Z+jets	Sherpa	NNLO	Sherpa	Sherpa
W+jets	Sherpa	NNLO	Sherpa	Sherpa
VV	Sherpa	NNLO	Sherpa	Sherpa
		$t\bar{t}$ samples for estimating modelling uncertainties (fast-sim)		
$t\bar{t}$	Powheg	NNLO+NNLL	Pythia8	A14
$t\bar{t}$	MadGraph5_aMC@NLO	NLO	Pythia8	A14
$t\bar{t}$	Powheg	NNLO+NNLL	Herwig	H7UE
$t\bar{t}$	Powheg	NNLO+NNLL	Pythia8-h_{damp}	A14

Table 4.1: The different Monte Carlo samples used in the analysis, including the generator used, the order of the cross section calculation, the showering programme and tune.

$t\bar{t}$ and Wt, the main signal samples, are processed using the full ATLAS simulation programme described in Section 2.8. Some of the samples used for systematic uncertainty studies do not use the full simulation framework, and are instead processed using the reduced but faster `AtlFast2` simulation programme. All events had additional minimum bias events generated with Pythia8 to simulate pile-up. The MC generated samples are reconstructed using the same algorithms as the data. They are reweighted with several scale factors to reproduce what is measured in data. The amount of pile-up in MC events is reweighted to produce the average number of interactions per bunch crossing in 2015+2016, and the actual number of interactions per bunch crossing in 2017 and 2018, as observed in data. The samples were reweighted with scale factors of order unity to reproduce the electron and muon reconstruction efficiencies, as well as the b-tagging efficiencies measured in data.

The data and simulation samples are analysed using the `AnalysisTop-21.2.99` software package. A bespoke software package named `ttau` was used to analyse the ntuples produced by `AnalysisTop`.

4.3.2 Signal processes

This analysis uses $t\bar{t}$ and Wt dileptonic final states, with decays $t \rightarrow Wq$, $W \rightarrow \ell\nu$ treated as signal. The processes include on-shell $W^{\pm}$ bosons that decay leptonically – either to τ or light leptons – such that they directly contribute to $R(\tau/\mu)$. The event selection produces an extremely pure sample of $t\bar{t}$ events, with some single top Wt events.

76

Normally in dileptonic $t\bar{t}$ studies, Wt would be treated as a background. However, since the parameter of interest is sensitive to $W \rightarrow \ell\nu$ decays, it is treated as a signal in this analysis. Additionally, there are a small number of events from $t\bar{t}$ events produced with either a $W^{\pm}$, Z^0 or $\ell^+\ell^-$ produced through Z^0/γ interference, which is referred to as $t\bar{t} + V$.

t$\bar{\text{t}}$:

The MC simulated $t\bar{t}$ events use the PowhegBox [97, 109, 120, 121] v2 generator. This provides matrix elements at next-to-leading order (NLO) with the NNPDF3.0NLO [122] parton distribution function. The Powheg h_{damp} factor (a model parameter that controls the matching between matrix element and the parton shower and effectively regulates the high p_{T} radiation) was set to $1.5m_{\text{top}}$ [123]. The default scale $\sqrt{m_{\text{top}}^2 + p_{\text{T}}^2}$ is used for the functional form of the renormalisation and factorisation scale. MadSpin [124, 125] is used to simulate top quark decays at LO whilst preserving spin correlations. For the parton shower and hadronisation, events are interfaced with Pythia8.230 [103] using the A14 set of tuned parameters [126] and the NNPDF2.3LO [122] PDF set. The EvtGenv1.6.0 programme [106] is used to simulate the decays of bottom and charm hadrons.

As described in Section 1.5.1, a higher order cross section prediction is available. This is used to correct the NLO $t\bar{t}$ inclusive cross section. It is calculated at next-to-next-to-leading-order (NNLO) accuracy in the strong coupling constant α_S, including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [58–63]. For pp collisions at $\sqrt{s} = 13$ TeV and assuming a fixed top mass of 172.5 GeV, the NNLO+NNLL prediction is 832 ± 51 pb as calculated using the Top++2.0 programme [64]. This cross-section prediction has uncertainties due to the PDF and α_{s}. These are calculated using the PDF4LHC prescription [127] with the MSTW2008 68% CL NNLLO [128, 129], CT10 NNLO [130, 131] and NNPDF2.3 5f FFN [132] PDF sets. These uncertainties are added in quadrature to the scale uncertainty.

In each event, either the top or anti-top particle p_{T} is reweighted to match the NNLO in QCD and NLO in α_{EW} differential prediction [133]. This is done to improve the data/MC agreement.

In order to assess the uncertainty due to the choice of generator, several additional $t\bar{t}$ samples are generated. These are detailed in Appendix A.

The dilepton sample is used in all cases with the exception of non-prompt background studies, where the non-all-hadronic sample is used. This ensures no overlap between samples and maximal exploitation of the available MC statistics.

Single top Wt:

The simulated single-top Wt events use the PowhegBox [97, 109, 120, 121] v2 generator. This provides matrix elements at NLO in α_s in the five-flavour scheme with the NNPDF3.0NLO [122] PDF set. The default scale $\sqrt{m_{\text{top}}^2 + p_{\text{T}}^2}$ is used for the functional form of the renormalisation and factorisation scale. When generating Wt at NLO there is some overlap with $t\bar{t}$. The diagram removal scheme [134] is used to handle this interference [123], removing all NLO diagrams that overlap with double resonant $t\bar{t}$ from the calculation of the Wt amplitude. MadSpin [124, 125] is used to simulate top quark decays at LO whilst preserving spin correlations. For the parton shower and hadronisation, events are interfaced with Pythia8.230 [103] using the A14 set of tuned parameters [126] and the NNPDF2.3LO [122] PDF set. The EvtGenv1.6.0 programme [106] is used to simulate the decays of bottom and charm hadrons.

A higher order cross section prediction is available and it is used to correct the NLO inclusive cross-section. It is calculated at NLO in QCD with NNLL soft gluon corrections [135, 136]. For pp collisions at $\sqrt{s} = 13$ TeV and assuming a fixed top mass of 172.5 GeV, the NLO+NNLL prediction is 71.7 ± 3.8 pb. This prediction has uncertainty due to the PDF which is calculated using the MSTW2008 90% CL [128, 129] NNLO PDF set. It is added in quadrature to the scale uncertainty.

In order to assess the uncertainty due to the choice of generator, several additional Wt samples are generated. These are detailed in Appendix A.

$t\bar{t}$+V:

The simulated $t\bar{t} + V$ events use the MadGraph5_aMC@NLOv2.3.3 [98] generator. This provides matrix elements at NLO with the NNPDF3.0NLO [122] PDF set. For the parton shower and hadronisation, events are interfaced with Pythia8.230 [103] using the A14 set of tuned parameters [126] and the NNPDF2.3LO [122] PDF set.

4.3.3 Background processes

There are two main groups of background processes in this analysis. The first consists of high p_{T} processes that produce two prompt leptons, while the second consists of non-prompt events where the *probe* muon comes from the decays of *b*-, *c*- or light hadrons. The first is a background for $W^{\pm} \rightarrow \mu\nu$ while the second is a background for $W^{\pm} \rightarrow \tau\nu \rightarrow \mu\nu\nu$.

t$\bar{\text{t}}$ and single top Wt:

Whilst these are both signal processes, they are also the main contribution to events where only one of the leptons comes from $W^{\pm} \to \ell\nu$ decay. The other non-prompt muon is then a background for $W^{\pm} \to \tau\nu \to \ell\nu\nu$. The simulation of these events is described above. For the background studies, the non-all-hadronic $t\bar{t}$ samples and inclusive Wt samples are used.

Single top s and t:

The simulated s and t channel single top events are generated with the PowhegBox [97, 109, 120, 121] v2 generator, using the same settings as described for single top Wt.

Z+jets:

The simulated QCD Z+jets events use the Sherpa v2.2 [9] generator. This provides matrix elements at NNLO with the NNPDF3.0NNLO [122] PDF set and uses a dedicated set of tuned parton-shower parameters developed by the Sherpa authors for this version. Additional hard parton emissions [137] are matched to a parton shower based on Catani-Seymour dipoles [100].

These simulated events use the ME+PS matching [138] for different jet multiplicities which are then merged into an inclusive sample using an improved CKKW matching procedure [99, 139]. This is extended to NLO accuracy using the MEPS@NLO prescription [101]. The particular samples used in this analysis are NLO accurate for up to two additional partons and LO accurate for up to four additional parton emissions. The sample uses a virtual QCD correction for matrix elements at NLO which is provided by the OpenLoops library [140, 141].

W+jets:

The simulated W+jets events use the Sherpa v2.2.1 generator, using the same settings as described for $Z+$ jets.

Fully leptonic dibosons:

The simulated fully leptonic diboson (with additional jets) events use the Sherpa v2.2 [9] generator. This provides matrix elements at NNLO with the NNPDF3.0NNLO [122] PDF set and uses a dedicated set of tuned parton-shower parameters developed by the Sherpa authors for this version. Additional hard parton emissions [137] are matched to a parton shower based on Catani-Seymour dipoles [100].

These simulated events use the ME+PS matching [138] for different jet multiplicities which are then merged into an inclusive sample using an improved CKKW matching procedure [99, 139]. This is extended to NLO accuracy using the MEPS@NLO prescription [101]. The particular samples used in this analysis are NLO accurate for up to one additional parton and LO accurate for up to three additional parton emissions. The sample uses a virtual QCD correction for matrix elements at NLO which is provided by the OPENLOOPS library [140, 141]. The correction is calculated in the G_μ scheme, which ensures optimal description of pure EW interactions at the EW scale.

Semileptonic dibosons:

The simulated semileptonic diboson events use the SHERPA v2.2 [9] generator, with the same configuration as for fully leptonic dibosons.

4.3.4   Monte Carlo weighting

Monte Carlo events are reweighted in various variables to match observations in data. These weights are of the order of unity. Each MC event has an overall weight applied. This is calculated from the following:

- Weight produced by the MC generators.

- The pile-up reweighting. The amount of pile-up in MC events is reweighted to produce the average number of interactions per bunch crossing in 2015+2016, and the actual number of interactions per bunch crossing in 2017 and 2018, as observed in data.

- (*Tag* and *probe*) lepton scale factors (reconstruction, identification, isolation, track-to-vertex-association). The individual scale factors are combined into one scale factor for this analysis.

- Single (*tag*) lepton trigger scale factor. This is the ratio of measured and simulated trigger efficiencies. It used to correct simulated samples to match the efficiencies observed in data.

- *B*-tagging scale factor. This is derived from comparisons between simulation and data. It is used to correct the *b*-tagging response and related uncertainties in simulation.

- Scale factor correcting the distribution of the Jet-Vertex-Tagger (JVT) variable [142]. This variable is used to identify and reject jets from pile-up interactions.

4.4 Event selection, object definition and calibration

This section documents the object definitions, event selection criteria and calibration used in this analysis. First this section details the different object definitions used in this analysis. Next the selection criteria using these objects is described for both of the e-μ and μ-μ channels. Finally, the calibration of the parameters to improve data/MC agreement is described.

4.4.1 Object definitions

Jets

Hadronic jets are reconstructed from energy deposits in the calorimeter cells (topoclusters) [143] at the electromagnetic energy scale using the anti-k_t algorithm [144]. This is done using a radius parameter of 0.4. The jets are calibrated to the energy scale of identical jets reconstructed from stable simulated (so-called "truth") particles, excluding muons and neutrinos, using the same algorithm [145]. A jet vertex tagger [142] is applied to suppress pile-up for jets with $25 < p_T^{\text{jet}} < 60$ GeV and $|\eta| < 2.4$. Reconstructed jets with $p_T^{\text{jet}} > 25$ GeV and $|\eta| < 2.5$ are used in this analysis.

Top quarks are expected to almost always decay to bottom quarks and $W^{\pm}$ bosons. The bottom quarks are identified by tagging reconstructed hadronic jets as containing b-hadrons. This is done using the MV2c10 b-tagging algorithm [146, 147] is used with a 70% efficiency working point.

Muons

Muons are reconstructed using a combination of data from the Muon Spectrometer and Inner Detector. Hits in muon chambers are combined with Inner Detector track segments, which are fitted together to determine muon trajectories [148]. Three categories of muons are defined for this analysis: *baseline*, *tag* and *probe*.

Baseline muons are required to pass the 'medium' quality criteria, be strictly isolated from surrounding activity and satisfy $p_T^{\mu} > 5$ GeV. The strict isolation is achieved through two requirements. First, within a cone of size $\Delta R = 0.3$[4] around the muon, the sum of the p_T of other tracks must be below 6% of the muon p_T^{μ}. Second, within a cone of size $\Delta R = 0.2$ around the muon, the sum of p_T calculated from calorimeter energy deposits must be below 6% of the muon p_T^{μ}. This isolation level is provided by the Muon Combined Performance group, and is known as `FCTight` (FixedCutTight) [86]. *Tag* and *probe* muons must meet the *baseline* requirements with additional requirements. *Baseline* muons are used for 3rd lepton vetoes.

---

[4]Where $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the angular separation, as described in Section 2.3.

Tag muons must satisfy the *baseline* requirements. They must be trigger matched. Their additional kinematic requirements ensure they are above trigger thresholds. The p_T^μ is required to either be greater than 27.3 GeV, or $p_T^\mu > 1.05 \cdot \texttt{threshold [GeV]}$, where `threshold` is the Level-1 trigger threshold, whichever is larger. The trigger details are provided below. The pseudorapidity is required to be $|\eta| < 2.5$. *Tag* muons must have $|z_0 \sin\theta| < 0.3$ mm, where θ is the polar angle of the track and z_0 is the longitudinal distance to the primary vertex, and $|d_0^\mu| < 0.5$ mm. These requirements reduce background from pile-up, cosmic muons and poorly reconstructed muons.

Probe muons must satisfy the *baseline* requirements. Unlike *tag* muons, *probe* do not need to be trigger matched. Therefore an unbiased sample over a wide p_T^μ spectrum can be obtained by applying loose kinematic requirements. These are $p_T^\mu > 5$ GeV and $|\eta| < 2.5$. *Probe* muons must have $|z_0 \sin\theta| < 0.3$ mm, where θ is the polar angle of the track and z_0 is the longitudinal distance to the primary vertex, and $|d_0^\mu| < 0.5$ mm. These requirements reduce background from pile-up, cosmic muons and poorly reconstructed muons. In order to reduce background from muons produced in the decays-in-flight of pions and kaons, the momentum balance significance[5] (MBS) must be less than two.

Electrons

Electrons are reconstructed from clusters in the ECal that are matched with tracks from the Inner Detector and refitted using a Gaussian Sum Filter. Two categories of electron are defined for this analysis: *baseline* and *tag*.

Baseline electrons are required to pass the 'tight' log-likelihood (TightLH) criteria, be strictly isolated from surrounding activity and satisfy $p_T^e > 5$ GeV. The isolation criteria are designed to give an efficiency of 90% at $p_T^e = 25$ GeV and 99% at $p_T^e = 60$ GeV, uniform in η, using the so-called Gradient working point [85]. *Baseline* electrons are used for 3rd lepton vetoes.

Tag electrons must satisfy the *baseline* requirements. They must be trigger matched. Their additional kinematic requirements ensure they are above trigger thresholds. The p_T^e is required to be greater than 27 GeV, or $p_T^e + \texttt{threshold [GeV]}$ where `threshold` is the Level-1 trigger threshold, whichever is larger. The trigger details are provided below. The pseudorapidity is required to be $|\eta| < 2.47$, with electrons in the crack region $1.37 < |\eta| < 1.52$ excluded. *Tag* electrons must have $|z_0 \sin\theta| < 0.3$ mm, where θ is the polar angle of the track and z_0 is the longitudinal distance to the primary vertex, and $|d_0^e| < 0.5$ mm. These requirements reduce pile-up and poorly reconstructed electrons.

---

[5]Momentum balance significance (MBS) is the difference in measured momentum of a muon between the ID and MS standalone measurements, with respect to the uncertainty (σ) on energy lost in the system.

Overlap removal

An overlap procedure is employed in order to avoid cases where the detector response to a single physical object is reconstructed as two separate-final state objects. This stops lepton signals in the calorimeter from being reconstructed as hadronic jets. The procedure is as described in Ref. [149].

E_T^{miss}

This physics object is not used in this analysis. Using this parameter as a selection criterion would bias the selection of tops decaying to taus, therefore biasing the parameter of interest $R(\tau/\mu)$.

Electron and muon triggers

Single high p_T electron or muons are used to trigger the events. The *tag* electron or muon in each event must be matched to the trigger lepton.

For 2016-2018, tag muons use the either the HLT_mu26_ivarmedium or HLT_mu50 high-level triggers. Both of these are seeded by the 20 GeV level-1 trigger (L1_MU20). The HLT_mu26_ivarmedium single-muon trigger selects isolated muons with a threshold of $p_T > 26$ GeV. The HLT_mu50 single-muon trigger has no isolation requirements and selects muons with a threshold of $p_T > 50$ GeV. Full details are provided in [150]. For 2015, tag muons use the HLT_mu20_iloose or HLT_mu50 high-level triggers. HLT_mu20_iloose is seeded by the 15 GeV level-1 trigger (L1MU15) and selects single loosely isolated muons with a threshold of $p_T > 20$ GeV. HLT_mu50 is seeded by the 20 GeV level-1 trigger (L1_MU20) and selects single muons with a threshold of $p_T > 50$ GeV. Full details of the 2015 triggers are provided in Ref. [8].

For 2016-2018, tag electrons use either the HLT_e26_lhtight_nod0_ivarloose, HLT_e60_lhmedium_nod0 or HLT_e140_lhloose_nod0 high-level triggers. All of these are seeded by the 22 GeV single electron level-1 trigger (L1_EM22VHI), which has isolation requirements for $p_T < 50$ GeV. The HLT_e26_lhtight_nod0_ivarloose single-electron trigger selects isolated electrons with a threshold of $p_T > 26$ GeV. The HLT_e60_lhmedium_nod0 and HLT_e140_lhloose_nod0 single-electron triggers select non-isolated electrons with p_T thresholds of 60 GeV and 140 GeV respectively. Full details are provided in Ref. [151]. For 2015, tag electrons use the HLT_e24_lhmedium, HLT_e60_lhmedium, or HLT_e120_lhloose high-level triggers. These triggers are seeded by the 20 GeV level-1 trigger (L1_EM20VH) and select electrons without isolation requirements, at p_T thresholds of 24 GeV, 60 GeV and 120 GeV respectively. Full details of the 2015 triggers are provided in Ref. [8].

4.4.2 Event selection

This section details the selection criteria for this analysis. Different selections are applied to define a signal region, used to extract $R(\tau/\mu)$, calibration regions, used to determine calibration (described in Section 4.4.4), and control regions, used to determine background scale factors (described later in Sections 4.6 and 4.7).

The principal goal of the signal region criteria are to obtain a pure sample of dileptonic $t\bar{t}$ events. The key requirements are exactly two leptons of opposite charge, and at least two b-tagged jets. One lepton must satisfy *tag* and the other *probe* criteria. The b-tagged jets must have $p_{\mathrm{T}}^{\mathrm{jet}} > 25$ GeV.

This analysis selects events with either one electron and one muon (e-μ channel), or two muons (μ-μ channel). Events must be triggered by the *tag* electron in the e-μ channel or the *tag* muon in the μ-μ channel. This ensures that there is no trigger bias on the *probe* muons used to determine the parameter of interest. In the μ-μ channel, both muons can satisfy *tag* and *probe* criteria. In this case, both muons are used as *probe* muons. This eliminates any bias in the p_{T}^{μ} distribution of the *probe* muons.

In order to reduce background from Drell-Yan $Z^0(\to \mu\mu) + (b\text{-})$jets, events with dimuon mass in the range $85 < M(\mu\mu) < 95$ GeV are excluded in the μ-μ channel. Finally, low mass hadronic resonances are excluded in both channels by requiring the dilepton mass $M(\ell\ell) < 15$ GeV.

This selection results in a high-purity $t\bar{t}$ (and Wt) sample. There are small backgrounds where the *probe* muon originates either from non-prompt hadron decays and prompt (non-top) decays. Applying the event selection to the MC samples described in Section 4.3 shows these backgrounds constitute less than 5% of events in the e-μ sample and less than 15% of events in the μ-μ sample. The larger background component of the μ-μ sample is due to Drell-Yan dimuon production.

4.4.3 Definition of $|d_0^{\mu}|$

d_0 definition

The transverse impact parameter of the *probe* muons is one of the two essential parameters of this analysis. This is measured as the closest approach of the track to the beamline in the transverse plane, which is standard in ATLAS. Defining it relative to the beamline as opposed to the primary vertex ensures that the resolution of $|d_0|$ is independent of the transverse resolution of the vertex, which is dependent on the physics process (this can be seen in Section 3.8.4, Figure 3.12). This in turn allows the $|d_0^{\mu}|$ calibration methods in Section 4.4.4.

Data taking period:	α
2015 data	$+0.00422(\pm 0.000014)$ mm
2016 data A, B	$+0.00343(\pm 0.000011)$ mm
2016 data C	$-0.00335(\pm 0.0000003)$ mm

Table 4.2: Charge-independent shifts applied to d_0 in different data taking periods.

$|d_0^\mu|$ alignment bias

The Inner Detector tracking group is responsible for ensuring that the detector is fully aligned in data [152]. Corrections to alignment can be made by reprocessing the data. Sometimes, even after reprocessing there are residual imperfections in alignment. Such imperfections may result in charge-dependent or charge-independent biases in the d_0 distribution. The average d_0 as a function of Run Number is shown in Figure 4.1. This figure shows that there are significant biases in 2015 and 2016 data that require correction. The correction is applied simply as a charge-independent shift in the value of d_0:

$$d_0' = d_0 + \alpha \tag{4.2}$$

where α is determined as the mean value of a Gaussian fit of the d_0 distribution. The values of α for the different data-taking periods are shown in Table 4.2. Corrections are only applied to 2015 and 2016 data, since correcting the sub-micron level biases will have little-to-no impact when the resolution is $\mathcal{O}(10 \ \mu\mathrm{m})$. Checks were performed to ensure the corrections can be applied independently of p_T, η and ϕ. Improved data/MC agreement is seen in the d_0 resolution after applying the correction. The correction is applied as a first step in all the following studies and throughout this analysis.

The bias in 2015-2016 was introduced by a change in the underlying geometry description of the ATLAS inner detector and a mis-configuration of the beamspot constraint. These biases are to be corrected in future data reprocessing campaigns [152].
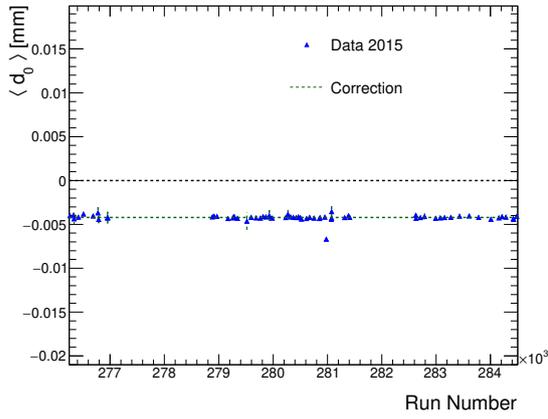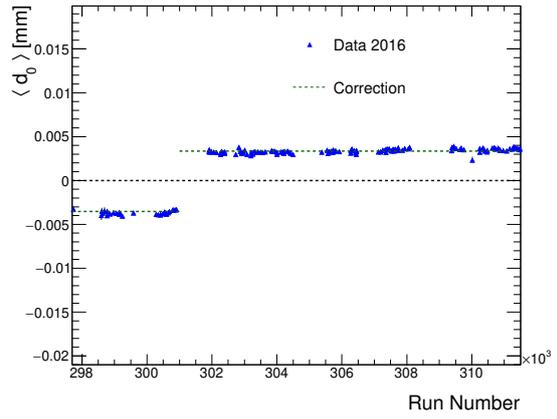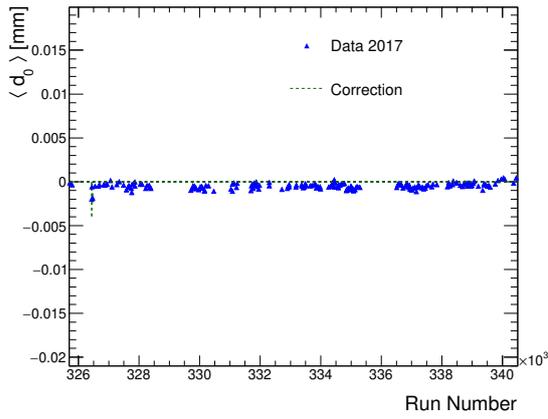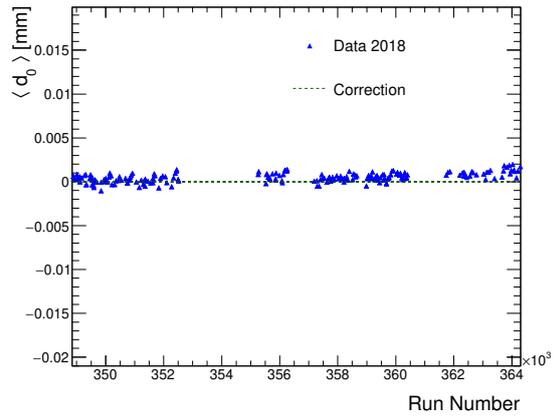
Figure 4.1: The average d_0 of *probe* muons in the $Z \rightarrow \mu\mu$ selection, as a function of RunNumber, for (a) 2015, (b) 2016, (d) 2017, (d) 2018 data. Significant biases are seen in 2015 and 2016 data that require correction.

4.4.4 $|d_0^\mu|$ calibration

The fit to extract $R(\tau/\mu)$ is sensitive to differences between MC and data in the *probe* muon $|d_0^\mu|$ distributions. The expected $|d_0^\mu|$ distributions of prompt muons in MC are therefore calibrated using data, to remove any disagreement. This calibration is performed using a $Z^0 \to \mu\mu$ calibration region. In both $t \to bW^\pm(\to \mu\nu)$ and $Z^0 \to \mu\mu$, the muons come from the primary interaction point, therefore their $|d_0^\mu|$ distributions are expected to be identical to the first approximation. The η and p_T^μ dependent $|d_0^\mu|$ distributions are extracted in this region and applied as templates to the μ_{prompt} component of the $t\bar{t}$ signal region.

The $Z^0 \to \mu\mu$ calibration region is defined by modifying the nominal event selection in two ways. First, hadronic jet requirements are dropped and, second, the invariant mass is *required* to be in the Z^0 peak region $85 < M(\mu\mu) < 100$ GeV. Removing the hadronic jets requirement maximises statistics whilst keeping the $|d_0^\mu|$ distribution consistent with $t\bar{t}$. The selection results in a $Z^0 \to \mu\mu$ sample of approximately 94 million prompt muons with 99.9% purity (as estimated in simulation). Thus the systematic uncertainties from the modelling of background processes are negligible.

The $|d_0^\mu|$ distributions are dependent on the kinematic variables η and p_T^μ. In order to account for this dependence, the $|d_0^\mu|$ distributions are determined separately in several kinematic bins of the η and p_T^μ of the *probe* muon in data. This is done in 11 p_T^μ bins and 3 η bins for a total of 33 bins. The bin boundaries in η are (0, 0.8, 1.5, 2.5) and in p_T^μ are (5, 10, 15, 20, 25, 30, 40, 50, 65, 100, 250, >250) GeV. The calibration region event selection yields a sample with sufficient statistics in all bins. In each kinematic bin:

- The expected contributions of muons with significant lifetimes (as estimated in simulation) are subtracted. These processes include $t \to bW^\pm(\to \tau(\mu\nu\nu)\nu)$, $Z^0(\to \tau\tau) + (b\text{-})$jets and muons from hadron decays.

- The resulting distribution is normalised to unity.

- The normalised $|d_0^\mu|$ distributions are scaled to the expected number of prompt events in the signal region.

The $|d_0^\mu|$ *templates* are then obtained by summing over all bins. This can be expressed simply as:

$$F^{\mathrm{pr}}(|d_0^\mu|) = \sum_{i=0}^{10} \sum_{j=0}^{2} r_{ij}^{\mathrm{pr}} F_{ij}^{\mathrm{pr}}(|d_0^\mu|),$$

where ij denotes a given kinematic bin, $F_{ij}^{\mathrm{pr}}(|d_0^\mu|)$ is the normalised $|d_0^\mu|$ distribution in the calibration region in each bin, r_{ij}^{pr} is the fraction of prompt events in the signal region in each bin, and $F^{\mathrm{pr}}(|d_0^\mu|)$ is the resulting template. Templates are derived separately across years of data taking to account for differences in the beam conditions and Inner Detector alignment.

(a) Before applying $|d_0^\mu|$ templates (b) After applying $|d_0^\mu|$ templates
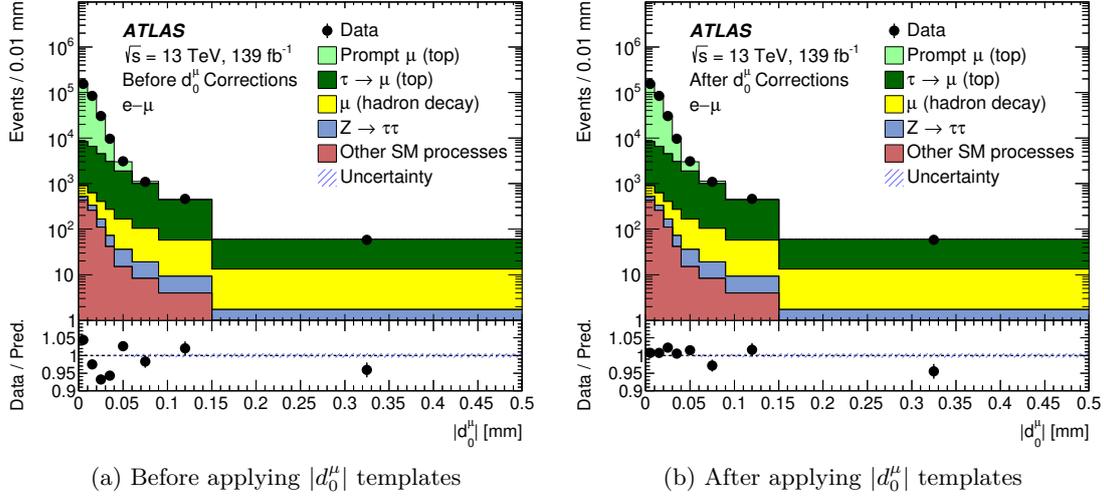
Figure 4.2: Data/MC agreement in the signal region, inclusive in p_T^μ, (a) before and (b) after applying the estimated $|d_0^\mu|$ templates for prompt muons, in the e-μ channel. By applying the the templates, the agreement between observation and prediction is seen to improve. The bias corrections described in Section 4.4.3 are applied in both (a) and (b) as a first step, such that the effect of this particular correction is not compared in this figure.

The $|d_0^\mu|$ templates are applied to the μ_{prompt} component in the fit used to extract $R(\tau/\mu)$. The impact of the corrections are shown in Figures 4.2 and 4.3. An uncertainty is applied due to the fact the templates are derived in a $Z^0 \to \mu\mu$ calibration region, but applied to a $t\bar{t}$ final state, which have very different track environments. This is described in Section 4.9.1.

$|d_0^\mu|$ resolution

The Inner Detector $|d_0^\mu|$ resolution can be determined from the $|d_0^\mu|$ distribution of the *probe* muons. The fit to extract $R(\tau/\mu)$ is affected by differences in the $|d_0^\mu|$ resolution between MC and data, in processes with significant $|d_0^\mu|$. Corrections are applied to the $|d_0^\mu|$ distributions of muons coming from τ and hadron decays to correct for the differences in $|d_0^\mu|$ resolution.

The corrections are determined using the same $Z^0 \to \mu\mu$ calibration region described above, as well as the same kinematic bins. The core of the resolution can be approximated by a Gaussian curve. This is done by fitting a Gaussian curve in each kinematic bin in the range $0 < |d_0^\mu| < 0.02$ mm. In the tail region $|d_0^\mu| > 0.02$ mm, the distribution is less well described by a Gaussian, so it is excluded from the fit. An example of the distribution in one kinematic bin is shown in Figure 4.4. As can be seen in the figure, extrapolating the Gaussian over the full range of $|d_0^\mu|$ describes almost all tracks. Across all kinematic bins, the difference is around 2% less using the Gaussian description

(a) Before $|d_0^\mu|$ corrections (b) After $|d_0^\mu|$ corrections
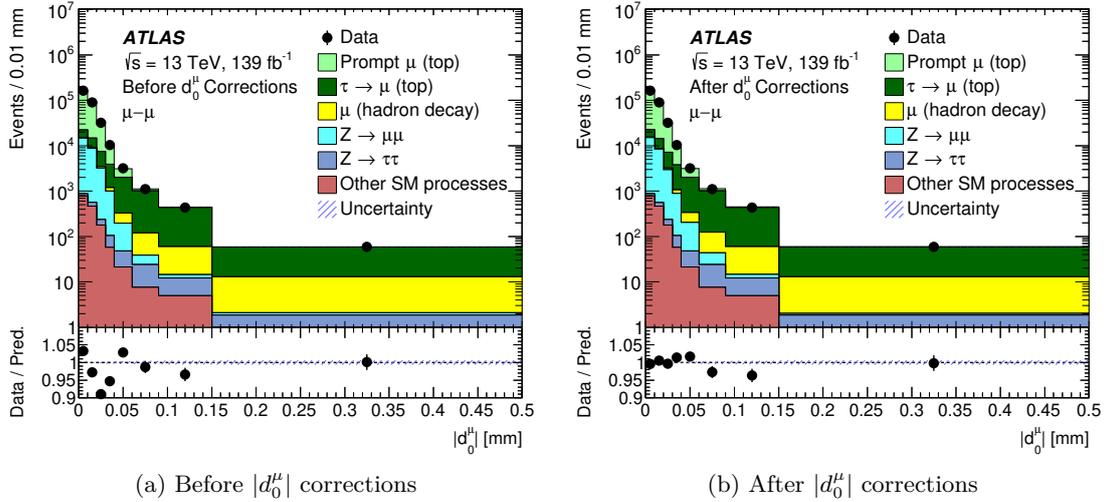
Figure 4.3: Data/MC agreement in the signal region, inclusive in p_T^μ, (a) before and (b) after applying $|d_0^\mu|$ corrections to prompt muons, in the μ-μ channel. By applying the the templates, the agreement between observation and prediction is seen to improve. The bias corrections described in Section 4.4.3 are applied in both (a) and (b) as a first step, such that the effect of this particular correction is not compared in this figure.

compared to observed data. Furthermore, the calculated corrections are of the order 1%. Therefore the Gaussian approximation is sufficient here. The calculations are described below.

In data, the $|d_0^\mu|$ resolution is not constant across years. It is wider in 2015+2016 and narrower in 2018. This is due to the narrowing of the beamspot in the transverse plane. This is not seen in MC, where all years show the same resolution. This results in better resolution in data than in MC in many of the kinematic bins. The difference in resolution must be accounted for in the templates used in the fit to extract $R(\tau/\mu)$.

The templates are determined in each kinematic bin as follows:

- The value of the shift is calculated as follows:

  - In each MC event, the $|d_0^\mu|$ of the *probe* muon is randomly smeared by a Gaussian with standard deviation defined as

  $$\sigma_{ij}^{\mathrm{sm}} = \sqrt{|\sigma_{ij}^2(\mathrm{data}) - \sigma_{ij}^2(\mathrm{MC})|}.$$

  - The difference between the resulting smeared (denoted by superscript "sm") and non-smeared (denoted by superscript "ns") $|d_0^\mu|$ distributions is computed as

  $$\delta_{ij} = f_{ij}^{\mathrm{sm}} - f_{ij}^{\mathrm{ns}}.$$

89
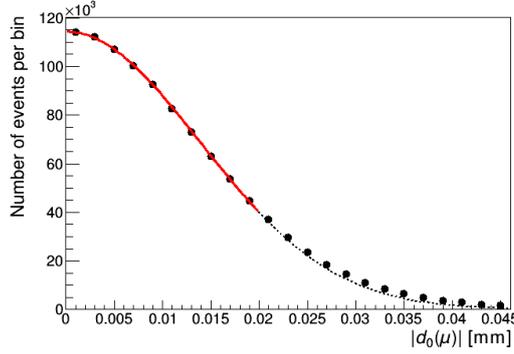
Figure 4.4: The $|d_0^\mu|$ distribution of *probe* muons in the kinematic bin with $25 < p_T^\mu < 30$ GeV and $|\eta| < 0.8$ in the 2017 sample. The black dots show the data points, whilst the curve shows the Gaussian fit with mean fixed at zero in range $|d_0^\mu| \leqslant 0.02$. The dashed curve is the extrapolation of the fitted curve beyond the fit range. The Gaussian approximation describes the distribution very well in the region $|d_0^\mu| \leqslant 0.02$. Beyond this range a difference of few percent is observed.

- – In order to avoid random fluctuations, the procedure is repeated 10,000 times and the average value of δ_{ij} is used in the following steps.

- If the resolution is wider in data than MC such that $\sigma_i^2(\text{data}) > \sigma_i^2(\text{MC})$, the correction is added to the non-smeared distribution, which is just equal to the average smeared distribution:

$$F_{ij}^\tau = f_{ij}^{\text{ns}} + \delta_{ij} = f_{ij}^{\text{sm}}.$$

- If the resolution is narrower in data MC such that $\sigma_i^2(\text{data}) < \sigma_i^2(\text{MC})$, the correction is subtracted from the non-smeared distribution:

$$F_{ij}^\tau = f_{ij}^{\text{ns}} - \delta_{ij}.$$

- By applying the correction in this way, it is assumed that the variations in the $|d_0^\mu|$ distribution are symmetric with respect to the change of sign of $\sigma_{ij}(\text{data}) - \sigma_{ij}(\text{MC})$.

- The F_{ij}^τ distributions are normalised to unity.

- The resulting $|d_0^\mu|$ distributions are scaled to the expected number of events in that kinematic bin.

Once the corrections have been applied in each kinematic bin, the *templates* are obtained by summing over all bins, such that

$$F^\tau(|d_0^\mu|) = \sum_{i=0}^{11} \sum_{j=0}^{2} r_{ij}^\tau F_{ij}^\tau(|d_0^\mu|),$$

90

where r^{τ}_{ij} is the fraction of $\mu_{\tau(\to\mu)}$ events in the kinematic bin ij of the signal region. The same method is also applied to smear the resolution of the μ_{had} background. The templates are applied to the $\mu_{\tau(\to\mu)}$ and μ_{had} components in the fit to extract $R(\tau/\mu)$.

This method produces relatively small corrections however they improve the data/MC agreement. A cross check is performed in a control region with a $Z^0 \to \tau\tau$ selection applied. Good agreement data/MC agreement is seen with all the $|d^{\mu}_0|$ templates applied, across all values of p^{μ}_{T} and $|d^{\mu}_0|$.

An uncertainty on the templates is applied to account for the differences between MC and data. This is simply half the correction, symmetrised. This is described in more detail in Section 4.9.1.

4.5   Selection optimisation

Whilst the signal $t\bar{t}$ ($+$ single top Wt) sample is quite pure, there is still a significant background contribution from hadron decays. This background requires scaling (described in the next chapter), and the associated scale factors contribute to the overall uncertainty on $R(\tau/\mu)$. Therefore, studies were performed with the goal of reducing the backgrounds but having little effect on signal, in order to improve the precision of $R(\tau/\mu)$. This was done by searching many discrimination variables for suitable cuts. These studies considered only the $\mu_{\tau(\to\mu)}$ contribution as signal, since the μ_{prompt} contribution is not statistically limited, relatively speaking. In addition, the fit to extract $R(\tau/\mu)$ is configured to have regions with different $\mu_{\tau(\to\mu)}$ and μ_{prompt} contributions. Therefore, where μ_{prompt} contributions occur in the same phase space as $\mu_{\tau(\to\mu)}$ it can also be considered a background. These studies are described in this chapter.

4.5.1   Discrimination variables considered

Several different variables are used to look at the activity near the *probe* leptons to discriminate between signal and background muons. These are:

- $\Delta R(\ell, \text{jet})$: The distance between the 4-vectors of the *probe* muon and nearest jet of any flavour.

- $\Delta R(\ell, b\text{-jet})$: The distance between the 4-vectors of the *probe* muon and nearest b-jet

- $z_0 \sin\theta$: The distance of closest approach in the r-z plane.

- η: pseudorapidity, as described in Section 2.3. Defined as

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right).$$
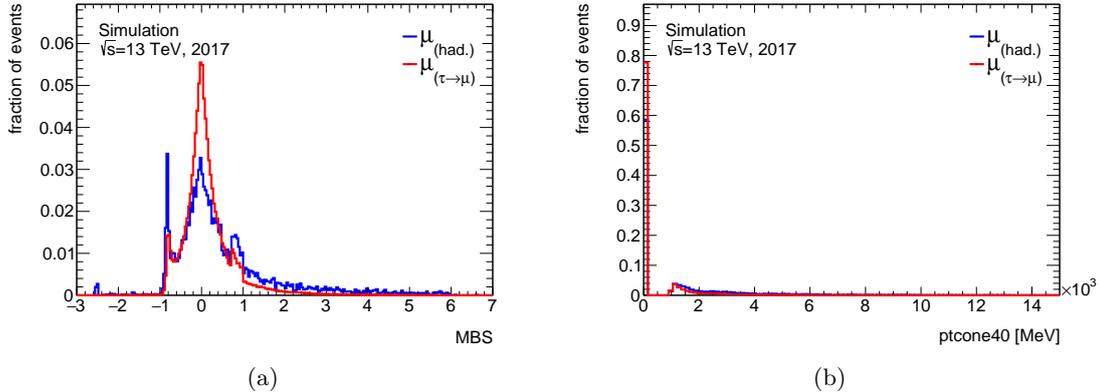
Figure 4.5: Distributions of (a) MBS and (b) *ptcone*40 for μ_{had} and $\mu_{\tau(\to\mu)}$, normalised to unity. The differences in these distributions can be used to better discriminate signal from background.

- *Momentum balance significance* (MBS): The significance of the difference in momentum between the ID and MS standalone measurements, with respect to the uncertainty on energy lost in the calorimeter system. This is defined as $(p_{\mathrm{ID}} - p_{\mathrm{MS}} - E_{\mathrm{loss}})/\sigma(E_{\mathrm{loss}})$, where p_{ID} and p_{MS} are the momentum as measured in the Inner Detector and Muon Spectrometer respectively, E_{loss} is the energy lost in the calorimeter system and $\sigma(E_{\mathrm{los}})$ is the uncertainty on E_{loss}.

The following variables are provided in three cone sizes: $\Delta R = 0.2$, 0.3 and 0.4 around the object of interest:

- *etcone*: the transverse energy calculated from the calorimeter cells in the given cone.

- *ptcone*: the sum of transverse momenta of the tracks in the given cone.

- *ptvarcone*: uses cone size of $\min(10[\mathrm{GeV}]/p_{\mathrm{T}}[\mathrm{GeV}], \Delta R)$.

- *topoetcone*: the sum of transverse energy of the topoclusters in the given cone.

4.5.2 Method

This section describes the method used to search for improved cuts. First, the distributions of the discriminating variables ($D(\mathrm{var})$) were plotted for the μ_{had} background and $\mu_{\tau(\to\mu)}$ signal, as shown in Figure 4.5. In order to search for improved cuts, the reduction in yield as a function of nominal yield was computed. For distributions that are asymmetrical but centred around 0, the reduction in yield as a fraction of nominal yield is estimated for "greater than" cuts using
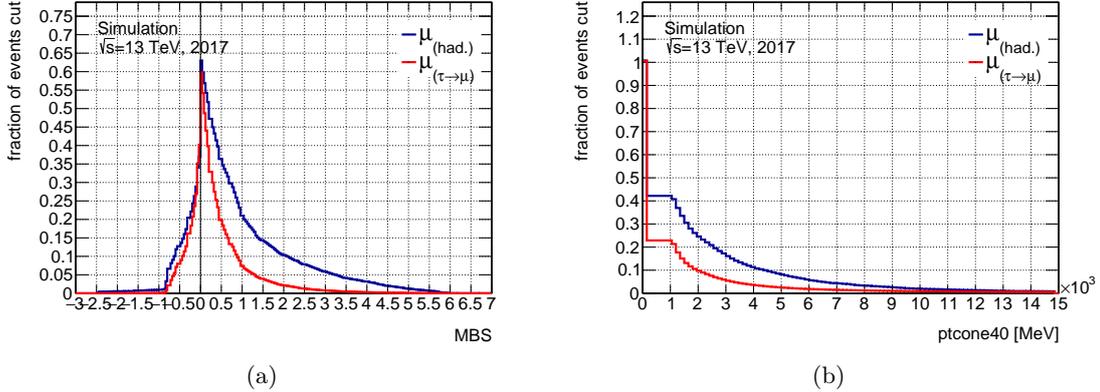
Figure 4.6: Distributions of the estimated signal loss and background reductions in (a) MBS and (b) *ptcone*40 for μ_{had} and $\mu_{\tau(\to\mu)}$. This suggests additional cuts will have very different effects on the background and signal yields. For example, MBS< 2 could reduce the background by $\sim$10% whilst only having a $\sim$2% impact on signal.

$$\frac{\int_{\min}^{\max} D(\text{var}) - \int_{\text{cut}}^{\max} D(\text{var})}{\int_{\min}^{\max} D(\text{var})}, \text{ var} < 0,$$

whilst the reduction in yield as a fraction of nominal yield is estimated for "less than" cuts using simply

$$\frac{\int_{\text{cut}}^{\max} D(\text{var})}{\int_{\min}^{\max} D(\text{var})}, \text{ var} > 0.$$

The study was performed using MC simulated data, using only the non-all-hadronic $t\bar{t}$ sample (see Section 4.3), which is the dominant source of muons from hadron decays. The study was performed using only the 2017 MC simulated data, which is representative of the whole dataset.

The loss of signal and reduction of background were calculated at 100 evenly-spaced cut values using this method. The comparisons for all variables considered are shown in Section 4.5.4. The two most promising variables were found to be MBS and *ptcone*40, which corresponds to *ptcone* with $\Delta R = 0.4$. The other track isolation variables showed similarly shaped separation as *ptcone*40 but they were not as powerful. The $z_0 \sin\theta$ variable was able to further reduce pile-up contributions, however these contributions are extremely small to begin with, meaning it is not worthwhile tightening the cut. The ΔR variables showed some discrimination due to the fact that muons coming from hadron decays are nearer jets. However, their discrimination is not as powerful as the track isolation variables, so they were not tested in the profile-likelihood fit to extract $R(\tau/\mu)$.

After promising cuts had been identified from these plots, they were applied sequentially. The nominal event selection was applied with the new requirements of MBS < 2,

93

Origin	Default	MBS < 2	MBS < 2, $ptcone40 < 3000$ MeV
μ_{had}	1183.86	1062.88 (-10.2%)	890.824 (-24.7%)
μ_{prompt}	159709	158748 (-0.60%)	146849 (-8.05%)
$\mu_{\tau(\to\mu)}$	18927.3	18539 (-2.05%)	17525.3 (-7.4%)
Prompt other	29.583	29.282 (-10.2%)	26.8156 (-9.35%)
Charge mis-ID	2.04261	2.04261 (-0%)	1.92547 (-5.73%)

Table 4.3: Reduction in background and signal after applying new selection cuts. The cut on MBS is seen to reduce the μ_{had} background without significant impact on signal.

and MBS $< 2, ptcone40 < 3$ GeV. The reduction in background and signal with these additional requirements (compared to the nominal selection) is shown in Table 4.3. Applying these additional selection criteria was shown to have a large impact on the number of muons from hadron decays without significant impact on the $\mu_{\tau(\to\mu)}$ contribution. Additional cuts that could reduce the number of muons from hadron decays, without affecting signal, include a very loose cut on η (e.g. $|\eta| < 2.2$) and $etcone20$ (e.g. $etcone20 < 10$ GeV).

4.5.3 Summary

A cut requiring MBS< 2 was shown to produce a small improvement on the overall precision of $R(\tau/\mu)$, and specifically reduced the number of background muons coming from pion and kaon decays, by 40% and 50% respectively. This is due to the fact that when a pion or kaon decays outside the Inner Detector, it is possible that the Inner Detector track is matched to a lower-momentum Muon Spectrometer track produced by the resulting muon, resulting in a larger MBS. As such, this cut is retained for the event selection for the analysis. Whilst $ptcone40$ and $etcone20$ were actively considered, cutting on these parameters would require recalibration of the muon scale factors. Therefore it was decided not to use these to modify the event selection at this time.

4.5.4 Distribution and effect of cut comparisons

This section shows all plots from the study. As described above, the plots come in two flavours. The first shows the difference between the shape of the distributions for muons from hadron decays background and $\mu_{\tau(\to\mu)}$ signal (here labelled $t \to \tau \to \mu$), whilst the second shows the reduction in background and loss in signal for 100 evenly-spaced cut values. Due to the tight isolation requirements already applied to the *probe* muons, several of the distributions of discrimination variables show little difference between $\mu_{\tau(\to\mu)}$ and μ_{had}. The ranges of the plots were selected to be consistent between variables

of the same type (e.g. same ranges for all *etcone* distributions), and to have the same scales for both the distributions and the effect of cuts plots. The y-axis scales are chosen to prevent any clipping, and to avoid switching back and forth between logarithmic scales.

The figures are organised according to the discrimination variables. Figures 4.13 and 4.16 are also shown in the above section for illustration, and included here again for completeness. The discrimination variables shown are:

- $\Delta R(\ell, \text{jet})$: The distance between the 4-vectors of the *probe* lepton and nearest jet of any flavour. Shown in Figure 4.7.

- $\Delta R(\ell, b\text{-jet})$: The distance between the 4-vectors of the *probe* lepton and nearest b-jet. Shown in Figure 4.8.

- $z_0 \sin \theta$: The distance of closest approach in the r-z plane. Shown in Figure 4.21.

- η: pseudorapidity, as described in Section 2.3. Defined as

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right).$$

  Shown in Figure 4.9.

- *Momentum balance significance* (MBS): The significance of the difference in momentum between the ID and MS standalone measurements, with respect to the uncertainty on energy lost in the calorimeter system. Shown in Figure 4.13.

The following variables are provided in three cone sizes: $\Delta R = 0.2$, 0.3 and 0.4 around the object of interest:

- *etcone*: the transverse energy calculated from the calorimeter cells in the given cone. Shown in Figures 4.10, 4.11 and 4.12 for each of the cone sizes respectively.

- *ptcone*: the sum of transverse momenta of the tracks in the given cone. Shown in Figures 4.14, 4.15 and 4.16 for each of the cone sizes respectively.

- *ptvarcone*: uses cone size of $\min(10[\text{GeV}]/p_\text{T}[\text{GeV}], \Delta R)$. Shown in Figures 4.17 and 4.18 for cone sizes of $\Delta R = 0.2$ and 0.4 respectively.

- *topoetcone*: the sum of transverse energy of the topoclusters in the given cone. Shown in Figures 4.19 and 4.20 for cone sizes of $\Delta R = 0.3$ and 0.4 respectively.
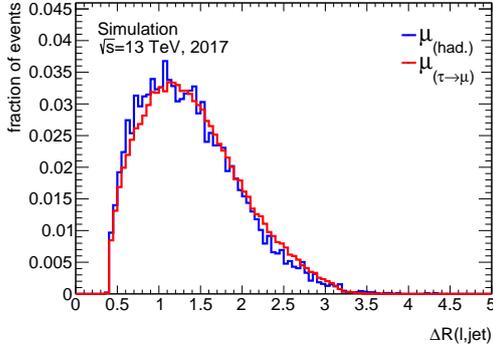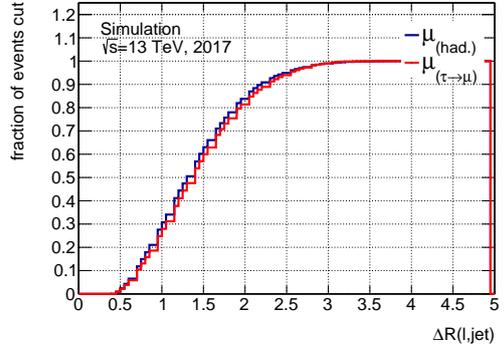
(a)   (b)

Figure 4.7: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable $\Delta R(\ell, \mathrm{jet})$. The distributions are very similar in both cases such that this variable was not considered for further study.
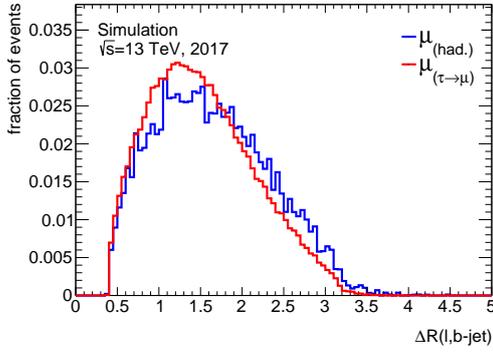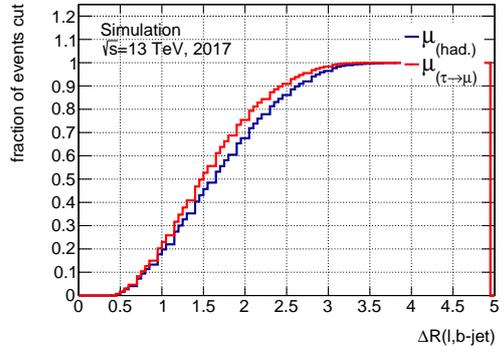
(a)   (b)

Figure 4.8: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable $\Delta R(\ell, b-\mathrm{jet})$. The distributions are very similar in both cases such that this variable was not considered for further study.
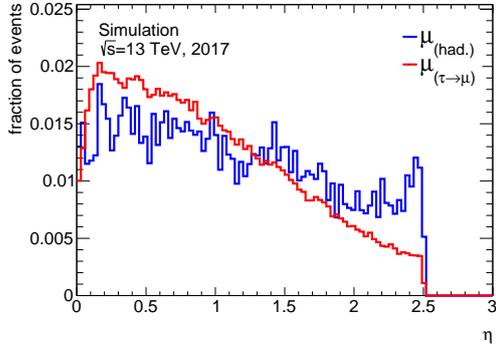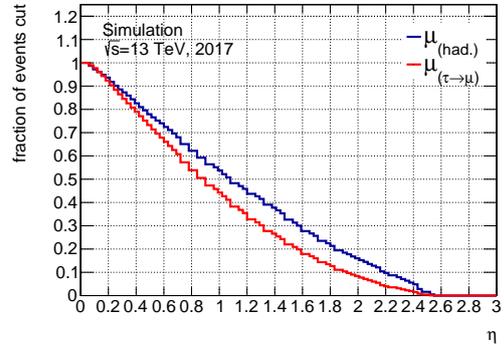
(a)   (b)

Figure 4.9: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable η. The main differences are in the tails of the distributions. A loose cut of around $\eta < 2.2$ could remove around 10% of background with around a signal loss of less than 5%.
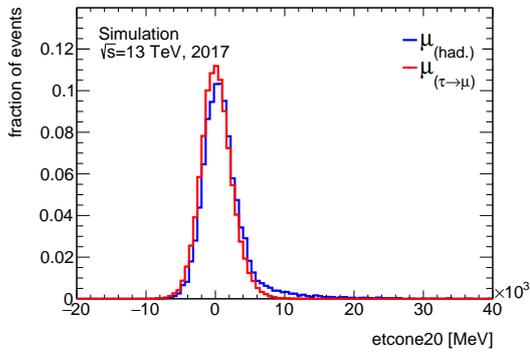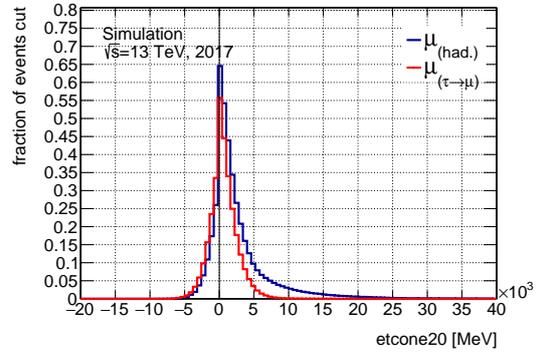
(a)   (b)

Figure 4.10: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *etcone*20. There is little discriminating power around *etcone*20= 0, however there is a clearer difference in the positive tails of the distributions.

97

(a)   (b)
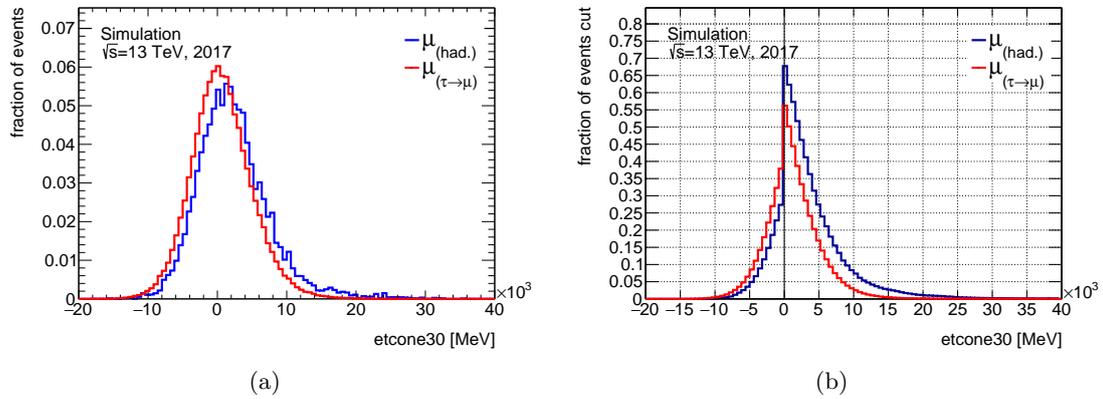
Figure 4.11: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *etcone*30. The distributions are very similar in both cases such that this variable was not considered for further study.

(a)   (b)
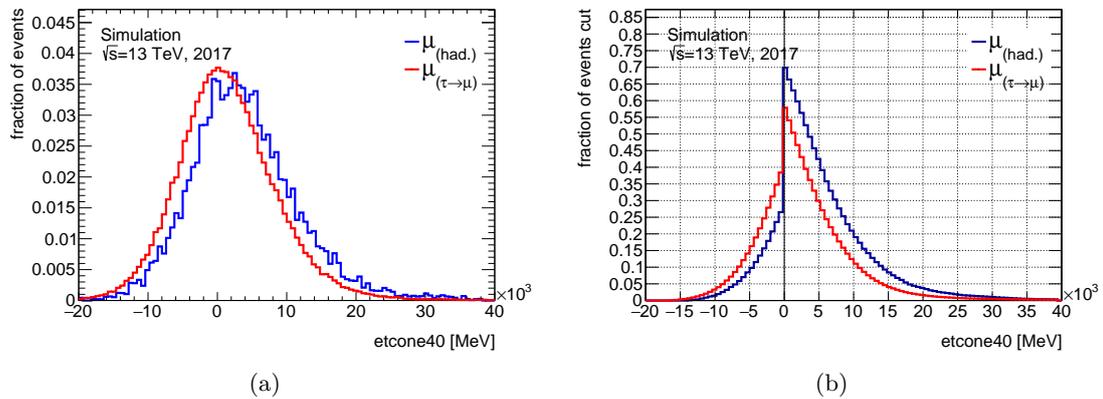
Figure 4.12: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *etcone*40. The distributions are very similar in both cases such that this variable was not considered for further study.

(a)   (b)
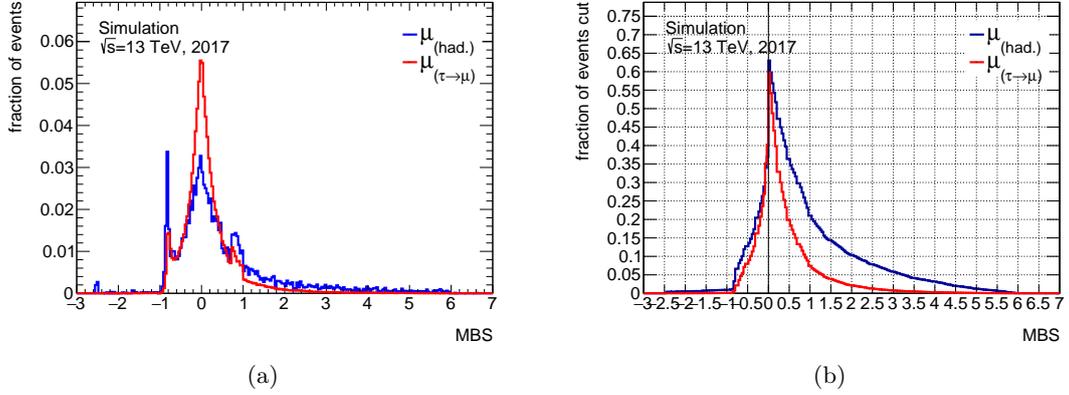
Figure 4.13: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable momentum balance significance. As mentioned in the previous section, the distributions are quite different between signal and background, with μ_{had} having a larger tail. A cut of around MBS< 2 was implemented, which reduces background by around 10%, with only around a 2% loss of signal.

(a)   (b)
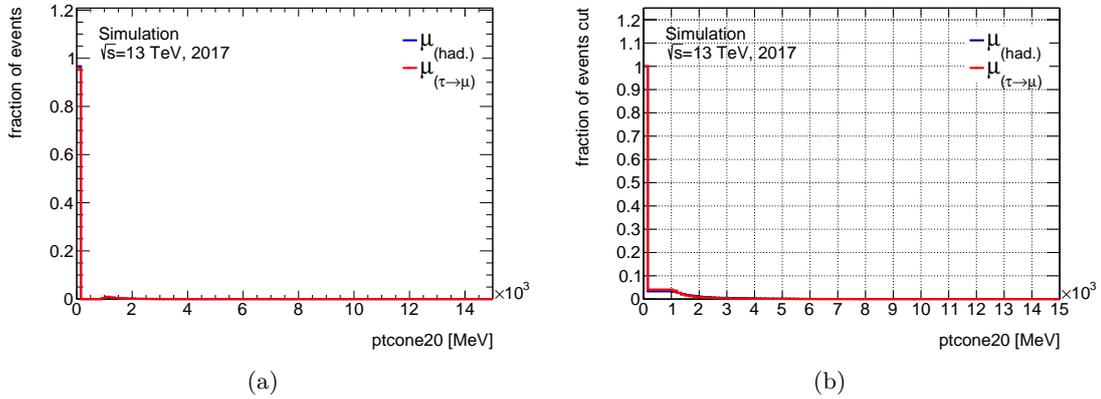
Figure 4.14: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable $ptcone20$. The distributions are very similar in both cases such that this variable was not considered for further study.

(a)   (b)
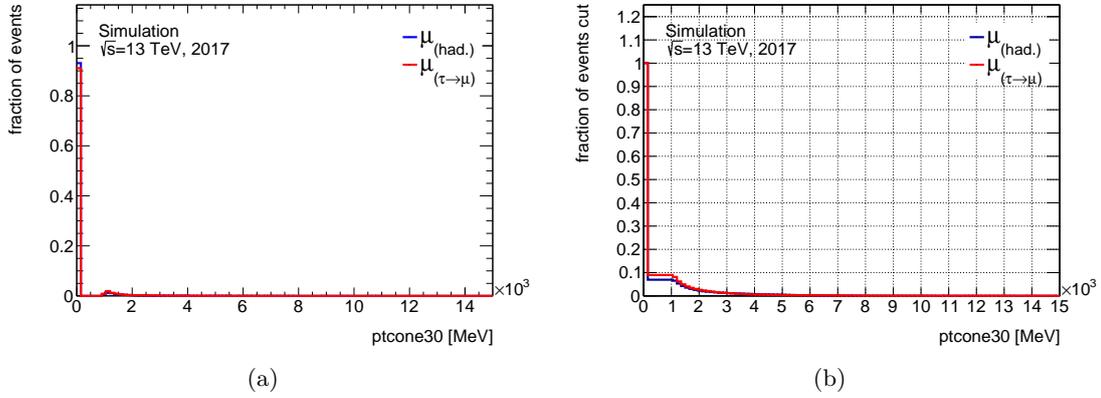
Figure 4.15: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *ptcone*30. The distributions are very similar in both cases such that this variable was not considered for further study.

(a)   (b)
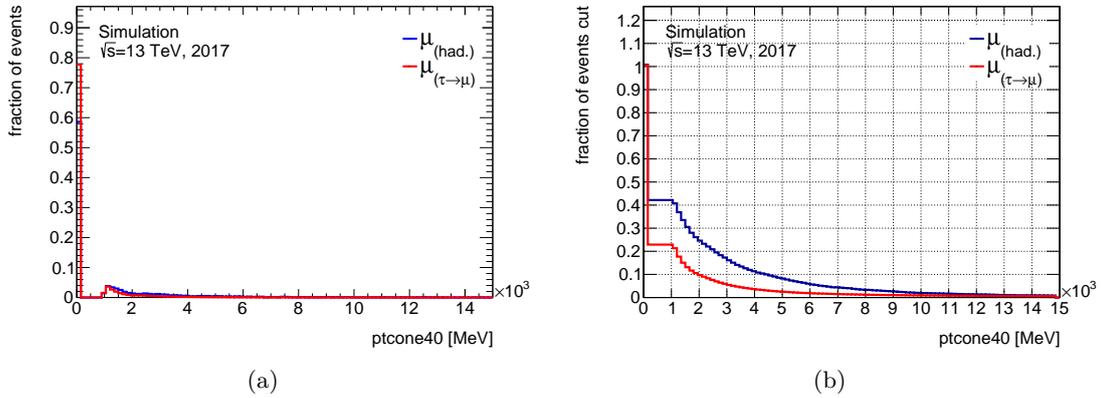
Figure 4.16: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *ptcone*40. The tails of this distribution are quite different between μ_{had} and $\mu_{\tau(\to\mu)}$, a feature which could be used to reduce background whilst maintaining signal.

100
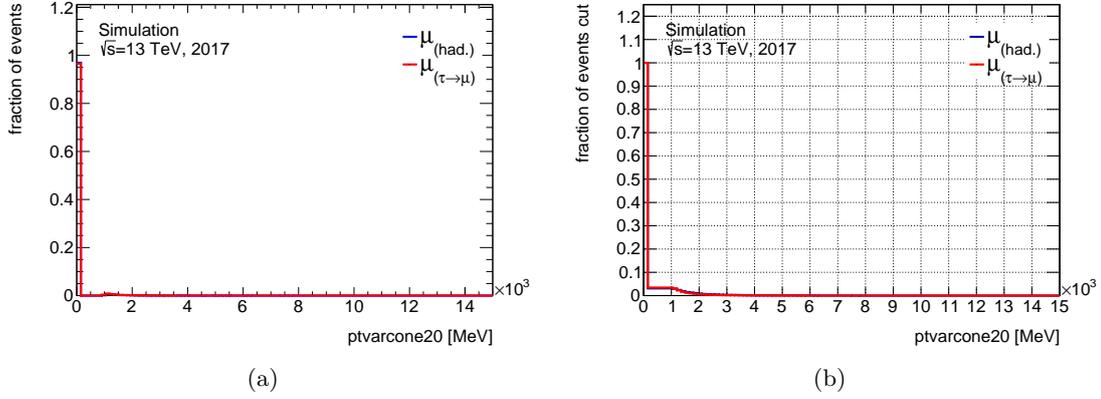
Figure 4.17: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *ptvarcone*20. The distributions are very similar in both cases such that this variable was not considered for further study.
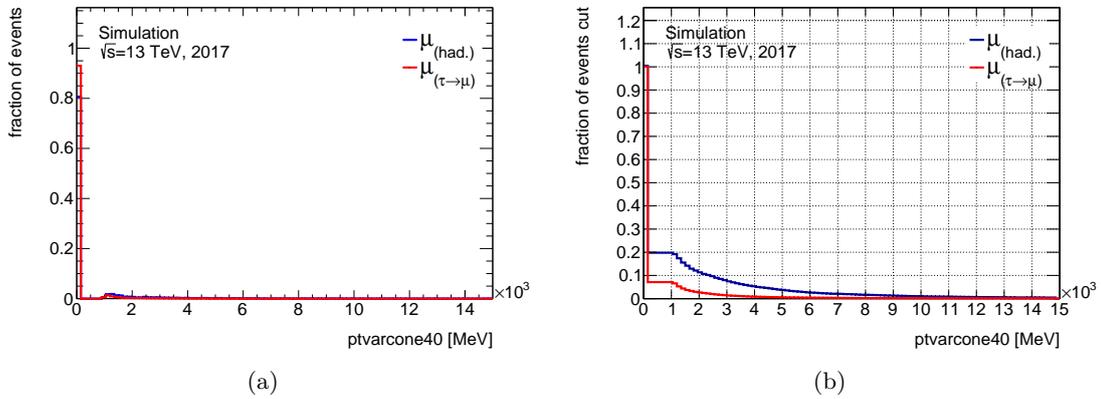
Figure 4.18: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *ptvarcone*40. The tails of this distribution are quite different between μ_{had} and $\mu_{\tau(\to\mu)}$, a feature which could be used to reduce background whilst maintaining signal.
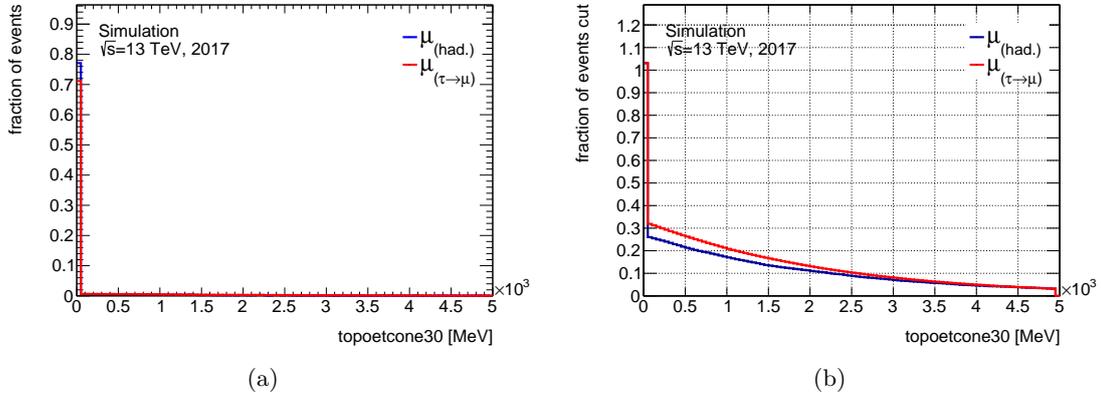
Figure 4.19: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *topoetcone*30. The distributions are very similar in both cases such that this variable was not considered for further study.
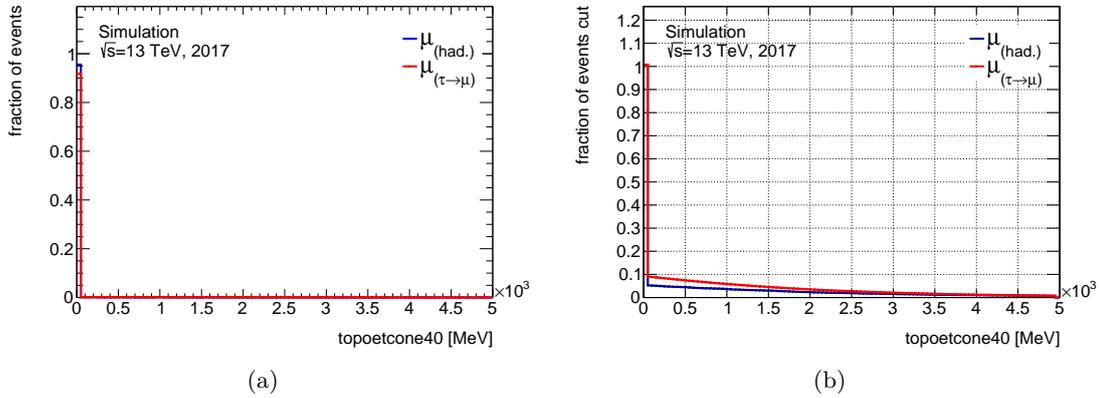
Figure 4.20: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable *topoetcone*40. The distributions are very similar in both cases such that this variable was not considered for further study.

102
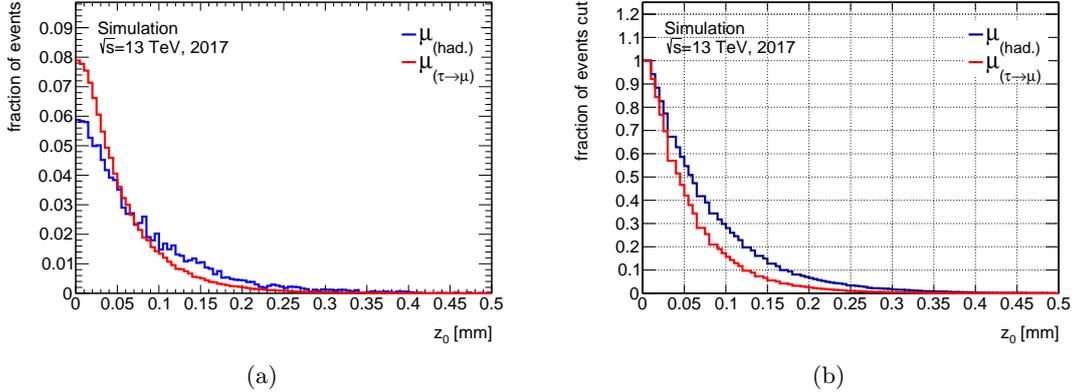
Figure 4.21: (a) the difference in shape between $\mu_{\tau(\to\mu)}$ signal and μ_{had} background and (b) the reduction in background and loss in signal for the variable $z_0 \sin\theta$. The distributions are very similar in both cases such that this variable was not considered for further study.

4.6 Muons from hadron decays background scale factors

Probe muons with significant displacement that come from a source other than $t \to bW^{\pm}(\to \tau(\mu\nu\nu)\nu)$ are a primary background to this analysis. The majority of these muons are produced in heavy flavour (*b*- and *c*-)hadron decays, and more rarely by the decays-in-flight of pions and kaons. There are also very minor contributions from other Standard Model processes. These background muons are labelled as μ_{had} throughout this section. The significant displacement of these muons populates the tails of the $|d_0^{\mu}|$ distributions in the region particularly sensitive to the fraction of muons produced via intermediate tau ($\mu_{\tau(\to\mu)}$), making them the most important background at high $|d_0^{\mu}|$. The estimated contributions of this background are corrected using data. Making use of a control region for each channel, scale factors are calculated that can be applied in the full profile-likelihood fit to extract $R(\tau/\mu)$ as nuisance parameters. This section describes the calculation of these scale factors.

4.6.1 Control regions

The nominal event selection produces a sample of $t\bar{t}$ events with a dileptonic final state, with the two leptons being of opposite charge. Estimating the scale factors of muons from hadron decays makes use of the fact that these are dominant in a dileptonic $t\bar{t}$ final state where both leptons are of the same charge, at a similar rate to the signal region but with much lower μ_{prompt} and $\mu_{\tau(\to\mu)}$ contributions. Therefore, a sample of same-sign events can be used as a control region to determine the scale factors for the opposite-sign sample. The scale factors are calculated individually for e-μ and μ-μ, such that there is a same-sign control region for each channel. The same-sign control regions
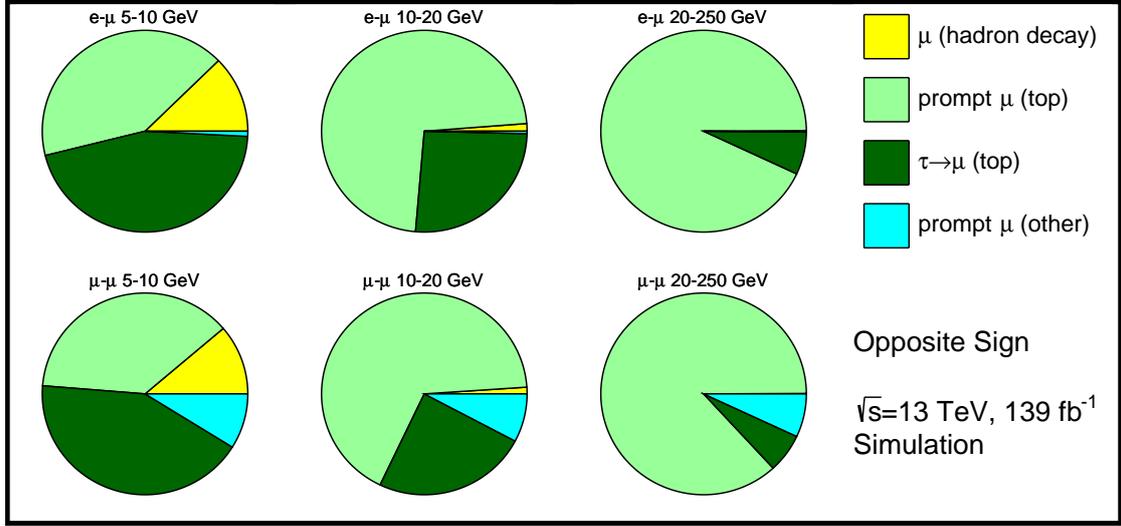
Figure 4.22: Proportions of *probe* muons in the opposite-sign signal region that are from hadron decays, from intermediate tau decays, prompt, or charge-flipped. The largest contributions from μ_{had} are in the lowest p_{T} bins, where they are a significant background to the $\mu_{\tau(\to\mu)}$ signal.
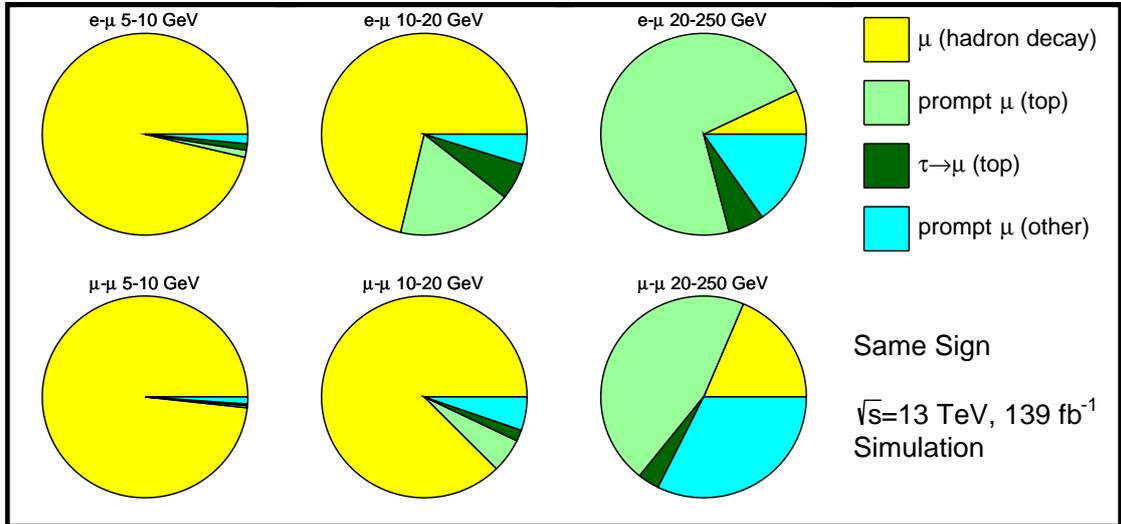
Figure 4.23: Proportions of *probe* muons in the same-sign control region that are from hadron decays, from intermediate tau decays, prompt, or charge-flipped. The μ_{had} contributions dominate the same-sign region in all but the highest p_{T} bin. At higher p_{T} there are prompt contributions arising from $t\bar{t} + V$. The e-μ channel also sees contributions from events with *tag* electrons that have misidentified charge.
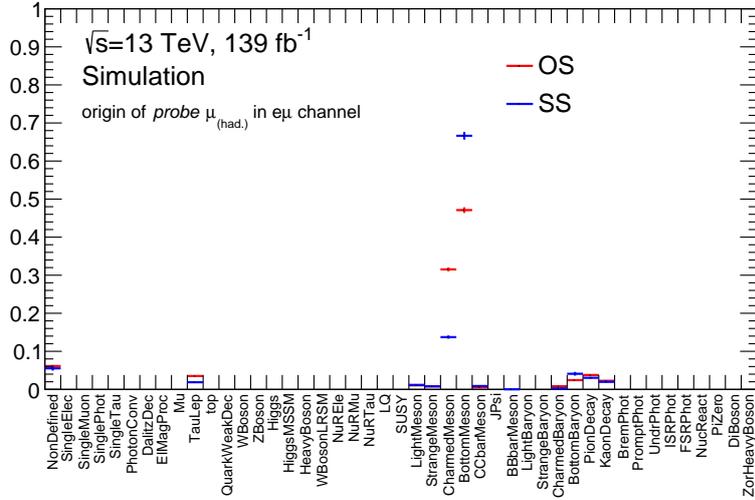
Figure 4.24: Sources of muons from hadron decays for same-sign (SS) and opposite-sign (OS) e-μ events. The dominant contributions are seen to come from the decays of charmed and bottom hadrons. There is a difference in the contributions in the same-sign and opposite-sign regions. The larger charmed meson contribution in the opposite-sign region arises from a $W^{\pm}$ decaying to $c\bar{s}$, as muons produced in the c decay will have opposite charge to the *tag* lepton. Little difference is seen between the e-μ and μ-μ channels, which can be seen by comparing with Figure 4.25.
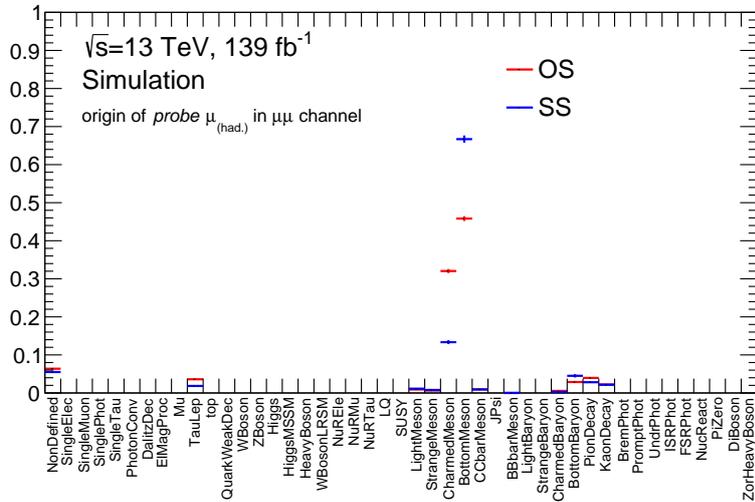
Figure 4.25: Sources of muons from hadron decays for same-sign (SS) and opposite-sign (OS) μ-μ events. The dominant contributions are seen to come from the decays of charmed and bottom hadrons. There is a difference in the contributions in the same-sign and opposite-sign regions. The larger charmed meson contribution in the opposite-sign region arises from a $W^{\pm}$ decaying to $c\bar{s}$, as muons produced in the c decay will have opposite charge to the *tag* lepton. Little difference is seen between the e-μ and μ-μ channels, which can be seen by comparing with Figure 4.24.

105

produce high purity samples of muons from hadron decays. The proportions of signal and background muons in the signal and control regions are shown in Figures 4.22 and 4.23.

The composition of *probe* muons from different sources of hadron decays are shown in Figures 4.24 and 4.25. These plots show that *probe* muons produced via heavy flavour decays dominate, with a higher rate of *probe* muons produced via charmed meson decays in the opposite-sign signal region compared to the same-sign control region. This difference is understood, since $W^{\pm}$-bosons decay to $c\bar{s}$ around 30% of the time [32]. The leptons that are produced in the decay chain of the c-mesons will have the opposite charge to the lepton from the other $W^{\pm}$-boson produced in the initial $t\bar{t}$ decay, therefore contributing in the opposite-sign selection. The other primary source of muons from hadron decays come from the b and $\bar{b}$ produced in all top quark decays (including any charmed hadrons produced in their decay chains). These contribute equally to the background in same-sign and opposite-sign events. It is expected that the contributions will be the same in both channels, since the *probe* muon is independent of the *tag* lepton, and this is observed in the figures.

4.6.2 Method

The normalisation scale factors are determined using a data-driven method in each of the e-μ and μ-μ same-sign control regions. The extrapolation from same-sign control region to opposite-sign signal region is estimated using MC simulated data. In both of the control regions, at high p_{T}^{μ} and low $|d_0^{\mu}|$, there are background contributions from $t\bar{t} + V$, and in the e-μ same-sign region there is a background contribution at high p_{T}^{μ} from charge misidentification of the electron. The $p_{\mathrm{T}}^{\mu} > 30$ GeV region is dominated by these muons though it is still a small number of events. Before the scale factors are calculated, the background contributions are separately scaled according to the number of *probe* muons with $p_{\mathrm{T}}^{\mu} > 30$ GeV.

Putting this all together, one obtains the following equations:

$$N_{\mathrm{data}}^{\mu_{\mathrm{had}},\mathrm{OS}} = R N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{OS}} \tag{4.3}$$

where $N_{\mathrm{data}}^{\mu_{\mathrm{had}},\mathrm{OS}}$ is the estimated total number of μ_{had} events in the signal region in data, $N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{OS}}$ is the number of μ_{had} events in simulation and R is the scale factor. R is defined as:

$$R = \frac{N_{\mathrm{data}}^{\mathrm{SS}} - N_{\mathrm{MC}}^{\mathrm{prompt},\mathrm{SS},\mathrm{scaled}}}{N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{SS}}}, \tag{4.4}$$

where $N_{\mathrm{data}}^{\mathrm{SS}}$ is the observed number of same-sign μ_{had} events in data, $N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{SS}}$ is the is the number of same-sign events in simulation, and $N_{\mathrm{MC}}^{\mathrm{prompt},\mathrm{SS},\mathrm{scaled}}$ is the scaled number of prompt *probe* muons. $N_{\mathrm{MC}}^{\mathrm{prompt},\mathrm{SS},\mathrm{scaled}}$ is calculated as:

$$N_{\mathrm{MC}}^{\mathrm{prompt},\mathrm{SS},\mathrm{scaled}} = S N_{\mathrm{MC}}^{\mathrm{prompt},\mathrm{SS}} \tag{4.5}$$

where S is defined as the number of *probe* muons with $p_\mathrm{T}^\mu > 30$ GeV:

$$S = \left[\frac{N_\mathrm{data}^\mathrm{SS} - N_\mathrm{MC}^{\mu_\mathrm{had},\mathrm{SS}}}{N_\mathrm{MC}^\mathrm{prompt,SS}} \right]_{p_\mathrm{T}^\mu > 30\,\mathrm{GeV}} . \tag{4.6}$$

These calculations are performed separately for e-μ and μ-μ. The calculated values and their inputs are shown in Tables 4.4 and 4.5.

4.6.3 Systematic Uncertainties

The scale factors have systematic uncertainties due to the size of the same-sign dataset, due to the choice of MC generators used and due to the uncertainty in the subtraction of other processes in the same-sign control region. These systematic uncertainties are not calculated on a year-by-year basis, instead they use the full dataset.

The uncertainty due to the choice of MC generator is estimated by using a range of $t\bar{t}$ Monte Carlo configurations and comparing the total estimated number of muons from hadron decays using equation (4.3). Eight different configurations are compared against POWHEG+PYTHIA8-AF2 (ATLAS fast 2). These configurations are the following:

- POWHEG+HERWIG uses the HERWIG v7.04 [104,105], H7UE tune [105], `MMHT2014LO` PDF set [153], which is an alternative generator to PYTHIA8. This varies the parton shower and hadronisation model.

- POWHEG+PYTHIA8-h_damp is the configuration wherein the POWHEG h_damp parameter is varied from $1.5m_\mathrm{top}$ to $3m_\mathrm{top}$, which varies the resummation scale.

- Final State Radiation up/down (FSR up/down) varies the renormalisation scale for QCD emission in the FSR up (down) by a factor of 2 (0.5), to vary the FSR tuning.

- Initial State Radiation up/down (ISR up/down) uses the up or down variation of the A14 eigen-tune variation [126] of the strong coupling constant (α_s), to vary the ISR tuning.

- $\mu_{R,F}$ up/down (muRF up/down) varies the factorisation and renormalisation scales up (down) by a factor of 2 (0.5), in order to simulate changes to the amount of parton radiation and potential missing higher-order corrections.

The alternate configurations are all generated using the faster ATLAS fast 2 detector simulation programme (see Section 2.8.2), so POWHEG+PYTHIA8-AF2 is used for comparison to avoid folding in the differences between the different simulation types. The comparison of the estimated μ_had yields in data is shown in Table 4.7. The systematic uncertainty due to top modelling for each channel is taken from the largest difference as a percentage. The tables showing all input and output values can be found in Appendix C.1.

The differences in OS/SS ratios between the Powheg+Pythia8-AF2 and Powheg+Herwig configurations for each of the different μ_{had} sources is shown in Figures 4.26 and 4.27. Very good agreement is shown in the dominant contributions from b- and c-hadrons. A small number of sources of μ_{had} have different opposite-sign to same-sign ratios in the different generators. However the differences are at most 30% and these are seen in components that contribute only a very small amount (at most a 5%, as shown in Figures 4.24 and 4.25) to the overall yield of μ_{had} so are not a cause for concern. Therefore there is no additional uncertainty estimated to account for differences in composition between the two configurations. The effect of any differences should be covered by the generator uncertainty for the scale factors, and in by the parton shower and hadronisation uncertainty for the μ_{had} templates.

The systematic uncertainty due to the subtraction of other processes in the same-sign control region is estimated by using a number of different p_{T}^{μ} cuts and comparing the scale factors against the nominal scale factor. Four different p_{T}^{μ} cuts are used in total: $p_{\text{T}}^{\mu} < 20$ GeV, $p_{\text{T}}^{\mu} < 30$ GeV (nominal), $p_{\text{T}}^{\mu} < 40$ GeV, and $p_{\text{T}}^{\mu} < 50$ GeV. The scale factors are recalculated for each different p_{T}^{μ} cut, and the estimated number of μ_{had} events in data is compared against the nominal estimate. The comparison of the estimated μ_{had} yield in data is shown in Table 4.8. The systematic uncertainty due to the p_{T} cut for each channel is taken from the largest difference as a percentage. The tables showing all input values can be found in Appendix C.2.

Uncertainties on the μ_{had} templates, including top modelling and PDF uncertainties, are described in Section 4.9.
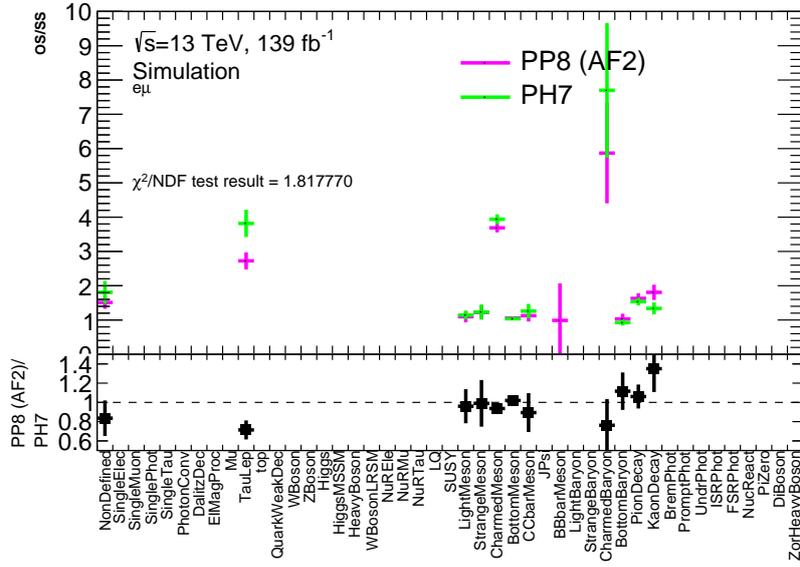
Figure 4.26: OS/SS for each muons from hadron decays background lepton source for e-μ channel for POWHEG+PYTHIA8 and POWHEG+HERWIG. Very good agreement is seen for the dominant bottom and charmed meson components.
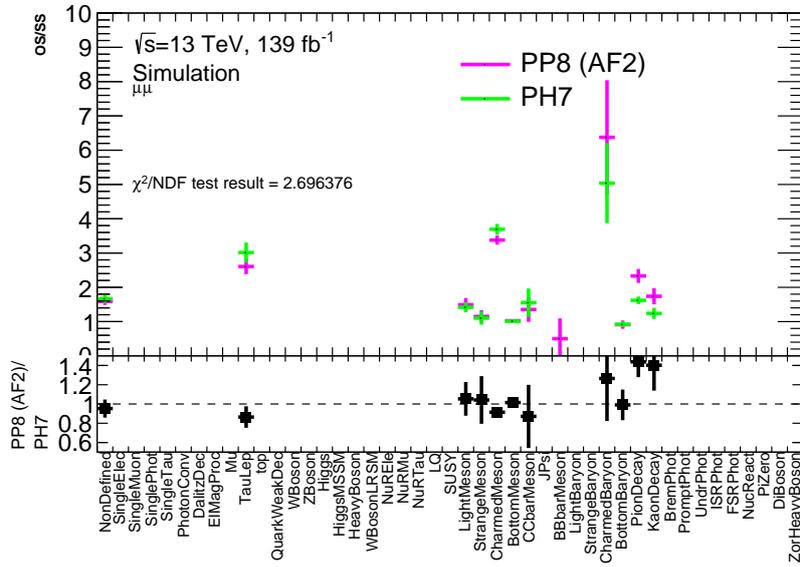
Figure 4.27: OS/SS for each muons from hadron decays background lepton source for μ-μ channel for POWHEG+PYTHIA8 and POWHEG+HERWIG. Very good agreement is seen for the dominant bottom and charmed meson components.

4.6.4 Results

The normalisation scale factors for the muons from hadron decays background are found to be 1.39 (1.37) in the e-μ (μ-μ) channel. The values used to calculate these are summarised in Tables 4.4 and 4.5.

The uncertainties in the e-μ (μ-μ) control region normalisation are 4% (4%) due to the size of the same-sign dataset; 8% (3%) due to the choice of MC generators used; and 1.0% (1.3%) due to the uncertainty in the subtraction of the other processes in the same-sign control region. These are summarised in Table 4.6. The estimates of the number of OS μ_{had} events in data as calculated with the different $t\bar{t}$ modelling configurations is shown in Table 4.7, and as calculated with the different p_{T} cuts is shown in Table 4.8.

tag-probe	e-μ	μ-μ
$N_{MC}^{\mathrm{SS,prompt}}$	776.77 ± 8.41	284.15 ± 2.80
$N_{\mathrm{data}}^{\mathrm{SS},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	666.00 ± 25.81	309.00 ± 17.58
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	21.38 ± 1.67	25.26 ± 1.88
$N_{MC}^{\mathrm{SS,prompt},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	489.57 ± 6.71	171.77 ± 2.06
S	1.32 ± 0.06	1.65 ± 0.10
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	1022.79 ± 44.75	469.37 ± 30.14

Table 4.4: Values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events.

tag-probe	e-μ	μ-μ
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2134.00 ± 46.20
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1207.47 ± 16.39	1217.88 ± 15.21
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1767.67 ± 17.67	1736.61 ± 16.83
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	1022.79 ± 44.75	469.37 ± 30.14
R	1.39 ± 0.06	1.37 ± 0.05
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2453.87 ± 108.57	2373.64 ± 87.14

Table 4.5: Values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events.

	Choice of MC generator	Background Subtraction	Statistical
e-μ	7.66%	0.96%	4.32%
μ-μ	3.02%	1.32%	3.65%

Table 4.6: Summary of the uncertainties on the normalisation scale factors in the two channels.

Configuration	$N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$	Δ	%
e-μ			
Powheg+Pythia8 (full-sim)	2453.87 ± 108.57	n/a	n/a
Powheg+Pythia8 (fast-sim)	2546.37 ± 112.71	n/a	n/a
Powheg+Herwig	2643.60 ± 117.67	-97.22	-3.82%
Powheg+Pythia-h_{damp}	2502.31 ± 108.56	44.06	1.73%
FSR up	2616.18 ± 120.98	-69.80	-2.74%
FSR down	2351.33 ± 121.22	195.04	7.66%
ISR up	2542.12 ± 112.58	4.25	0.17%
ISR down	2547.72 ± 113.47	-1.35	-0.05%
$\mu_{R,F}$ up	2535.33 ± 113.23	11.04	0.43%
$\mu_{R,F}$ down	2541.78 ± 113.81	4.59	0.18%
μ-μ			
Powheg+Pythia8 (full-sim)	2373.64 ± 87.14	n/a	n/a
Powheg+Pythia8 (fast-sim)	2460.66 ± 91.01	n/a	n/a
Powheg+Herwig	2535.72 ± 94.67	-75.06	-3.05%
Powheg+Pythia-h_{damp}	2460.46 ± 90.88	0.20	0.01%
FSR up	2526.98 ± 97.53	-66.32	-2.70%
FSR down	2396.82 ± 101.56	63.84	2.59%
ISR up	2442.14 ± 90.43	18.52	0.75%
ISR down	2473.10 ± 91.91	-12.44	-0.51%
$\mu_{R,F}$ up	2469.2 ± 91.25	-8.55	-0.35%
$\mu_{R,F}$ down	2444.06 ± 91.14	16.60	0.67%

Table 4.7: Comparisons of the estimated number of μ_{had} events using each of the different configurations. The values are compared against the nominal configuration (Powheg+Pythia8) using the ATLAS Fast-2 (AF2) detector simulation. The nominal configuration using the full ATLAS simulation is also included for reference. The largest differences are seen with FSR down in the *e-μ* channel and with Powheg+Herwig for the *μ-μ* channel.

p_T cut	$N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$	Δ	%
	e-μ		
$p_T > 20$ GeV	2476.20 ± 105.58	-22.32	-0.91%
$p_T > 30$ GeV (nominal)	2453.87 ± 108.57	0	0%
$p_T > 40$ GeV	2445.24 ± 113.04	8.63	0.35%
$p_T > 50$ GeV	2477.33 ± 119.27	-23.46	-0.96%
	μ-μ		
$p_T > 20$ GeV	2346.96 ± 86.14	26.68	1.12%
$p_T > 30$ GeV (nominal)	2373.64 ± 87.14	0	0%
$p_T > 40$ GeV	2370.81 ± 89.44	2.83	0.12%
$p_T > 50$ GeV	2404.99 ± 92.22	-31.34	-1.32%

Table 4.8: Comparisons of the estimated number of μ_{had} events using each of the different p_T cuts, for each of the e-μ and μ-μ channels. The values are compared against the nominal configuration ($p_T > 30$ GeV). The largest differences are seen with $p_T > 50$ GeV in both the e-μ and μ-μ channels.

4.6.5 Cross-checks

In order to prove the validity of this method, a number of cross checks are performed. Firstly, the $|d_0^\mu|$ and p_T^μ distributions of muons from hadron decays are compared in same-sign events in data. These distributions are estimated in data by subtracting the estimated (using MC simulated data) contributions coming from $\mu_{\tau(\to\mu)}$ and prompt backgrounds, with the prompt backgrounds scaled using equation (4.5). The distributions are then normalised to unity. A systematic uncertainty due to the choice of MC generator is applied for these cross-checks. This is estimated in each bin of each histogram as the spread of the generators, and this is added in quadrature to the statistical uncertainty. These comparisons are shown in Figures 4.28, 4.29 and 4.30. Good agreement is seen in the shape of the distribution, giving confidence that the two variables of interest, p_T^μ and $|d_0^\mu|$ are well modelled by the simulation. This justifies the use of a scale factor calculated from data, whilst taking the shape in p_T^μ and $|d_0^\mu|$ in simulation.

Figures 4.31 and 4.32 show the $|d_0^\mu|$ and p_T^μ distributions in the same-sign control region after the fit to extract $R(\tau/\mu)$ has been performed, in each of the e-μ and μ-μ channels. Whilst it is worth noting that the background normalisation scale factors are used as inputs to the fit (see Section 4.8), the data and MC distributions agree well.

Additionally, the scale factors are calculated for each year of data-taking, as well as separately depending on the charge of the *probe* muon. Figure 4.33 compares the background scale factors for the muons from hadron decays background in each of the data-taking periods against the nominal value. Figure 4.34 compares the background scale factors for the muons from hadron decays background calculated separately according to the charge of the *probe* muon, against the inclusive value.
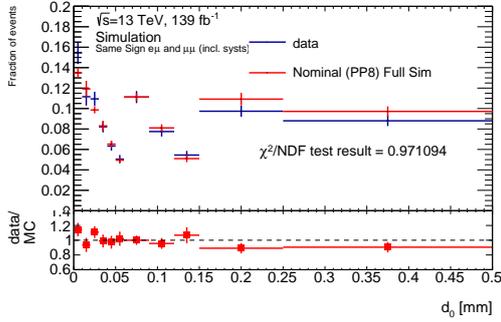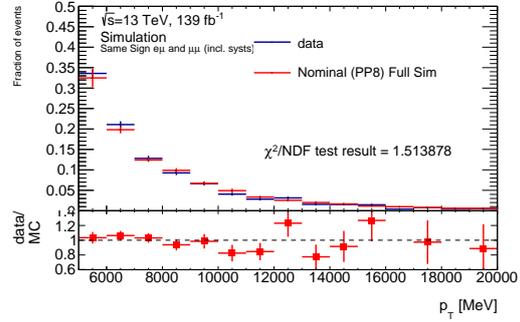
Figure 4.28: Comparison of distributions of (a) $|d_0^\mu|$ and (b) p_T^μ in same-sign e-μ and μ-μ events in data and simulation for the full dataset. The $|d_0^\mu|$ distributions are not normalised to the bin widths. The systematic uncertainty in each bin of this histogram is calculated as half the difference between the highest and lowest MC generator. The χ^2 test is performed after adding this systematic uncertainty in quadrature with the statistical uncertainty. The results of the χ^2 test are shown in the legend. This systematic is applied only to these same-sign distributions for this comparison.
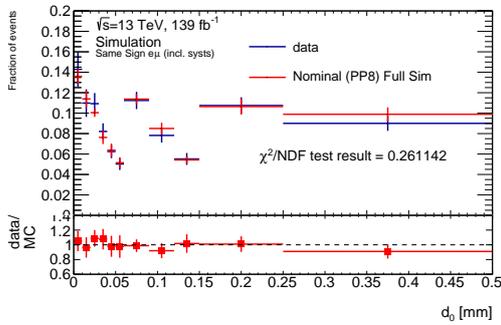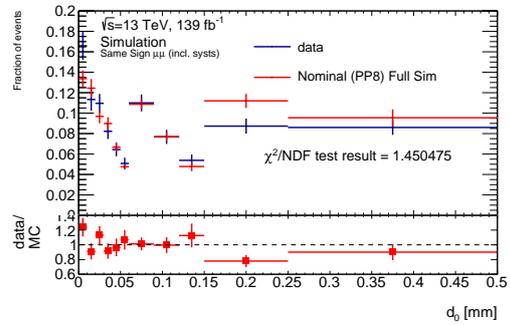
Figure 4.29: Comparison of distributions of $|d_0^\mu|$ in same-sign (a) e-μ and (b) μ-μ events in data and simulation for the full dataset. The distributions are not normalised to the bin widths. The systematic uncertainty in each bin of this histogram is calculated as half the difference between the highest and lowest MC generator. The χ^2 test is performed after adding this systematic uncertainty in quadrature with the statistical uncertainty. The results of the χ^2 test are shown in the legend. This systematic is applied only to these same-sign distributions for this comparison.

115
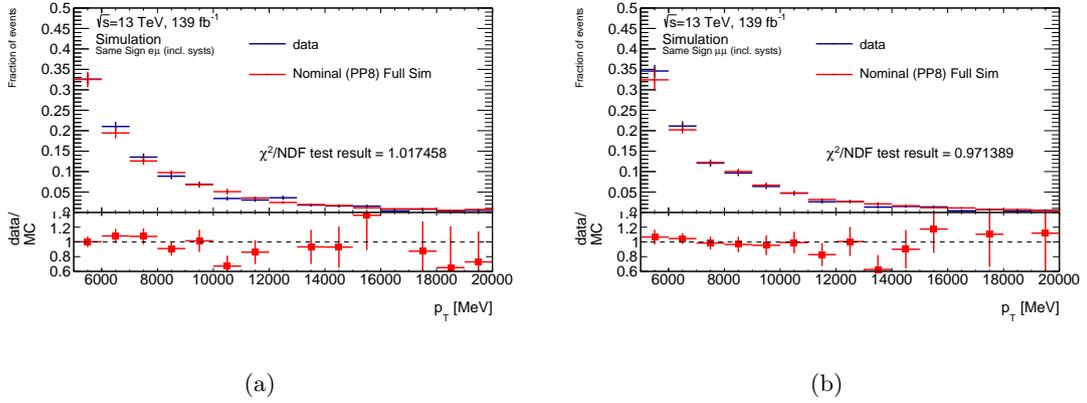
Figure 4.30: Comparison of distributions of p_T^μ in same-sign (a) e-μ and (b) μ-μ events in data and simulation for the full dataset. The systematic uncertainty in each bin of this histogram is calculated as half the difference between the highest and lowest MC generator. The χ^2 test is performed after adding this systematic uncertainty in quadrature with the statistical uncertainty. The results of the χ^2 test are shown in the legend. This systematic is applied only to these same-sign distributions for this comparison.
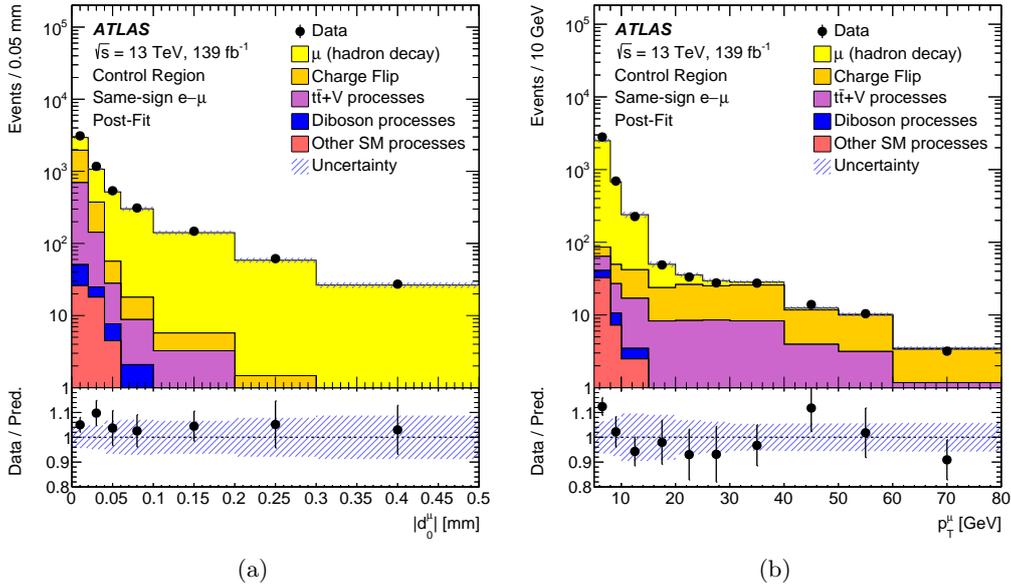
Figure 4.31: The *probe* muon (a) $|d_0^\mu|$ and (a) p_T^μ distributions in the e-μ channel same-sign μ(hadron decay) control region, post fit. The calculated μ_{had} scale factor has been applied. The bottom panel shows the data/MC ratio. The blue bands indicate the systematic uncertainties with the constraints from the analysis fit of the signal region data applied. Good agreement is seen between data and prediction, which gives confidence in the method used to calculate the scale factors.

116
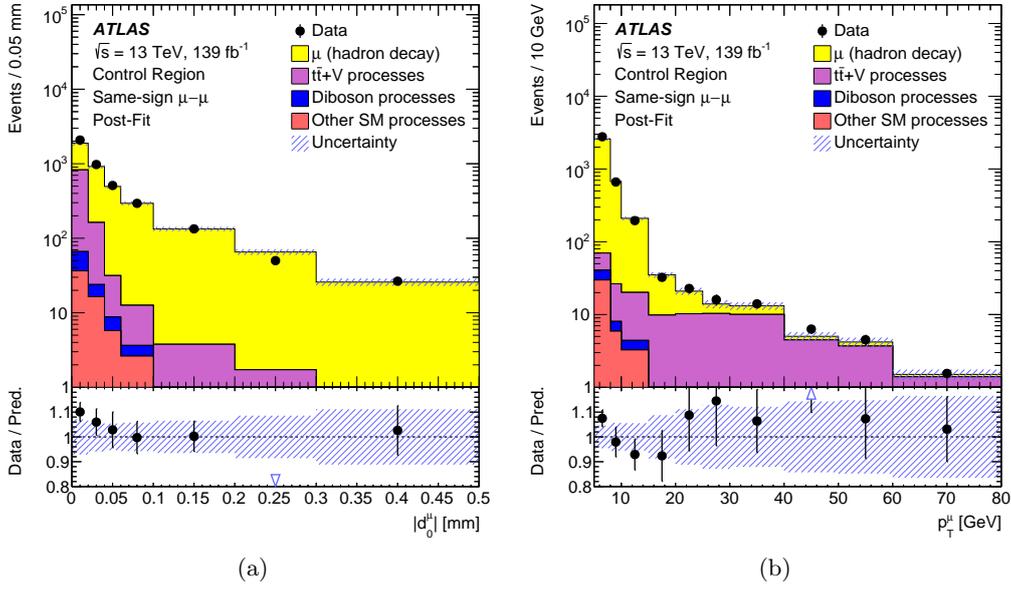
Figure 4.32: The *probe* muon (a) $|d_0^\mu|$ and (a) p_T^μ distributions in the μ-μ channel same-sign μ(hadron decay) control region, post fit. The calculated μ_{had} scale factor has been applied. The bottom panel shows the data/MC ratio. The blue bands indicate the systematic uncertainties with the constraints from the analysis fit of the signal region data applied. Good agreement is seen between data and prediction, which gives confidence in the method used to calculate the scale factors.
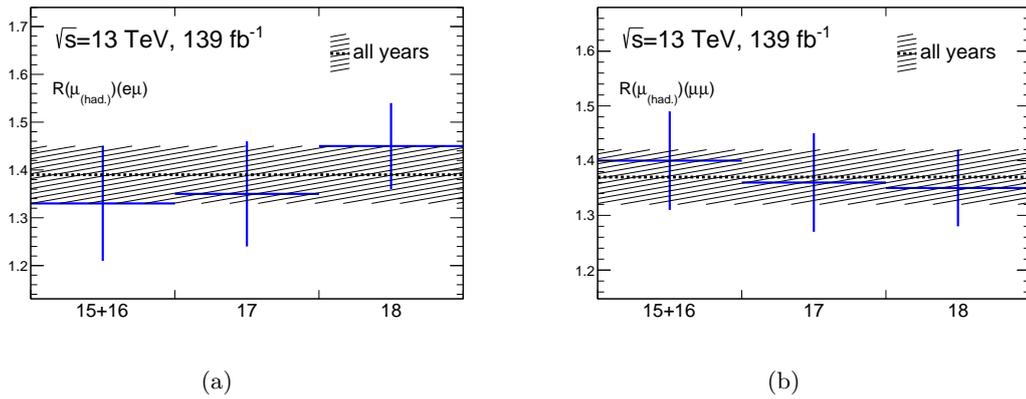
Figure 4.33: The muon from hadron decays background scale factors calculated in the different data-taking periods, compared against the inclusive value. Only statistical uncertainties are shown. The values obtained in each of the periods all agree well within uncertainties.

117
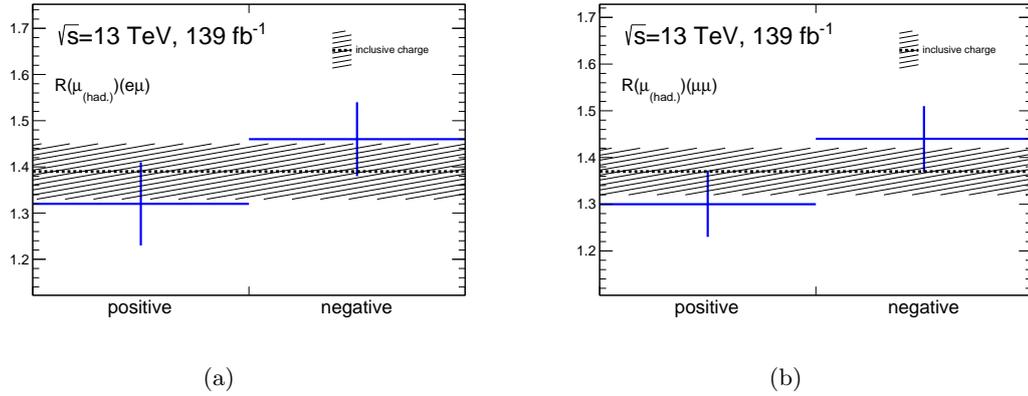
Figure 4.34: The muon from hadron decays background scale factors calculated exclusively in charge, compared against the inclusive value. Only statistical uncertainties are shown. The values obtained for both charges agree well within uncertainties.

4.6.6 Additional cross-check

An additional cross-check was performed in a sub-region of the signal region (i.e. opposite-sign) with a high proportion of muons from hadron decays. Cutting on the *ptcone*40 variable is able to produce a relatively pure sample of muons from hadron decays within the signal region. Good agreement is seen by applying the normalisation scale factors. This is shown in Figure 4.35.
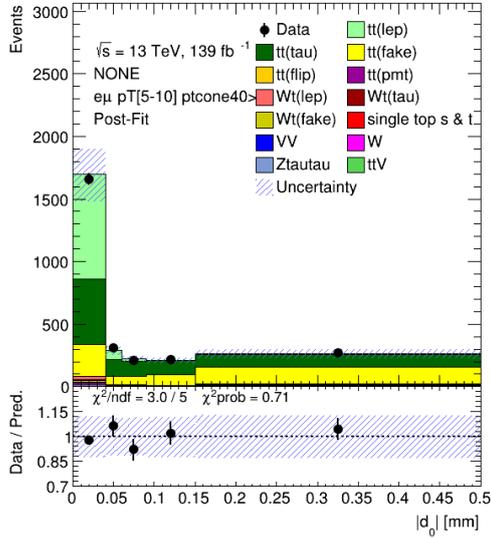
(a) $e - \mu$ channel, ptcone40> 1 GeV
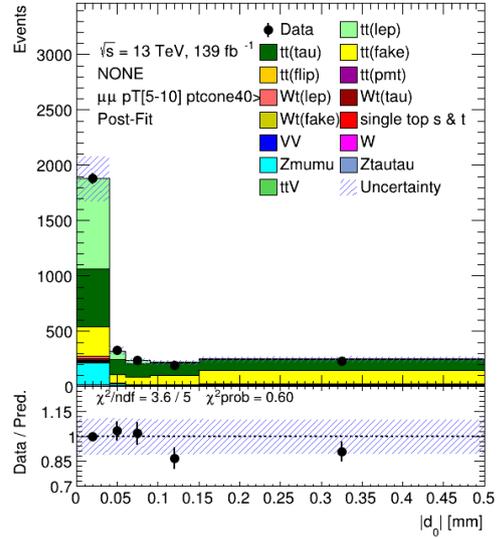
(b) $\mu - \mu$ channel, ptcone40> 1 GeV
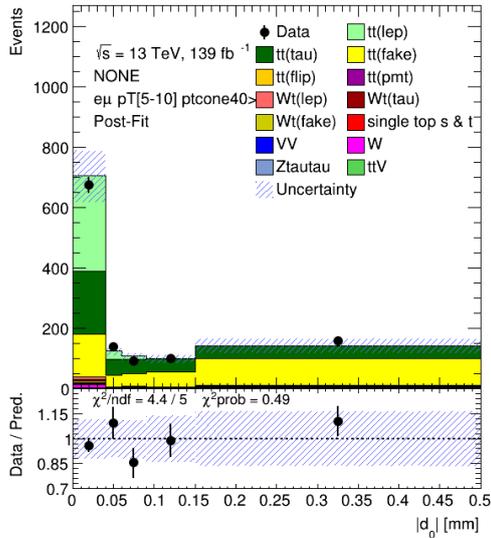
(c) $e - \mu$ channel, ptcone40> 2 GeV
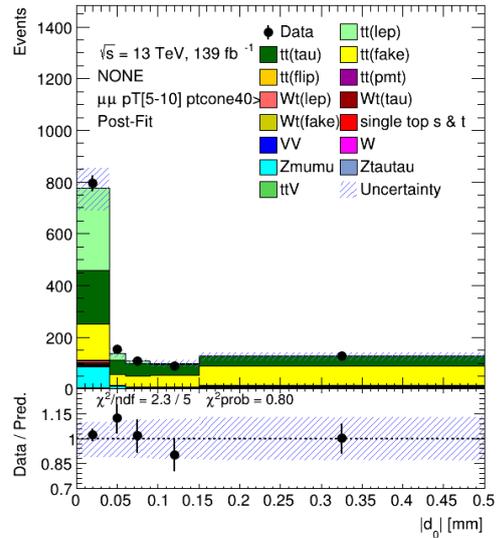
(d) $\mu - \mu$ channel, ptcone40> 2 GeV

Figure 4.35: The distribution of $|d_0^\mu|$ for a region enriched with muons from hadron decays (here labelled tt(fake). For the $ptcone40> 1\ GeV$ ($ptcone40> 1\ GeV$) region, a scaling of 25% (30%) is applied to to other muon sources. This is to account for mis-modelling due to stray tracks and pile-up. Good agreement is seen after the muons from hadron decays scale factors have been applied.

119

4.7 Z background scale factors

The Drell-Yan process [154] is the production of a virtual Z^0 boson or photon in hadron collisions, occurring when a quark of one hadron annihilates with the anti-quark of another. The dileptonic decays of these Z^0 bosons are another primary background of this analysis. They populate the low $|d_0^\mu|$ region, where they occupy the same phase space as $t \to bW^\pm(\to \mu\nu)$ in the μ-μ channel. The estimated contributions from this process are corrected using data. Making use of a control region, a scale factor is calculated that can be applied in the full profile-likelihood fit to extract $R(\tau/\mu)$ as a nuisance parameter. This section describes the calculation of this scale factor.

4.7.1 Control Regions

In the nominal selection, resonant $Z^0(\to \mu\mu) + (b\text{-})$jets events are removed by excluding events with dilepton invariant mass in the range $85 < m(\mu\mu) < 95$ GeV. However, the Z^0 resonance line-shape is well known, and can be used to compare data and MC. As such, a control region is defined by dropping this invariant mass requirement, whilst keeping all other selection criteria the same. This produces a sample with a clear resonant $m(\mu\mu)$ spectrum. This control region has different criteria compared to the $Z^0 \to \mu\mu$ calibration region used in Section 4.4.4, most notably the (b-)jet requirements.

The study is performed using the full dataset and signal and background MC sample. This is to ensure that the shape of the invariant mass distribution is fully compatible between data and MC.

4.7.2 Method

A χ^2-fit is performed across the invariant mass distribution in the Z^0-peak region. The fit is performed separately for both data and MC over the range $50 < m(\mu\mu) < 140$ GeV. A Voigt profile [155] (a convolution of a Breit-Wigner and a Gaussian) is used for the $Z^0 \to \mu\mu$ resonance and a third-order Chebychev polynomial is used for all other processes. The Breit-Wigner component of the Voigt profile describes the $Z^0 \to \mu\mu$ resonance, and the Gaussian component describes the detector resolution. The Chebychev polynomial provides a smooth background to the resonant part of the $m(\mu\mu)$ spectrum.

The fit is configured using the RooFit toolkit for data modelling [156]. This provides a Voigt profile with the following parameters:

- Decay width (Γ_{Z^0})

- Central value of the peak corresponding to both the Z^0 mass and Gaussian mean of the resolution

- Gaussian width (σ)

Parameter	Starting value
N_V and N_C	1.
All Chebychev coefficients	0
The central value of the peak	90 GeV
The Gaussian width (σ)	5 GeV
Z^0 decay width Γ_{Z^0}	$\Gamma_{Z^0} = 2.4952$ GeV

Table 4.9: The parameters and their starting values used in the fit. All parameters apart from the Z^0 decay width are free-floating in the fit.

The Chebychev polynomial is provided with four coefficients. Both functions take the invariant mass ($m(\mu\mu)$) as input. The Voigt profile ($V(m)$) and Chebychev polynomial ($C(m)$) are combined as a composite model probability density function (p.d.f.) ($M(m)$) as:

$$M(m) = N_V V(m) + N_C C(m)$$

where N_V and N_C refer to the number of events in the Voigt profile and the Chebychev polynomial respectively. During a χ^2-fit, the probability density of the composite model is scaled by the number of events in the data or MC sample to obtain the fit function. The χ^2-fit is performed with all parameters free-floating, except for the Z^0 decay width. These parameters and their starting values are shown in Figure 4.9. The Z^0 boson decay width is fixed to a value corresponding to the PDG average [32]. The fit begins with only N_V and N_C free-floating, with the other free parameters released iteratively and the fit re-performed after each iteration. The fitted distributions are shown in Figure 4.36.

The scale factor is then simply the ratio of the number of events in the Voigt profile in data and in MC:

$$R_{Z^0} = \frac{N_V^{(\text{data})}}{N_V^{(\text{MC})}}. \tag{4.7}$$

4.7.3 Systematic Uncertainties

The systematic uncertainty on R_{Z^0} due to the choice of fitting function is estimated by varying the fitting function. This systematic uncertainty is not calculated on a year-by-year basis, instead they use the full dataset. Three combinations of Voigt profiles and Crystal ball functions were investigated: Single Crystal Ball, Double Voigt, Crystal Ball + Voigt. The largest difference between alternative fitting configuration and the nominal is taken as the systematic uncertainty.
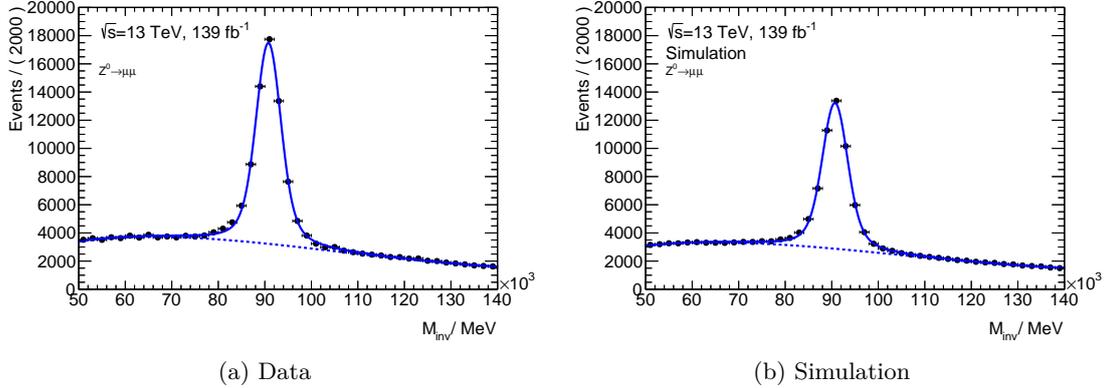
(a) Data

(b) Simulation

Figure 4.36: The invariant mass distribution in the $Z^0 \to \mu\mu$ control region used to extract the $Z^0(\to \mu\mu) + (b\text{-})$jets scale factor, in (a) data and (b) simulation. The solid line represents the sum of the fitted Voigt profile and 3rd-order Chebychev polynomial, whilst the dashed line represents the 3rd-order Chebychev polynomial only.

For the Double Voigt and Crystal Ball + Voigt, a composite p.d.f. is made with two Voigt profiles and a Crystal Ball and Voigt profile respectively. The fits are configured in the same way as the nominal fit, with a 3rd-order Chebychev polynomial describing all processes other than the Z^0 resonance. The parameters are progressively freed and the fit repeated as this happens.

A comparison of the resulting R_{Z^0} obtained using the different configurations is shown in Table 4.10.

4.7.4 Results

The normalisation scale factor obtained is

$$R_{Z^0} = 1.3602 \pm 0.0111(\text{stat}) \pm 0.0102(\text{syst}).$$

It is not concerning that the ratio is significantly away from unity, since the rate of Z^0 with additional heavy flavour jets is being measured, where the cross-section is not well known (though this scale factor should not be considered a cross-section measurement). A similar deviation is observed in the ATLAS $Z^0(\to \mu\mu) + b$-jets measurement [157] when accounting for the different jet momenta (see, for example, the red curve Figure 8 of Ref. [157]). The largest difference in R_{Z^0} in the different fitting configurations was with the double Voigt fit, hence this is applied as the systematic uncertainty.

The scale factor is also applied to the $Z^0(\to \tau\tau) + (b\text{-})$jets background, though with an uncertainty of 5%. This has very little impact on the uncertainty of $R(\tau/\mu)$, such that the analysis is not sensitive to the $Z^0(\to \tau\tau) + (b\text{-})$jets cross-section.
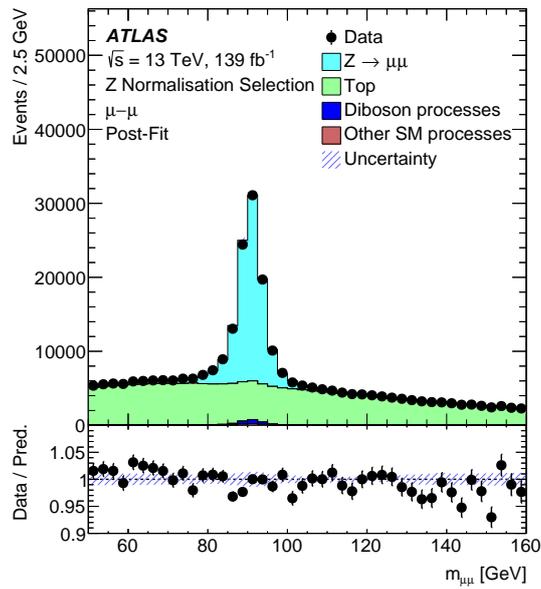
Figure 4.37: The invariant mass distribution in the $Z^0 \to \mu\mu$ control region used to extract the $Z^0(\to \mu\mu) + (b\text{-})$jets normalisation, with the calculated scale factor applied. The bottom panel shows the data/MC comparison. The blue bands indicate the systematic uncertainties with the constraints from the analysis fit of the signal region data applied. Good agreement is seen between data and prediction which gives confidence in the method used to calculate the scale factor.

Configuration	R_{Z^0}	|Difference|
Single Voigt (nominal)	$1.3602 \pm 0.0111(\text{stat})$	0
Single Crystal Ball	$1.3585 \pm 0.0153(\text{stat})$	0.0017
Double Voigt	$1.3704 \pm 0.0232(\text{stat})$	0.0102
Voigt + Crystal Ball	$1.3523 \pm 0.0535(\text{stat})$	0.0079

Table 4.10: The $Z^0(\to \mu\mu) + (b\text{-})$jets scale factors obtained using different fitting configurations. The largest differences are at the percent level which gives confidence in the fitting procedure. The largest difference compared to the nominal is seen with the double Voigt configuration, where the absolute difference is 0.00102. This is then the estimated systematic uncertainty on R_{Z^0}.

4.7.5 Cross-checks

This result is primarily validated in two ways. Firstly, the same method is applied to extract the $Z^0(\to ee) + (b\text{-})$jets scale factor. The fitted invariant mass distribution is shown in Figure 4.38. The scale factor obtained is

$$R_{Z^0}^{(e)} = 1.3074 \pm 0.0137(\text{stat}) \pm 0.00106(\text{syst}).$$

This is around 2σ away from the muon result. Secondly, data-MC agreement is assessed after the fit has been performed. This is shown in Figure 4.37. The agreement is shown to be very good.

A comparison of the resulting R_{Z^0} obtained using the different configurations is shown in 4.10. This is used to estimate the systematic uncertainty due to choice of fitting function.

The result is additionally validated individually in each year of data-taking, and separately based on the charge of the *probe* muon. Figure 4.40 compares the $Z^0(\to \mu\mu) + (b\text{-})$jets background scale factor in each of the data-taking periods against the nominal value. Figure 4.39 compares the background scale factor for the $Z^0(\to \mu\mu) + (b\text{-})$jets background separately according to the charge of the *probe* muon, against the inclusive value.
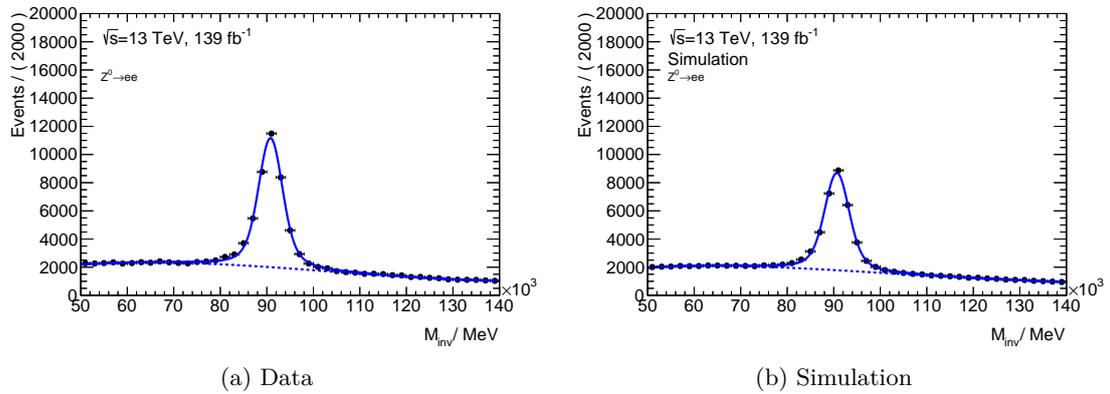
(a) Data

(b) Simulation

Figure 4.38: The invariant mass distribution in the $Z^0(\to ee) + (b\text{-})$jets control region used to extract the $Z^0(\to ee) + (b\text{-})$jets scale factor, in (a) data and (b) simulation. The solid line represents the sum of the fitted Voigt profile and 3rd-order Chebychev polynomial, whilst the dashed line represents the 3rd-order Chebychev polynomial only.
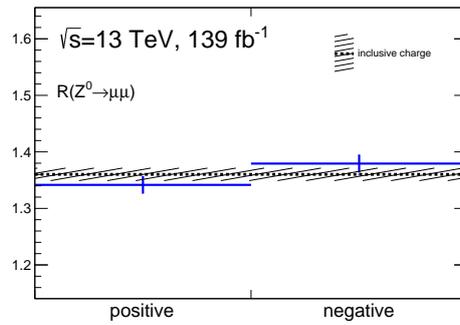
Figure 4.39: The $Z^0(\to \mu\mu) + (b\text{-})$jets background scale factors calculated exclusively in charge. Only statistical uncertainties are shown. The values obtained for both charges agree well within uncertainties.
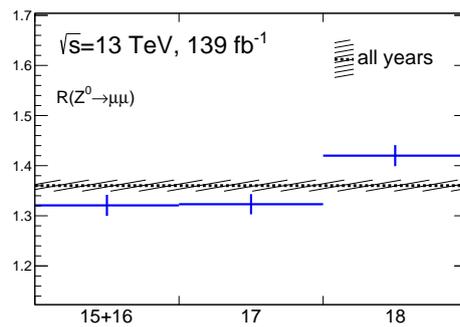
Figure 4.40: The $Z^0(\to \mu\mu) + (b\text{-})$jets background scale factors calculated in the different data-taking periods, compared against the inclusive value. Only statistical uncertainties are shown. Good agreement is seen for 2015+16 and 2017, though in 2018 the value is higher.

4.8 Fit procedure

In this analysis, the parameter of interest, $R(\tau/\mu)$, is extracted using a profile likelihood fit. ATLAS's TREXFITTER package is used for this.

4.8.1 Profile likelihood fitting

A profile likelihood fit is a statistically meaningful way of including systematic uncertainties in a maximum likelihood fit. A global likelihood function is constructed that includes all bins and fit parameters used in the measurement. The configuration of the fit is such that a negative-log-likelihood minimisation is performed with several unknown parameters, the parameter of interest ($R(\tau/\mu)$) and systematic uncertainties included as nuisance parameters θ. The likelihood function is defined in the standard way for a binned likelihood fit. Each bin in each measurement region is described by a Poisson distribution, and the likelihood function is simply the product across all bins, with a probability density function for systematics:

$$L(n, \theta^0 | R(\tau/\mu), \theta) = \prod_{i \in \text{bins}} P(n_i | R(\tau/\mu)(\theta)) \times \prod_{j \in \text{NPs}} G(\theta_j^0 | \theta_j)$$

$G(\theta_j^0 | \theta_j)$ is a Gaussian prior that requires continuous interpolation between variations and nominal templates.

The fit is set up with two floating parameters: $k(t\bar{t})$ and the parameter of interest $R(\tau/\mu)$. $k(t\bar{t})$ is the ratio of the normalisation of both the μ_{prompt} and $\mu_{\tau(\to\mu)}$ components of the $t\bar{t}$ and Wt processes to the total predicted events (using the theoretical cross sections). $R(\tau/\mu)$ only affects the $\mu_{\tau(\to\mu)}$ components. It therefore controls the relative contributions of the μ_{prompt} and $\mu_{\tau(\to\mu)}$ templates – it is the ratio of the two.

The normalisation scale factor for the muons from hadron decays background is not a free floating parameter, rather it is treated as a nuisance parameter. It has a nominal value and 1σ uncertainty as derived in Section 4.6.

The normalisation scale factor for the $Z^0(\to \mu\mu) + (b\text{-})$jets background is not a free floating parameter, rather it is treated as a nuisance parameter. It has a nominal value and 1σ uncertainty as derived in Section 4.7.

The other backgrounds are much smaller, so do not require data-driven normalisation scale factors. They are instead normalised to their higher order cross-section and treated as nuisance parameters with 1σ uncertainties given by the uncertainties on these higher order cross sections.

Additional sources of uncertainty arising from shape and acceptance effects are also treated as nuisance parameters in the fit.

4.8.2   Fit configuration

The fit uses three p_T^μ bins: $p_T^\mu = $ 5-10 GeV, $p_T^\mu = $ 10-20 GeV and $p_T^\mu = $ 20-250 GeV. Each p_T^μ bin is divided into eight $|d_0^\mu|$ bins, with boundaries $|d_0^\mu| = [0, 0.01, 0.02, 0.03, 0.04, 0.06, 0.09, 0.15, 0.5]$ mm. There are two channels (e-μ and μ-μ), therefore the 2-D fit has $3 \times 8 \times 2 = 48$ bins in total.

The three p_T^μ regions were selected to have significantly different compositions of μ_{prompt}, $\mu_{\tau(\to\mu)}$ and μ_{had}. The low and middle p_T^μ bins are most sensitive to the $\mu_{\tau(\to\mu)}$ component whilst the highest bin is most sensitive to the μ_{prompt} component. The lowest bin has the highest background contribution from muons from hadron decays, whilst the highest bin has almost none.

The eight $|d_0^\mu|$ bins were also optimised to have varying compositions, but also to ensure it is easy to get good convergence in the fit. Finer binning could improve sensitivity to $R(\tau/\mu)$, however it can overcomplicate the fit and make it harder to get good convergence.

4.8.3   Blinding procedure

The analysis was developed blind to the central value of parameter of interest $R(\tau/\mu)$. The uncertainty on $R(\tau/\mu)$ was not blinded. The analysis procedure, including all inputs to the fit such as the calibration and background normalisation scale factors, were all finalised before $R(\tau/\mu)$ was unblinded. The fit was developed using Asimov data before being finalised using real data but keeping $R(\tau/\mu)$ blinded.

After unblinding, the fit result was validated by performing the fit in separate regions: separately in each of the e-μ and μ-μ channels, in the charge of the *probe* muon, and in the data taking period. This ensures that the result is robust.

4.9   Systematic Uncertainties

This section describes the systematic uncertainties with the largest impact on $R(\tau/\mu)$. Many of the systematic uncertainties are correlated between the $\mu_{\tau(\to\mu)}$ and μ_{prompt} templates. This means that they mostly cancel in the calculation of $R(\tau/\mu)$. These include uncertainties due to jet reconstruction, flavour tagging and trigger efficiencies. The remaining uncertainties are due to the transverse impact parameter calibration, background scale factors, modelling and reconstruction. The ones with significant impact are described below.

4.9.1 Transverse impact parameter calibration uncertainties

The templates for prompt muons include the calibrated $|d_0^\mu|$ distributions, described in Section 4.4.4. There is an uncertainty on these templates due to the fact they are derived using a $Z^0 \to \mu\mu$ calibration region, but the measurement uses the $t\bar{t}$ signal region. This means that there are differences in the hadronic environment around the *probe* muons used for the calibration and measurement respectively. The full shape information might not be fully encapsulated in the templates due to coarse binning in p_T^μ and η. This means that small biases are present in the template distributions. The size of the bias can be estimated from the full difference between the $|d_0^\mu|$ templates from Z^0 and $t\bar{t}$ in simulation. The uncertainty is then split into a "core" component corresponding to the region $|d_0^\mu| < 0.05$ mm, and a "tail" component corresponding to the region $|d_0^\mu| > 0.05$ mm. This prevents the data from over constraining the uncertainty using the full $|d_0^\mu|$ distribution.

The templates for non-prompt ($\mu_{\tau(\to\mu)}$ and μ_{had}) muons include the calibrated resolution, described in Section 4.4.4. This method is not statistically limited. The differences between data and MC are understood to arise from the beamspot size at high p_T^μ, as well as material, alignment and modelling at low p_T^μ. An uncertainty is applied as half the full correction symmetrised. This is around five times the size of the statistical uncertainty of the correction and is therefore considered highly conservative. This uncertainty has little impact on $R(\tau/\mu)$, therefore a less conservative uncertainty is not currently motivated.

The uncertainty on the $|d_0^\mu|$ calibration for prompt muons accounts for around has a large impact on $R(\tau/\mu)$ (around 35% of the total, excluding correlations), whilst the uncertainty on the $|d_0^\mu|$ resolution calibration for non-prompt muons has less impact (around 16% of the total, excluding correlations).

4.9.2 Background scale factor uncertainties

There are systematic uncertainties due to the calculation of the background scale factors. The estimation of these uncertainties are described for muons from hadron decays in Section 4.6.3 and for $Z^0(\to \mu\mu) + (b\text{-})$jets in Section 4.7.3. The $Z^0(\to \mu\mu) + (b\text{-})$jets scale factor is also applied to the $Z^0(\to \tau\tau) + (b\text{-})$jets background, with a 5% uncertainty.

The uncertainty due to the muons from hadron decays background scale factors has more impact on $R(\tau/\mu)$ than the uncertainty due to the $Z^0(\to \mu\mu) + (b\text{-})$jets scale factor. This is mainly due to the fact that the μ_{had} background dominates at larger $|d_0^\mu|$ values, occupying the same phase space as the $\mu_{\tau(\to\mu)}$ contribution. The $Z^0(\to \mu\mu) + (b\text{-})$jets background, however, dominates at lower values of $|d_0^\mu|$, in the same phase space as the very large μ_{prompt} contribution.

There are also differences due to the methods used to calculate the scale factors. The control region used to calculate $Z^0(\to \mu\mu) + (b\text{-})$jets scale factor has high statistics and the resonance peak in the invariant mass distribution is well described by a Voigt profile.

Uncertainty	Variation
Initial- and final-state radiation	A14 eigen-tune variations [126] of the strong coupling constant (α_s)
Missing higher-order QCD corrections	Factorisation and renormalisation up or down by a factor of 2 (multiplied or divided)
Resummation scale uncertainty	POWHEG h_{damp} parameter varied from 1.5 to 3 m_{top}
Parton shower and hadronisation model	HERWIG v7.04 [104, 105], H7UE tune [105], `MMHT2014LO` PDF set [153]
Top p_{T} spectrum	Symmetrised effect of removing the NNLO in QCD NLO in EW in top p_{T} reweighting.

Table 4.11: Different generator configurations used to estimate systematic uncertainties due to modelling.

Therefore, the number of events in the peak can be precisely determined. The muons from hadron decays background scale factors are dependent on the $t\bar{t}$ modelling and have fewer statistics, meaning there is a greater uncertainty on the number and distribution of events used to calculate the scale factors.

The muons from hadron decays scale factor uncertainty has a large impact on $R(\tau/\mu)$ (around 26% of the total, excluding correlations) and the $Z^0(\to \mu\mu)+(b\text{-})$jets scale factor uncertainties have less impact (around 8%, excluding correlations).

4.9.3 Modelling uncertainties

Uncertainties due to modelling of MC simulated samples are derived. The most significant of these are due to the $t\bar{t}$ modelling, which affect both signal (μ_{prompt} and $\mu_{\tau(\to\mu)}$) and background (μ_{had}) contributions. The combined yields of the signal contributions are allowed to float in the fit, but the p_{T}^{μ} and $|d_0^{\mu}|$ distributions depend on the generator setup, meaning different generators lead to different yields in each bin of the templates for μ_{prompt} and $\mu_{\tau(\to\mu)}$ contributions. The templates for muons from hadron decays also have an uncertainty due to changes in the muon p_{T}^{μ} modelling, and relative fractions of muon from different sources. These in turn affect the $|d_0^{\mu}|$ distribution. In all of these cases, the uncertainties are estimated using several different generator variations. These are described in Table 4.11. The effects on both of the signal contributions are considered correlated, but the effects on muons from hadron decays are treated separately. The parton shower and hadronisation uncertainty is separated into four nuisance parameters in the fit, corresponding to low, medium and high probe-muon p_{T}^{μ}, with the high p_{T}^{μ} split into normalisation and shape components.

As mentioned in Section 4.3.2, the $t\bar{t}$ MC samples are reweighted to NNLO in QCD and NLO in EW in p_T of one of the top quarks, in order to improve modelling of the lepton p_T spectrum. The systematic uncertainty due to this method is estimated as the symmetrised difference between applying and not applying the correction. This uncertainty has little impact on $R(\tau/\mu)$.

Other background processes not mentioned in this section are normalised to their higher order cross-section and treated as nuisance parameters with 1σ uncertainties given by the uncertainties on these higher order cross sections.

The uncertainty due to the choice of parton distribution function (PDF) is estimated by following the recommendation procedure. The measurement is repeated with 30 different variations of PDF which are defined by the PDF4LHC15 set [127], and the quadrature sum of $R(\tau/\mu)$ is determined for the different PDFs. The resulting variation of $R(\tau/\mu)$ is 0.04%.

The $t\bar{t}$ modelling uncertainties have a very large impact $R(\tau/\mu)$. The parton shower variations are the dominant modelling uncertainties (around 33% of total, excluding correlations).

4.9.4 Instrumentation uncertainties

The measurement of $R(\tau/\mu)$ using tau to muon decays relies on precise muon identification and reconstruction. This means that the uncertainties on muon efficiency corrections have significant impact on $R(\tau/\mu)$. The corrections are applied as p_T dependent scale factors that correct MC to data. This means that they affect μ_{prompt} and $\mu_{\tau(\to\mu)}$ components differently, which has an effect on the measurement of $R(\tau/\mu)$. The muon efficiencies are determined in dimuon ($Z^0 \to \mu\mu$ and $J/\psi \to \mu\mu$) data and MC using a tag and probe method [86]. The p_T dependent scale factors are applied to MC to account for differences in efficiency between data and MC. The uncertainty on these scale factors are included in the analysis. The uncertainty on muon isolation has the third-highest impact $R(\tau/\mu)$ (around 30% of total, excluding correlations) and the uncertainty on low p_T muon identification scale factors has the fourth-highest impact (around 28% of total, excluding correlations).

The uncertainties due to pile-up modelling also have an impact on $R(\tau/\mu)$. These are estimated by reweighting MC samples to change the amount of pile-up. The impact on $R(\tau/\mu)$ is mostly due to the residual effect on p_T modelling. The impact on $R(\tau/\mu)$ due to pile-up modelling is around 16% of total, excluding correlations.
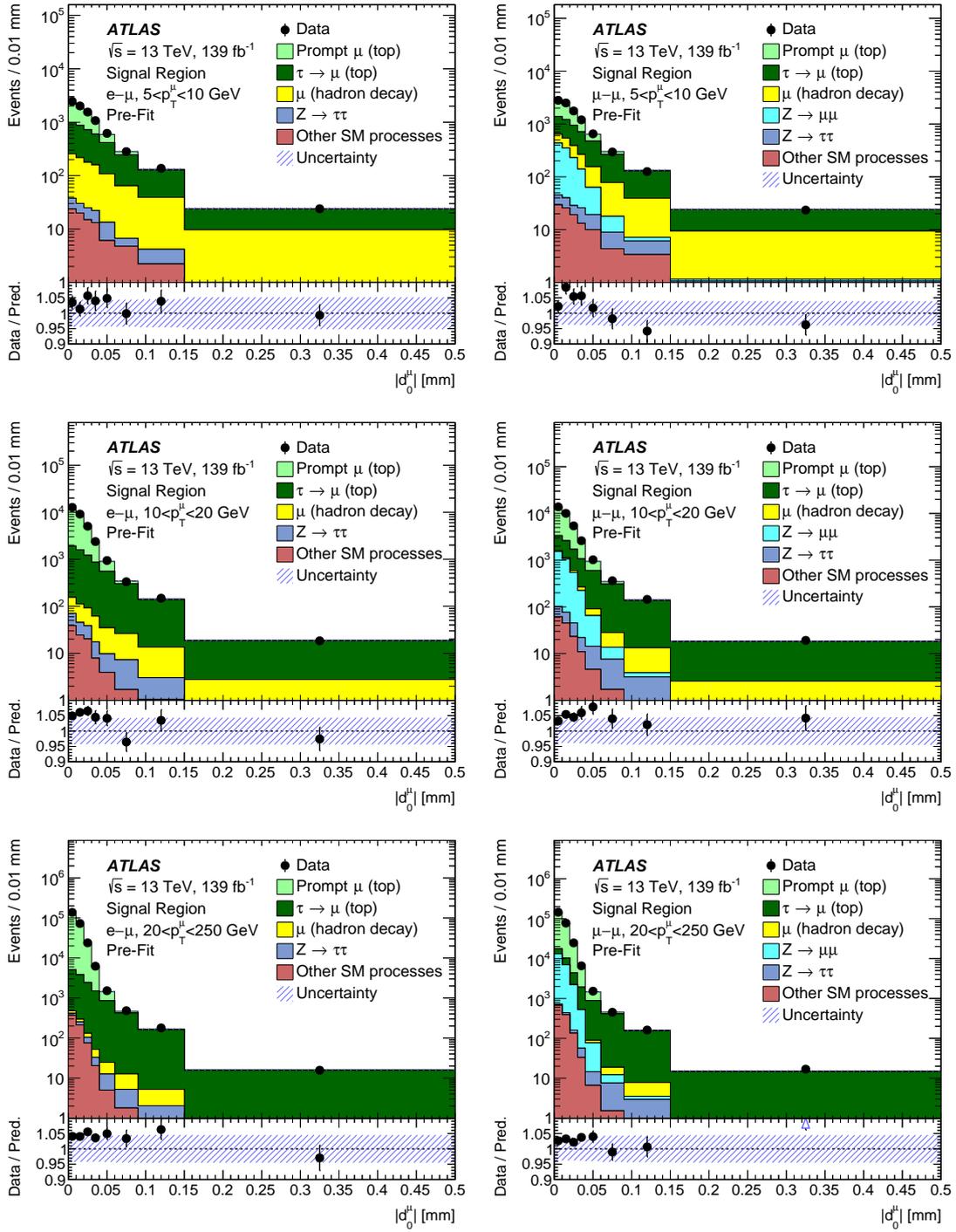
Figure 4.41: Pre-fit $|d_0^\mu|$ distributions for data and MC. The background scale factors have already been applied. The bottom panel shows the ratio of the data to the expectation, with the blue bands representing the uncertainty. The $t\bar{t}$ normalisation uncertainty is not included as this is a free-floating parameter in the fit. Data and prediction are seen to agree within the uncertainties.
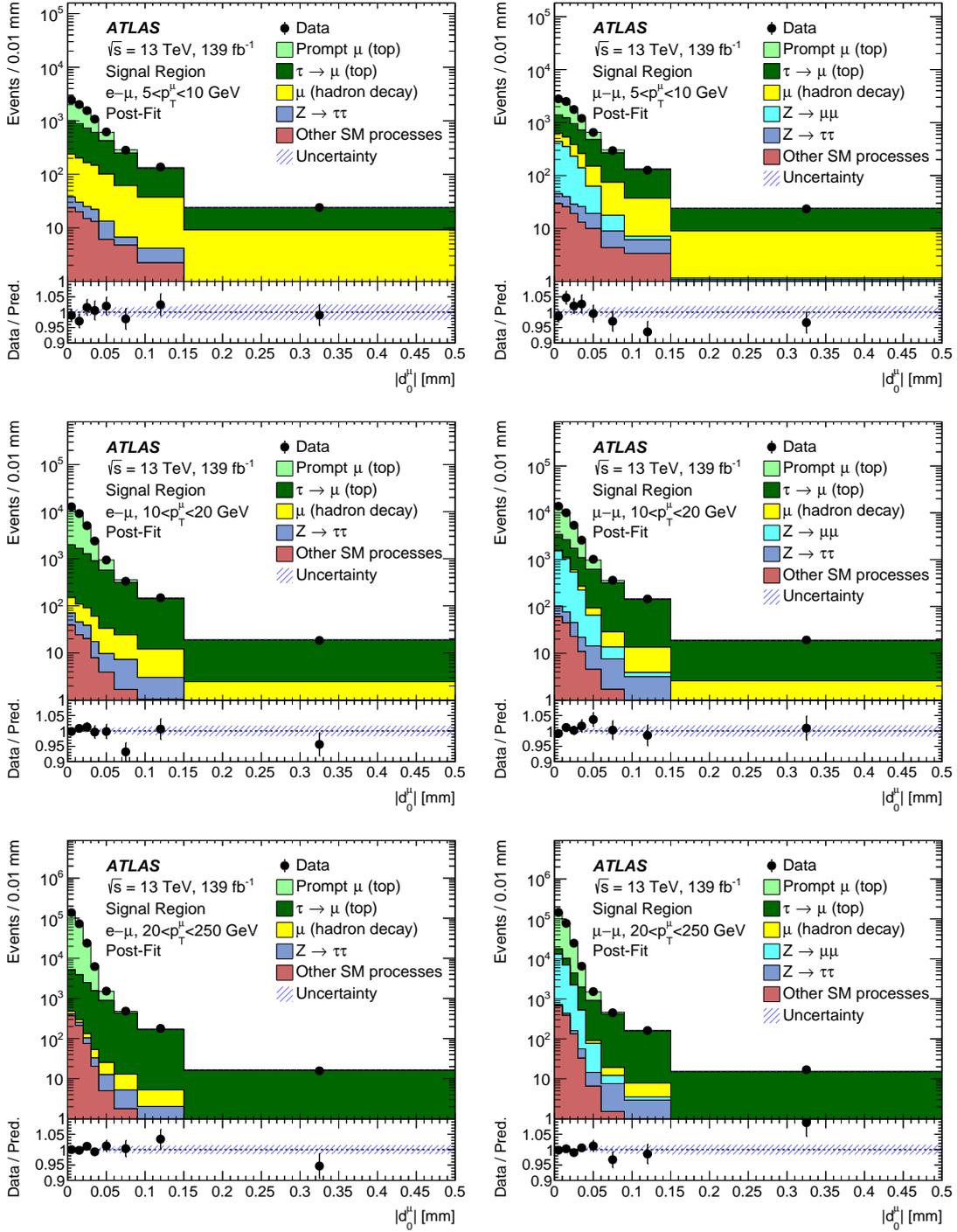
Figure 4.42: Post-fit $|d_0^\mu|$ distributions for data and MC. The bottom panel shows the ratio of the data to the expectation, with the blue bands representing the uncertainty. After the fit, the agreement between prediction and data is seen to improve.

4.10   Results

Figures 4.41 and 4.42 show the pre- and post-fit $|d_0^\mu|$ distributions respectively, in the six signal regions for data and MC. The separation between the μ_{prompt}, $\mu_{\tau(\to\mu)}$ and μ_{had} components can be clearly seen in the different p_T^μ regions. The μ_{prompt} contributions dominate at low values of $|d_0^\mu|$ whilst the $\mu_{\tau(\to\mu)}$ contributions dominate at high $|d_0^\mu|$. There are also significant contributions from muons from hadron decays at high $|d_0^\mu|$, however these mainly contribute only at low p_T^μ. The pre-fit distributions already show good agreement between data and MC, and this is improved after the fit.

The quality of the fit of expected distributions in MC to observed distributions in data can be quantified with a goodness of fit metric. Here, this is defined as twice the negative log-likelihood relative to a fit performed assuming the pre-fit expectation per degree of freedom. Averaged across all bins, this has a value of 1.11, which corresponds to a p-value of 0.29.

Biasing of the result was minimised by following the blinding procedure described in Section 4.8. The robustness of the result was ensured by checking that the extracted value of $R(\tau/\mu)$ is consistent across different sub-regions of the signal region. The different sub-regions are the channel (e-μ or μ-μ), the different kinematic bins, different data-taking periods and the charge of the *probe* muon. In the cases of the different channels, data-taking periods and charge of the *probe* muon, the background scale factors are calculated for these specific sub-regions. The comparisons to the nominal result are shown in Section 4.10.1.

The total systematic uncertainty on $R(\tau/\mu)$ is 0.0109. This includes the uncertainty on the branching ratio $\text{Br}(\tau \to \mu\nu\nu)$. The statistical uncertainty arising from the size of the dataset is 0.007. A categorised breakdown of the different contributions to the total uncertainty is shown in Table 4.12. The leading impacts come from imperfect knowledge of the tail of the $|d_0^\mu|$ distributions, the parton shower and hadronisation model uncertainty and the muon selection uncertainties. This is shown in Figure 4.43.

The best-fit observed value of the parameter of interest, $R(\tau/\mu)$, is:

$$R(\tau/\mu) = 0.992 \pm 0.013[\pm 0.007(\text{stat}) \pm 0.011(\text{syst})],$$

which is the most precise measurement of this parameter to date. This result is in very good agreement with the Standard Model assumption of equal couplings for different lepton flavours and provides good evidence for the hypothesis of Lepton Flavour Universality. The result is shown in comparison to the previous LEP result in Figure 4.44. The present result differs from the LEP result, which had shown some tension with the Standard Model expectation, with agreement at the level of 2.7 standard deviations.

Source	Impact on $R(\tau/\mu)$	Impact as %
Prompt $|d_0^\mu|$ templates	0.0038	35%
$\mu_{(\text{prompt})}$ and $\mu_{(\tau\to\mu)}$ parton shower variations	0.0036	33%
Muon isolation efficiency	0.0033	30%
Muon identification and reconstruction	0.0030	28%
$\mu_{(\text{had.})}$ normalisation	0.0028	26%
$t\bar{t}$ scale and matching variations	0.0027	25%
Top p_{T} spectrum variation	0.0026	24%
$\mu_{(\text{had.})}$ parton shower variations	0.0021	19%
Monte Carlo statistics	0.0018	17%
Pile-up	0.0017	16%
$\mu_{(\tau\to\mu)}$ and $\mu_{(\tau\to\mu)}$ $|d_0^\mu|$ shape	0.0017	16%
Other detector systematic uncertainties	0.0016	15%
$Z^0(\to\mu\mu) + (b\text{-})$jets normalisation	0.009	8%
Other sources	0.0004	4%
$\text{Br}(\tau\to\mu\nu_\tau\nu_\mu)$	0.0023	21%
Total systematic uncertainty	0.0109	–
Data statistics	0.0072	n/a
Total	**0.013**	–

Table 4.12: A list of the different sources of uncertainty. The impact as a percentage is calculated as the impact as a fraction of the total systematic uncertainty, which will sum to more than 100% due to correlations.
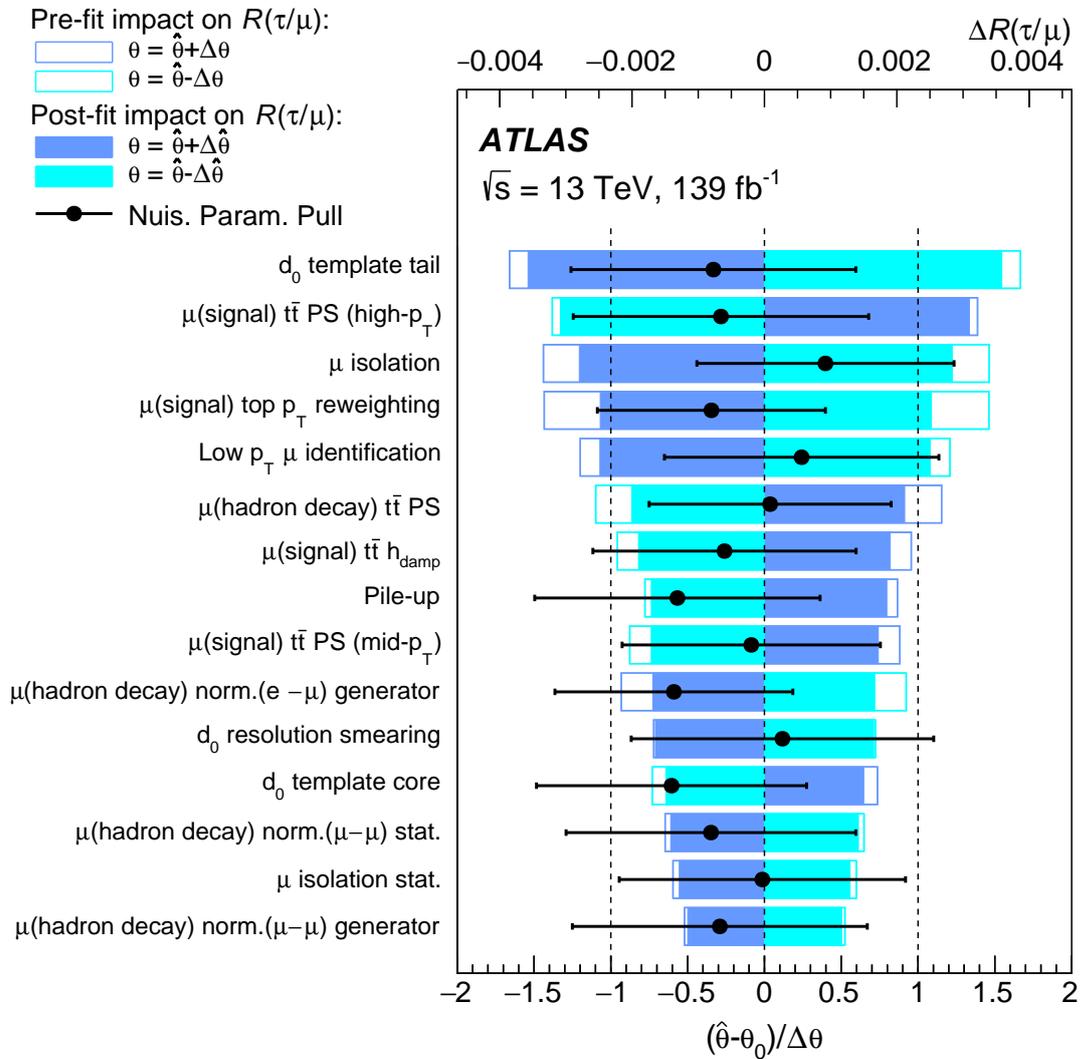
Figure 4.43: The ranked impacts of the different systematic uncertainties on the parameter of interest, $R(\tau/\mu)$.
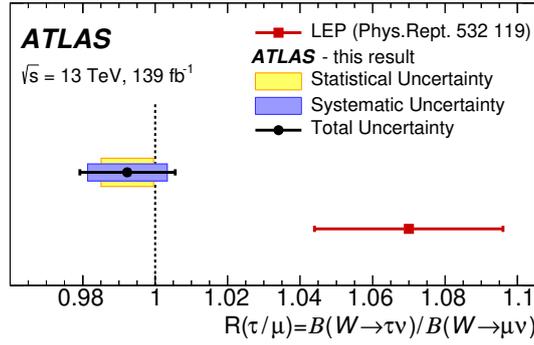
Figure 4.44: Measured value of $R(\tau/\mu)$, shown in comparison to the previous LEP result. The result obtained in this analysis shows good agreement with the Standard Model assumption of unity.

4.10.1 Cross checks

The robustness of the result was ensured by calculating $R(\tau/\mu)$ independently in several sub-regions of the signal region. Figure 4.45 shows the values of $R(\tau/\mu)$ calculated in each of the data-taking periods. Figure 4.46 shows the values of $R(\tau/\mu)$ when the profile likelihood fit is performed independently for negative and positive *probe* muons. The values of $R(\tau/\mu)$ calculated individually in the two channels is shown in Figure 4.47.
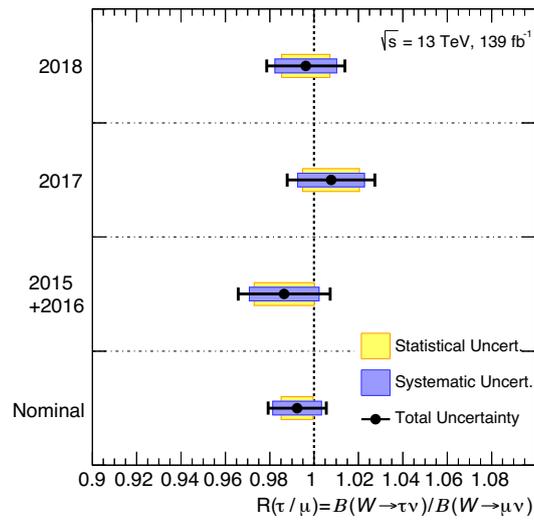
Figure 4.45: The nominal result (bottom) compared with the independent fits using only data from 2015+2016, 2017 and 2018 respectively. Good agreement is seen for the values obtained in each of the data-taking periods.
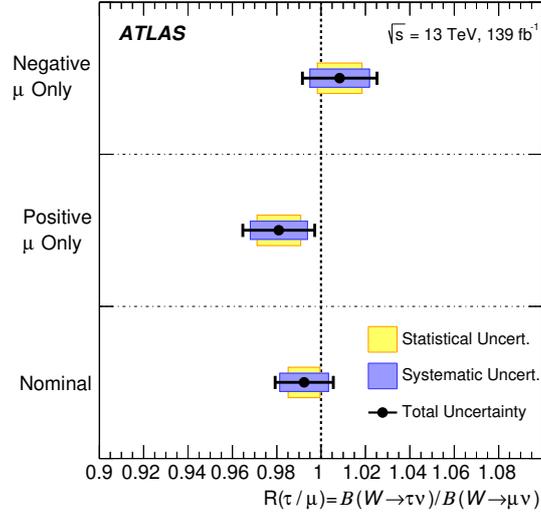
Figure 4.46: The nominal result (bottom) compared with the independent fits using only negatively (top) and positively (bottom) charged *probe* muons. The theory uncertainties are computed using the charge-inclusive sample, and the $|d_0^\mu|$ templates are also formed inclusive of charge. Good agreement is seen.
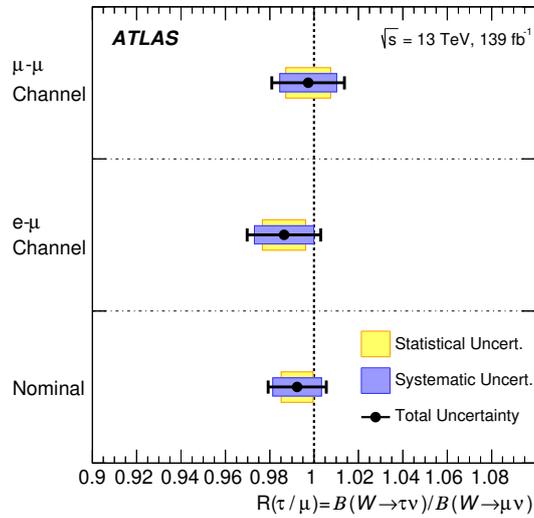
Figure 4.47: The nominal result (bottom) compared with the independent fits using only the μ-μ channel (top) and e-μ channel (middle). A single fit is performed using the same correlation model as the nominal fit, but using two parameters of interest for $R(\tau/\mu)$, one for each of the e-μ and μ-μ channels. Good agreement is seen between both channels.

4.11   Conclusions

The conclusions are split into two sections. The first presents the conclusions of the analysis as a whole, and the second more specifically reiterates the conclusions of the author's research and findings.

4.11.1   Conclusions of the analysis as a whole

Lepton Flavour Universality refers to the Standard Model assumption of universal lepton couplings to the vector bosons. This is a notable feature of the Standard Model, since leptons could also experience individual couplings. It is tested by comparing the decay rates of (semi-)leptonic processes that differ only in lepton flavour. A measurement deviating from this identity would be unambiguous evidence of new physics.

A measurement of

$$R(\tau/\mu) = \mathrm{Br}(W^\pm \to \tau\nu)/\mathrm{Br}(W^\pm \to \mu\nu)$$

has been presented here. The measurement was performed in the dileptonic decay modes of $t\bar{t}$ events, from a dataset corresponding to an integrated luminosity of 139 fb^{-1} of pp collision data recorded with the ATLAS detector at the LHC. This analysis provides the most precise measurement of $R(\tau/\mu)$ to date, which tests the fundamental assumption of the universality of the couplings of taus and muons to the $W^\pm$ boson in the Standard Model.

The best-fit observed value of $R(\tau/\mu)$ is:

$$R(\tau/\mu) = 0.992 \pm 0.013[\pm 0.007(\mathrm{stat}) \pm 0.011(\mathrm{syst})].$$

This agrees well with the Standard Model prediction.

This result reaches an unprecedented precision, surpassing the previous LEP result by a factor of two. It also sheds new light the long-standing 2.7σ deviation observed by LEP described in Section 1.3.3. This is notable due to the fact that such precision measurements have long been considered impossible in the complex environment of hadron collisions. This result is complementary and timely in the context of the recent observations of possible non-universal lepton couplings in b-hadron decays, such as those described in Section 1.3.2. The result is shown in comparison to previous tests of Lepton Flavour Universality using $W^\pm$ boson decays in Figure 4.48.

The previous LEP result showed some tension with the Standard Model assumption of universal couplings of taus and muons to $W^\pm$ bosons, whilst the result of this analysis does not. The previous LEP result is derived from the calculation of the total $W^\pm$-pair production cross section measurement using the full LEP dataset which comprises many different centre of mass energies, using many more channels (including hadronic tau decays), as well as using event selections from four different experiments. Furthermore,

138
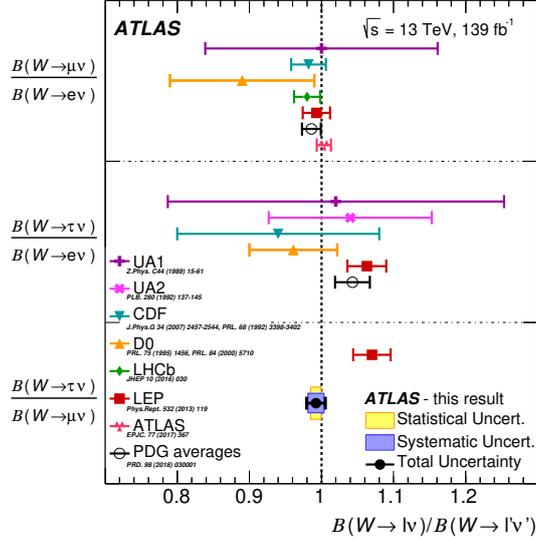
Figure 4.48: Measured value of $R(\tau/\mu)$, shown in comparison to previous tests of Lepton Flavour Universality using $W^{\pm}$ boson decays. The precision of the measured values has improved over time.

it is much more statistically limited. Taking all this into account, understanding why the two results differ in agreement with the Standard Model assumption is purely speculative. In the first instance, it could be a statistical fluctuation. A difference to expectation with a significance of 2.7σ would correspond to a probability of slightly less than 1% for such an occurrence. Alternatively, it might not be impossible that hadronic reconstruction uncertainties are underestimated.

Possibilities for future study include the test of LFU between taus and electrons. This was not included in this round of the analysis due to the fact that electrons require a different treatment compared to muons. For example, new $|d_0^e|$ calibrations would be required. Additionally, electrons are much more susceptible to misidentified charge. This means that the method for deriving the non-prompt background scale factors would either require modification or an entirely new method would be used. Including electrons would also allow for the combined test of taus against both light leptons $(2\mathrm{Br}(W^{\pm} \to \tau\nu)/(\mathrm{Br}(W^{\pm} \to \mu\nu) + \mathrm{Br}(W^{\pm} \to e\nu)))$. Another possibility for further study would be changing the event selection to include events with one b-tagged jet, which would increase sensitivity. This region sees increased background contributions (see, for example, Ref. [56]), so constraining these backgrounds would be particularly important here.

As described in Section 2.10, ATLAS is preparing for Run-3 and, later, the High Luminosity LHC (HL-LHC). This analysis would benefit from the additional statistics in Run-3, with improvements to the Muon Spectrometer ensuring good muon identification and reconstruction in spite of the higher event rate [158]. The planned ITk for HL-LHC may impact the resolution of low p_{T} muon tracks, since the innermost layer

of the proposed ITk will be slightly further away than the existing IBL, however the resolution of higher p_T tracks will be improved due to larger longitudinal pixel size to be used in the ITk [10]. Other than this, tracking performance should either be similar or improved compared to what has been seen in Run-2, in spite of the massive pile-up, with the benefit of significantly higher statistics.

4.11.2 Conclusions of the author's research and findings

Selection optimisation

The selection optimisation described in Section 4.5 resulted in modification of the selection criteria to exclude *probe* muons with momentum balance significance greater than two. This was to constrain muons produced in the decays in flight of pions and kaons by around 40% and 50% respectively. This was decided after it was shown that the cut improved the expected precision on $R(\tau/\mu)$ by a small amount.

Given more time, a more elaborate study could be performed. This could include more parameters useful in discriminating low p_T signal and background muons, in particular those described in Ref. [159]. These parameters could be used in a multi-variate analysis that also includes standard discrimination variables in order to identify background muons, for example. A similar, but separate, study could be performed for probe electrons if these are included in the analysis. The studies could be performed in different bins according to the number of b-tagged jets should the 1-b-tag region be included.

Muons from hadron decays scale factors

The calculation of the muons from hadron decays background scale factors described in Section 4.6 resulted in scale factors of $1.39 \pm 4.32\%(\text{stat}) \pm 7.66\%(\text{gen}) \pm 0.96\%$ (background subtraction) and $1.37 \pm 3.65\%(\text{stat}) \pm 3.02\%(\text{gen}) \pm 1.32\%$ (background subtraction) for e-μ and μ-μ respectively. It is not concerning that these scale factors deviate significantly from one. This is due to the fact that the topology of the muons from heavy flavour decays that pass the tight isolation criteria applied here is somewhat complicated, and as such hard to simulate. Therefore a scale factor far from one is not unexpected. The difference in size of the generator uncertainty can be attributed to the fact that the e-μ channel includes a contribution from electrons with misidentified charge. This is not modelled consistently across different generator configurations. The uncertainty on these scale factors have a relatively large impact on the measured value of $R(\tau/\mu)$ (26% of total, excluding correlations). This could be improved by reducing the background contributions in the signal region, as per the selection optimisation studies described above.

This method would require some modification if pursuing some of the potential future analysis goals described in the above section. In the case of including probe electrons

in the analysis, this method would have to be modified to include charge misidentified probe electrons, which occur at a much higher rate compared to muons. The 1-b-tag jet regions include higher background contributions. Therefore, the method could be modified to calculate scale factors in different bins according to the number of b-tagged jets.

Z background scale factors

The calculation of the $Z^0(\rightarrow \mu\mu)+(b\text{-})$jets scale factor described in Section 4.7 resulted in a scale factor of $1.3602\pm0.0111(\text{stat})\pm0.0102(\text{syst})$. The scale factor is significantly larger than one, however this is not concerning. This is due to the overall yield of $Z^0(\rightarrow \mu\mu)+(b\text{-})$jets. A similar discrepancy has been observed in Ref. [157]. The uncertainty on this scale factor does not have significant impact on $R(\tau/\mu)$. Whilst the $Z^0(\rightarrow \mu\mu)+(b\text{-})$jets background is statistically larger than the muons from hadron decays background, the former dominates at low $|d_0^\mu|$, in the same phase space as the numerous prompt muons. Muons from hadron decays dominate at low p_T^μ and high $|d_0^\mu|$, in the same phase space as muons from tau decays.

The $Z^0(\rightarrow ee)+(b\text{-})$jets scale factor is already calculated alongside the $Z^0(\rightarrow \mu\mu)+(b\text{-})$jets scale factor, where it provides a valuable cross-check. Therefore, the method would not need to be modified to include probe electrons. The scale factors could be calculated in different bins according to the number of b-tagged jets, should the 1-b-tag jet region be studied. The scale factors are expected to be different in the different bins.

5

Summary

This section provides a summary of the thesis with a focus on the author's research and findings. This is divided into two parts: first, a summary of improvements to vertex reconstruction in ATLAS, second, a summary of the test of Lepton Flavour Universality using leptonic tau decays in top quark decays.

The ATLAS detector at the LHC was described in Chapter 2. Both the collider and detector are formidable pieces of equipment forged through international collaboration. Without the great teamwork that has built the LHC and ATLAS, as well as the software framework to analyse the impressive dataset, the author's contributions described in this thesis would not be possible. The author has contributed to this effort by making improvements to the primary vertex reconstruction strategy used in ATLAS. This is described in Chapter 3. The main focus of this work was the Adaptive Multi-Vertex Finder (AMVF), with some additional focus on the Gaussian density seed finder. The author's research and findings worked on optimising the AMVF and seed finder setup, including identifying and fixing issues in the logic of the vertex fitting strategy. This involved modification of the simulated annealing setup, introduction of a significance-based track assignment cut, as well as a variable longitudinal constraint for the vertex fit. Performance gains were observed in almost all physics metrics used to quantify performance, which are especially significant at the high pile-up densities expected in Run-3 and beyond.

As introduced in Chapter 1, Lepton Flavor Universality (LFU) refers to the Standard Model assumption of universal lepton couplings to the vector gauge bosons. This is a notable feature of the Standard Model, since the different generations of leptons could experience individual, or non-universal, couplings. The assumption is experimentally verified to high precision across a wide range of momentum transfer, in both neutral current (involving a Z^0 boson) and charged current (involving a $W^{\pm}$ boson) interactions. The notable exceptions include the previous LEP averages of $\mathrm{Br}(W^{\pm} \to \ell\nu)$ ($\ell = \tau, \mu, \nu$), where tau-muon (tau-light-lepton) universality was in agreement with the Standard Model at the 2.7σ (2.6σ) level, as well as the recent so-called "flavour anomalies" seen in b-factory

experiments. These tests of LFU neither conclusively agree nor disagree with the Standard Model prediction and are therefore possible hints of new physics. LFU could be an accidental low-energy effect, with the different generations of leptons exhibiting different properties at higher energies, as indicated by their different masses. Thus, further study is well motivated.

If the central value of the LEP result were reproduced, it would need a measurement with a precision of less than two percent to confirm that it points to new physics. As demonstrated in Chapter 4, such a measurement has been possible in ATLAS, using the $W^{\pm}$ bosons produced in the numerous $t\bar{t}$ decays recorded there. The ratio of branching ratios

$$R(\tau/\mu) = \mathrm{Br}(W^{\pm} \to \tau\nu)/\mathrm{Br}(W^{\pm} \to \mu\nu)$$

was measured in order to test the assumption of Lepton Flavour Universality between taus and muons. The measurement was performed in dileptonic decay modes of $t\bar{t}$ events, using a dataset corresponding to 139 fb^{-1} of pp collision data recorded with the ATLAS detector at the LHC during Run-2.

The author contributed to this result with his own research and findings. The conclusions of these contributions are summarised in more detail in Section 4.11.2. This includes the event selection optimisation described in Section 4.5. Here, the author investigated improving discrimination between signal and background by modifying the event selection with additional cuts on discrimination variables. This resulted in the inclusion of a cut requiring the Momentum Balance Significance to be less than two. This cut reduces the number of muons from hadron decays produced in the decays-in-flight of pions and kaons by 40% and 50% respectively.

Another contribution is the calculation of background scale factors. The first set of these is described in Section 4.6, which scale the muons from hadron decays background. The calculation of these scale factors make use of a same-sign signal region in each of the e-μ and μ-μ channels. These control regions contain a relatively pure sample of muons from hadron decays, at a similar rate as observed in the signal region. Systematic uncertainties on these method were estimated by varying the MC generator configurations, by modifying the p_{T}^{μ} cut used in the prompt-background subtraction, and from the statistical size of the MC dataset. The robustness of the method was demonstrated through numerous cross-checks. The calculation of the scale factor for the $Z^0(\to \mu\mu) + (b\text{-})$jets background is described in Section 4.7. This makes use of a $Z^0(\to \mu\mu) + (b\text{-})$jets control region, which has the same selection criteria as the signal region but with the Z^0 veto cuts dropped. The scale factors are calculated by fitting a Voigt profile to the Z^0 resonance in the invariant mass distribution in both MC and data. The number of events in the Z^0 peak are extracted and used to scale the contributions in MC to what is observed in data. Systematics are estimated by using alternate functions to fit the Z^0 peak and comparing the difference in the central value of the scale factor.

The result of the analysis as a whole, the best-fit observed value of $R(\tau/\mu)$, is:

$$R(\tau/\mu) = 0.992 \pm 0.013[\pm 0.007(\mathrm{stat}) \pm 0.011(\mathrm{syst})].$$

which agrees well with the Standard Model prediction of unity. This result reaches an unprecedented precision, which is notable being produced by a hadron collider.

Bibliography

[1] ATLAS collaboration, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, Tech. Rep. ATL-PHYS-PUB-2019-015, CERN, Geneva (Apr, 2019).

[2] I. Sanderswood (on behalf of the ATLAS collaboration), *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, in *Connecting the Dots and Workshop on Intelligent Trackers*, 10, 2019 [1910.08405].

[3] ATLAS collaboration, *Test of the universality of τ and μ lepton couplings in W-boson decays from $t\bar{t}$ events with the ATLAS detector*, *Preprint (accepted by Nature Physics)* (2020) [2007.14040].

[4] I. Sanderswood (on behalf of the ATLAS collaboration), *Test of Tau and Muon Lepton Universality in W boson decays from $t\bar{t}$ in ATLAS*, in *19th International Conference on B-Physics at Frontier Machines*, Oct, 2020.

[5] HFLAV collaboration, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, *Preprint* (2019) [1909.12524].

[6] E. Mobs, *The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019*, Jul, 2019.

[7] ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[8] ATLAS collaboration, *Performance of the ATLAS Trigger System in 2015*, *Eur. Phys. J. C* **77** (2017) 317 [1611.09661].

[9] T. Gleisberg et al, *Event generation with SHERPA 1.1*, *JHEP* **0902** (2009) 007 [0811.4622].

[10] ATLAS collaboration, *Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC*, Tech. Rep. ATL-PHYS-PUB-2019-014, CERN, Geneva (Mar, 2019).

[11] A. Pich, *Precision Tau Physics*, *Prog. Part. Nucl. Phys.* **75** (2014) 41 [`1310.7922`].

[12] N. Kidonakis, *Theoretical results for electroweak-boson and single-top production*, *PoS* **DIS2015** (2015) 170 [`1506.04072`].

[13] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.

[14] ATLAS collaboration, *Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions*, *Eur. Phys. J. C* **70** (2010) 723 [`0912.2642`].

[15] ATLAS collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, *Eur. Phys. J. C* **70** (2010) 1193 [`1007.5423`].

[16] A. Artamonov, D. Bailey, G. Belanger, M. Cadabeschi, T.Y. Chen, V. Epshteyn et al., *The ATLAS forward calorimeter*, *Journal of Instrumentation* **3** (2008) .

[17] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [`1207.7214`].

[18] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [`1207.7235`].

[19] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) 2626 [`hep-ex/9503002`].

[20] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) 2632 [`hep-ex/9503003`].

[21] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s}$ = 540-GeV*, *Phys. Lett. B* **122** (1983) 103.

[22] UA2 collaboration, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, *Phys. Lett. B* **122** (1983) 476.

[23] UA1 collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider*, *Phys. Lett. B* **126** (1983) 398.

[24] UA2 collaboration, *Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ Collider*, *Phys. Lett. B* **129** (1983) 130.

[25] J. Greensite, *An introduction to the confinement problem*, vol. 821 of *Lecture Notes in Physics*, Springer (2011), 10.1007/978-3-642-14382-3.

[26] S.L. Glashow, J. Iliopoulos and L. Maiani, *Weak interactions with lepton-hadron symmetry*, *Phys. Rev. D* **2** (1970) 1285–1292.

[27] C. Rovelli, *Notes for a brief history of quantum gravity*, in *Recent developments in theoretical and experimental general relativity, gravitation and relativistic field theories. Proceedings, 9th Marcel Grossmann Meeting, MG'9, Rome, Italy, July 2-8, 2000. Pts. A-C*, p. 742–768, 6, 2000 [`gr-qc/0006061`].

[28] P. Peebles and B. Ratra, *The Cosmological Constant and Dark Energy*, *Rev. Mod. Phys.* **75** (2003) 559–606 [`astro-ph/0207347`].

[29] V. Trimble, *Existence and Nature of Dark Matter in the Universe*, *Ann. Rev. Astron. Astrophys.* **25** (1987) 425–472.

[30] C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes and R.P. Hudson, *Experimental test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) 1413–1415.

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

[32] Particle Data Group collaboration, *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001.

[33] J. Christenson, J. Cronin, V. Fitch and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, *Phys. Rev. Lett.* **13** (1964) 138–140.

[34] G. Isidori, *Flavour Physics: Old problems, recent hopes and challenges*, in *CERN Academic Training Lectures*, May, 2019, https://cds.cern.ch/record/2676021.

[35] CHARM collaboration, *Test of the Universality of the ν_e and ν_μ Coupling to the Charged Weak Current*, *Phys. Lett. B* **179** (1986) 301.

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

[37] LHCb collaboration, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, *JHEP* **08** (2017) 055 [`1705.05802`].

[38] HPQCD collaboration, *Standard Model Predictions for $B \to K\ell^+\ell^-$ with Form Factors from Lattice QCD*, *Phys. Rev. Lett.* **111** (2013) 162002 [`1306.0434`].

[39] LHCb collaboration, *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **113** (2014) 151601 [`1406.6482`].

[40] BaBar collaboration, *Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays*, *Phys. Rev. Lett.* **109** (2012) 101802 [`1205.5442`].

[41] BaBar collaboration, *Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons*, *Phys. Rev. D* **88** (2013) 072012 [`1303.0571`].

[42] BELLE collaboration, *Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle, Phys. Rev. D* **92** (2015) 072014 [1507.03233].

[43] LHCB collaboration, *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$, Phys. Rev. Lett.* **115** (2015) 111803 [1506.08614].

[44] BELLE collaboration, *Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$, Phys. Rev. Lett.* **118** (2017) 211801 [1612.00529].

[45] BELLE collaboration, *Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle, Phys. Rev. D* **97** (2018) 012004 [1709.00129].

[46] LHCB collaboration, *Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong τ-lepton decays, Phys. Rev. Lett.* **120** (2018) 171802 [1708.08856].

[47] LHCB collaboration, *Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays, Phys. Rev. D* **97** (2018) 072013 [1711.02505].

[48] BELLE collaboration, *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method, Preprint* (2019) [1904.08794].

[49] BABAR collaboration, *Precision measurement of the $\mathcal{B}(\Upsilon(3S) \to \tau^+\tau^-)/\mathcal{B}(\Upsilon(3S) \to \mu^+\mu^-)$ ratio, Phys. Rev. Lett.* **125** (2020) 241801 [2005.01230].

[50] ALEPH, DELPHI, L3, OPAL, LEP ELECTROWEAK collaboration, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP, Phys. Rept.* **532** (2013) 119 [1302.3415].

[51] ATLAS collaboration, *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 [1612.03016].

[52] H.D. Politzer, *Asymptotic Freedom: An Approach to Strong Interactions, Phys. Rept.* **14** (1974) 129.

[53] H.D. Politzer, *Reliable Perturbative Results for Strong Interactions?, Phys. Rev. Lett.* **30** (1973) 1346.

[54] D.J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories, Phys. Rev. Lett.* **30** (1973) 1343.

[55] W. Beenakker, A. Denner, W. Hollik, R. Mertig, T. Sack and D. Wackeroth, *Electroweak one loop contributions to top pair production in hadron colliders, Nucl. Phys. B* **411** (1994) 343.

[56] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in eμ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 528 [1910.08819].

[57] CMS collaboration, *Measurement of the $t\bar{t}$ production cross section, the top quark mass, and strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. **C79** (2019) 368 [1812.10505].

[58] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612 [1111.5869].

[59] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001 [1204.5201].

[60] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054 [1207.0236].

[61] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080 [1210.6832].

[62] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004 [1303.6254].

[63] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, J. Mazzitelli and H. Sargsyan, *Top-quark pair hadroproduction at next-to-next-to-leading order in QCD*, Phys. Rev. **D99** (2019) 051501 [1901.04005].

[64] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930 [1112.5675].

[65] ATLAS collaboration, *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **04** (2017) 086 [1609.03920].

[66] CMS collaboration, *Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **800** (2020) 135042 [1812.10514].

[67] ATLAS collaboration, *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS*, JHEP **01** (2018) 063 [1612.07231].

[68] CMS collaboration, *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **10** (2018) 117 [1805.07399].

[69] CDF, D0 collaboration, *Observation of s-channel production of single top quarks at the Tevatron*, *Phys. Rev. Lett.* **112** (2014) 231803 [`1402.5126`].

[70] J. Aguilar-Saavedra, *Top flavor-changing neutral interactions: Theoretical expectations and experimental detection*, *Acta Phys. Polon. B* **35** (2004) 2695 [`hep-ph/0409342`].

[71] O. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole et al., *LHC Design Report Vol.1: The LHC Main Ring*, CERN Yellow Reports: Monographs, CERN, Geneva (2004), 10.5170/CERN-2004-003-V-1.

[72] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole et al., *LHC Design Report Vol. 2: The LHC infrastructure and general services*, CERN Yellow Reports: Monographs, CERN, Geneva (2004), 10.5170/CERN-2004-003-V-2.

[73] M. Benedikt, P. Collier, V. Mertens, J. Poole and K. Schindl, *LHC Design Report Vol. 3: The LHC injector chain*, CERN Yellow Reports: Monographs, CERN, Geneva (2004), 10.5170/CERN-2004-003-V-3.

[74] L. Arnaudon et al., *Linac4 Technical Design Report*, Tech. Rep. CERN-AB-2006-084. CARE-Note-2006-022-HIPPI, CERN, Geneva (Dec, 2006).

[75] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[76] LHCв collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[77] ALICE collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[78] ATLAS collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19 (Sep, 2010).

[79] ATLAS collaboration, *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*, *Nucl. Instrum. Meth. A* **845** (2017) 257.

[80] ATLAS collaboration, *ATLAS computing: Technical design report*, Tech. Rep. CERN-LHCC-2005-022, ATLAS-TRD-017 (Jun, 2005).

[81] P. Calafiura, W. Lavrijsen, C. Leggett, M. Marino and D. Quarrie, *The Athena control framework in production, new developments and lessons learned*, in *14th International Conference on Computing in High-Energy and Nuclear Physics*, pp. 456–458, 2005.

[82] G. Barrand et al., *GAUDI - A software architecture and framework for building HEP data processing applications*, *Comput. Phys. Commun.* **140** (2001) 45.

[83] *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, Tech. Rep. ATL-PHYS-PUB-2015-018, CERN, Geneva (Jul, 2015).

[84] *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, Tech. Rep. ATL-PHYS-PUB-2015-026, CERN, Geneva (Jul, 2015).

[85] ATLAS collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14** (2019) P12006 [1908.00005].

[86] ATLAS collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} =13$ TeV*, Eur. Phys. J. C **76** (2016) 292 [1603.05598].

[87] ATLAS collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Preprint (2020) [2007.02645].

[88] ATLAS collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903 [1802.08168].

[89] A. Buckley, T. Eifert, M. Elsing, D. Gillberg, K. Koeneke, A. Krasznahorkay et al., *Implementation of the ATLAS Run 2 event data model*, J. Phys. Conf. Ser. **664** (2015) 072045.

[90] ATLAS collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823 [1005.4568].

[91] R. Jansky, *The ATLAS Fast Monte Carlo Production Chain Project*, J. Phys. Conf. Ser. **664** (2015) 072024.

[92] R. Bruce, N. Fuster-Martínez, A. Mereghetti, D. Mirarchi and S. Redaelli, *Review of LHC Run 2 Machine Configurations*, in *9th LHC Operations Evian Workshop*, (Geneva, Switzerland), pp. 187–197, 2019.

[93] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[94] S. Abdel Khalek et al., *The ALFA Roman Pot Detectors of ATLAS*, JINST **11** (2016) P11013 [1609.00249].

[95] ATLAS collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, Tech. Rep. ATLAS-CONF-2019-021, CERN, Geneva (Jun, 2019).

[96] G. Apollinari, I. Béjar, Alonso, O. Brüning, P. Fessia, M. Lamont, L. Rossi et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, Geneva (2017), 10.23731/CYRM-2017-004.

[97] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070 [0709.2092].

[98] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079 [1405.0301].

[99] S. Catani, F. Krauss, R. Kuhn and B.R. Webber, *QCD matrix elements + parton showers*, JHEP **11** (2001) 063 [hep-ph/0109231].

[100] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **0803** (2008) 038 [0709.1027].

[101] S. Höche, F. Krauss, M. Schönherr and F. S., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027 [1207.5030].

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

[103] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159 [1410.3012].

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

[105] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196 [1512.01178].

[106] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[107] G. Piacquadio, K. Prokofiev and A. Wildauer, *Primary vertex reconstruction in the ATLAS experiment at LHC*, J. Phys. Conf. Ser. **119** (2008) 032033.

[108] G. Piacquadio and K. Jakobs, *Identification of b-jets and investigation of the discovery potential of a Higgs boson in the $WH \to l\nu b\bar{b}$ channel with the ATLAS experiment*, Ph.D. thesis, 2010.

[109] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040 [hep-ph/0409146].

[110] P.Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018 [1005.3457].

[111] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145 [1406.3660].

[112] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P.M. Nadolsky, J. Pumplin et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024.

[113] ATLAS collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, Tech. Rep. ATL-PHYS-PUB-2016-017, CERN, Geneva (2016).

[114] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) 332 [`1611.10235`].

[115] R. Fruhwirth, W. Waltenberger and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys.* **G34** (2007) N343.

[116] ATLAS collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, Tech. Rep. ATLAS-CONF-2012-042, CERN, Geneva (2012).

[117] W. Waltenberger, *Development of Vertex Finding and Vertex Fitting Algorithms for CMS*, Ph.D. thesis, Vienna, Tech. U., 2004.

[118] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse and A. Salzburger, *The new ATLAS track reconstruction (NEWT)*, *Journal of Physics: Conference Series* **119** (2008) 032014.

[119] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, Tech. Rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, Geneva (Sep, 2017).

[120] S. Frixione, P. Nason and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126 [`0707.3088`].

[121] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [`1002.2581`].

[122] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [`1410.8849`].

[123] ATLAS collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, Tech. Rep. ATL-PHYS-PUB-2016-020, CERN, Geneva (Sep, 2016).

[124] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015 [`1212.3460`].

[125] S. Frixione, E. Laenen, P. Motylinski and B.R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, *JHEP* **04** (2007) 081 [`hep-ph/0702198`].

[126] *ATLAS Pythia 8 tunes to 7 TeV datas*, Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva (Nov, 2014).

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

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

[129] A.D. Martin, W. Stirling, R. Thorne and G. Watt, *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, *Eur. Phys. J. C* **64** (2009) 653 [0905.3531].

[130] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024 [1007.2241].

[131] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li et al., *The CT10 NNLO Global Analysis of QCD*, *Phys. Rev. D* **89** (2014) 033009 [1302.6246].

[132] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [1207.1303].

[133] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos and M. Zaro, *Top-pair production at the LHC through NNLO QCD and NLO EW*, *JHEP* **10** (2017) 186 [1705.04105].

[134] S. Frixione, E. Laenen, P. Motylinski, B.R. Webber and C.D. White, *Single-top hadroproduction in association with a W boson*, *JHEP* **0807** (2008) 029 [0805.3067].

[135] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer and M. Wiedermann, *HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*, *Comput. Phys. Commun.* **182** (2011) 1034 [1007.1327].

[136] P. Kant, O.M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, *Comput. Phys. Commun.* **191** (2015) 74 [1406.4403].

[137] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **0812** (2008) 039 [0808.3674].

[138] S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049 [1111.1220].

[139] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **0905** (2009) 053 [0903.1219].

[140] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601 [1111.5206].

[141] A. Denner, S. Dittmaier and L. Hofer, *Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations, Comput. Phys. Commun.* **212** (2017) 220 [1604.06792].

[142] ATLAS collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector, Eur. Phys. J. C* **76** (2016) 581 [1510.03823].

[143] ATLAS collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, Eur. Phys. J. C* **77** (2017) 490 [1603.02934].

[144] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm, JHEP* **04** (2008) 063 [0802.1189].

[145] ATLAS collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Phys. Rev. D* **96** (2017) 072002 [1703.09665].

[146] ATLAS collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV, JHEP* **08** (2018) 089 [1805.01845].

[147] ATLAS collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C* **79** (2019) 970 [1907.05120].

[148] ATLAS collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Tech. Rep. ATLAS-CONF-2020-030, CERN, Geneva (Aug, 2020).

[149] ATLAS collaboration, *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP* **07** (2020) 124 [2002.07546].

[150] ATLAS collaboration, *Performance of the ATLAS muon triggers in Run 2, JINST* **15** (2020) P09015 [2004.13447].

[151] ATLAS collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2, Eur. Phys. J. C* **80** (2020) 47 [1909.00761].

[152] ATLAS collaboration, *Alignment of the ATLAS Inner Detector in Run-2, Eur. Phys. J. C* **80** (2020) 1194 [2007.07624].

[153] L. Harland-Lang, A. Martin, P. Motylinski and R. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs, Eur. Phys. J. C* **75** (2015) 204 [1412.3989].

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

[155] W. Voigt, *Über das gesetz der intensitätsverteilung innerhalb der linien eines gasspektrums*, 1912.

155

[156] W. Verkerke and D.P. Kirkby, *The RooFit toolkit for data modeling*, in *13th International Conference for Computing in High-Energy and Nuclear Physics (CHEP 2003)*, 2003 [`physics/0306116`].

[157] ATLAS collaboration, *Measurements of the production cross-section for a Z boson in association with b-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **07** (2020) 044 [`2003.11960`].

[158] T. Kawamoto, S. Vlachos, L. Pontecorvo, J. Dubbert, G. Mikenberg, P. Iengo et al., *New Small Wheel Technical Design Report*, Tech. Rep. CERN-LHCC-2013-006. ATLAS-TDR-020 (Jun, 2013).

[159] ATLAS collaboration, *Identification of very-low transverse momentum muons in the ATLAS experiment*, Tech. Rep. ATL-PHYS-PUB-2020-002, CERN, Geneva (Feb, 2020).

Appendix A

List of Monte Carlo Samples

A comprehensive list of the MC samples used in this analysis is shown in Tables A.1 and A.2. The tables list the simulation type with "FS" referring to "full simulation" and "AF" referring to "ATLAS fast" simulation. In all cases the samples come from the MC16 MC production campaign. The samples use the relevant sub-campaigns for the data-taking period, with MC16a for the 2015+2016 sample, MC16d for 2017 sample and MC16e for 2018 sample. The ATLAS metadata interface p-tags are specified.

Table A.1: Top MC Samples

Sample	DSID	p-tag	Simulation
$t\bar{t}$ (non all-hadronic) Powheg + Pythia8	410470	3832	FS
$t\bar{t}$ (di-leptons) Powheg + Pythia8	410472	3832	FS
$t\bar{t}$ (non all-hadronic) Powheg + Pythia8	410470	3832	AF
$t\bar{t}$ aMC@NLO	410464, 410465	3832	AF
$t\bar{t}$ Powheg + Herwig	410457, 410458	3832	AF
$t\bar{t}$ Powheg + Pythia8 hdamp	410480, 410482	3832	AF
$t\bar{t} + W$ aMC@NLO	410155	3832	FS
$t\bar{t} + Z(\nu\nu)$ aMC@NLO	410156	3832	FS
$t\bar{t} + Z(q\bar{q})$ aMC@NLO	410157	3832	FS
$t\bar{t} + l\bar{l}$ aMC@NLO	410218–410220	3832	FS
Wt inclusive Powheg + Pythia8	410646,410647	3832	FS
Wt di-leptons Powheg + Pythia8	410648,410649	3832	FS
Wt di-leptons Powheg + Pythia8	410648,410649	3832	AF
Wt di-leptons Powheg + Herwig	411038,411039	3832	AF
Wt di-leptons aMC@NLO	412003	3832	AF
Wt inclusive Powheg + Pythia8 - DS	410654,410655	3832	FS
Wt di-leptons Powheg + Pythia8 - DS	410656,410656	3832	FS

Table A.2: Background MC Samples

Sample	DSID	p-tag	Simulation
W + jets Sherpa	364156 – 364197	3629	FS
l^+l^- + jets Sherpa	364100 – 364141	3629	FS
l^+l^- + jets (low mass) Sherpa	364198 – 364215	3629	FS
$ZZ \to q\bar{q}l\bar{l}$ Sherpa	363356	3629	FS
$WZ \to q\bar{q}l\bar{l}$ Sherpa	363358	3629	FS
$WW \to q\bar{q}l\nu$ Sherpa	363359,363360	3629	FS
$ZW \to q\bar{q}l\nu$ Sherpa	363489	3629	FS
$VV \to 4\,l$ Sherpa	364250	3629	FS
$VV \to 3\,l + \nu$ Sherpa	364253	3629	FS
$VV \to 2\,l + \nu\nu$ Sherpa	364254	3629	FS
Single t, s-channel Powheg + Pythia8	410644,410645	3832	FS
Single t, t-channel Powheg + Pythia8	410658,410659	3832	FS

Appendix B

Background scale factor cross checks

The robustness of the best-fit observed value of $R(\tau/\mu)$ is ensured by performing the profile likelihood fit in separate sub-regions of the signal region (separated according to data-taking period, charge of the *probe* muon and *e*-μ and μ-μ channels). In each case, the background scale factors (described in Sections 4.6 and 4.7) are calculated for the specific sub-region. The estimation of the systematic uncertainties on these values are not recalculated, with the nominal values deemed representative. This appendix contains the values used for estimated background scale factors, as well as the calculated scale factors, for the muons from hadron decays and $Z^0(\to \mu\mu) + (b\text{-})$jets backgrounds. The comparison plots of both the scale factors and $R(\tau/\mu)$ in the separate sub-regions compared to the nominal value are shown in the main text. The appendix begins with the separate data-taking periods, then the separate charges and finally the separate (*e*-μ,μ-μ) channels.

B.1   Separate data taking periods

B.1.1   2015+2016 samples only

The background scale factors were calculated for the 2015 and 2016 sample only. Table B.1 shows the values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, Table B.2 shows the values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, and Table B.3 shows the fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective data/MC ratios.

tag-probe	e-μ	μ-μ
$N_{MC}^{\mathrm{SS,prompt}}$	199.63 ± 4.15	73.82 ± 1.35
$N_{\mathrm{data}}^{\mathrm{SS},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	183.00 ± 13.53	73.00 ± 8.54
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	5.17 ± 0.81	5.36 ± 0.80
$N_{MC}^{\mathrm{SS,prompt},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	128.41 ± 3.40	46.44 ± 1.01
S	1.44 ± 0.12	1.54 ± 0.20
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	286.53 ± 23.86	113.97 ± 14.82

Table B.1: Values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, for the 2015 and 2016 sample.

tag-probe	e-μ	μ-μ
$N_{\mathrm{data}}^{\mathrm{SS}}$	702.00 ± 26.50	671.00 ± 23.90
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{SS}}$	312.07 ± 9.07	325.52 ± 8.16
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	466.84 ± 8.68	469.23 ± 8.04
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	286.53 ± 23.86	113.97 ± 14.82
R	1.33 ± 0.12	1.40 ± 0.09
$N_{\mathrm{data}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	621.53 ± 57.49	658.80 ± 45.20

Table B.2: Values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, for the 2015 and 2016 sample.

Sample	Ratio	Relative Uncertainty	N_Z (data)	N_Z (simulation)
$Z^0 \to ee$	1.3079 ± 0.0258	1.98%	9472 ± 154	7242 ± 81
$Z^0 \to \mu\mu$	1.3208 ± 0.0210	1.59%	15268 ± 197	11560 ± 107

Table B.3: Fitted numbers of $Z^0(\to ee)+(b\text{-})$jets and $Z^0(\to \mu\mu)+(b\text{-})$jets in data and simulation, along with their respective ratios for the 2015 and 2016 sample.

B.1.2   2017 samples only

The background scale factors were calculated for the 2017 sample only. Table B.4 shows the values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, Table B.5 shows the values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, and Table B.6 shows the fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective data/MC ratios.

tag-probe	e-μ	μ-μ
$N_{MC}^{\mathrm{SS,prompt}}$	252.91 ± 4.92	89.59 ± 1.60
$N_{\mathrm{data}}^{\mathrm{SS},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	230.00 ± 15.17	106.00 ± 10.30
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	8.06 ± 1.05	8.11 ± 1.08
$N_{MC}^{\mathrm{SS,prompt},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	161.04 ± 3.97	54.17 ± 1.08
S	1.38 ± 0.10	1.81 ± 0.19
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	348.56 ± 26.26	161.90 ± 17.70

Table B.4: Values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, for the 2017 sample.

tag-probe	e-μ	μ-μ
$N_{\mathrm{data}}^{\mathrm{SS}}$	879.00 ± 29.65	689.00 ± 26.25
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{SS}}$	392.39 ± 9.22	388.11 ± 9.10
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	582.05 ± 11.45	540.33 ± 10.28
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	348.56 ± 26.26	161.90 ± 17.70
R	1.35 ± 0.11	1.36 ± 0.09
$N_{\mathrm{data}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	786.83 ± 63.50	733.84 ± 49.33

Table B.5: Values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, for the 2017 sample.

Sample	Ratio	Relative Uncertainty	N_Z (data)	N_Z (simulation)
$Z^0 \to ee$	1.3348 ± 0.0249	1.87%	11163 ± 168	8363 ± 92
$Z^0 \to \mu\mu$	1.3231 ± 0.0201	1.52%	17772 ± 214	13432 ± 124

Table B.6: Fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective ratios for 2017 sample.

B.1.3   2018 samples only

The background scale factors were calculated for the 2018 sample only. Table B.7 shows the values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, Table B.8 shows the values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, and Table B.9 shows the fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective data/MC ratios.

tag-probe	e-μ	μ-μ
$N_{MC}^{\mathrm{SS,prompt}}$	324.23 ± 5.41	120.75 ± 1.86
$N_{\mathrm{data}}^{\mathrm{SS},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	253.00 ± 15.91	130.00 ± 11.40
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	8.15 ± 1.02	11.80 ± 1.32
$N_{MC}^{\mathrm{SS,prompt},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	204.63 ± 4.32	73.79 ± 1.42
S	1.20 ± 0.08	1.60 ± 0.16
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	387.96 ± 27.33	193.42 ± 19.38

Table B.7: Values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, for the 2018 sample.

tag-probe	e-μ	μ-μ
$N_{\text{data}}^{\text{SS}}$	1118.00 ± 33.44	874.00 ± 29.56
$N_{\text{MC}}^{\mu_{\text{had}},\text{SS}}$	503.01 ± 10.07	504.25 ± 9.06
$N_{\text{MC}}^{\mu_{\text{had}},\text{OS}}$	718.77 ± 10.29	727.05 ± 10.63
$N_{\text{MC}}^{\text{prompt,SS,scaled}}$	387.96 ± 27.33	193.42 ± 19.38
R	1.45 ± 0.09	1.35 ± 0.07
$N_{\text{data}}^{\mu_{\text{had}},\text{OS}}$	1043.18 ± 66.83	981.29 ± 55.81

Table B.8: Values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, for the 2018 sample.

Sample	Ratio	Relative Uncertainty	N_Z (data)	N_Z (simulation)
$Z^0 \rightarrow ee$	1.2837 ± 0.0210	1.64%	14458 ± 191	11283 ± 108
$Z^0 \rightarrow \mu\mu$	1.4200 ± 0.0210	1.21%	23474 ± 247	16532 ± 97

Table B.9: Fitted numbers of $Z^0(\rightarrow ee) + (b\text{-})$jets and $Z^0(\rightarrow \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective ratios for 2018 sample.

B.2 Separate electric charge

B.2.1 Positive *probe* muons only

The background scale factors were calculated using only positive *probe* muons. Table B.10 shows the values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, Table B.11 shows the values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, and Table B.12 shows the fitted numbers of $Z^0(\rightarrow ee) + (b\text{-})$jets and $Z^0(\rightarrow \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective data/MC ratios.

tag-probe	e-μ	μ-μ
$N_{MC}^{\mathrm{SS,prompt}}$	421.91 ± 6.03	169.11 ± 2.05
$N_{\mathrm{data}}^{\mathrm{SS},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	386.00 ± 19.65	200.00 ± 14.14
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	9.20 ± 1.09	12.61 ± 1.36
$N_{MC}^{\mathrm{SS,prompt},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	268.61 ± 4.84	102.35 ± 1.50
S	1.40 ± 0.08	1.83 ± 0.14
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	591.86 ± 33.77	309.62 ± 24.20

Table B.10: Values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, using only positive *probe* muons.

tag-probe	e-μ	μ-μ
$N_{\mathrm{data}}^{\mathrm{SS}}$	1395.00 ± 37.35	1126.00 ± 33.56
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{SS}}$	609.06 ± 11.48	627.27 ± 12.05
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	877.17 ± 12.39	866.13 ± 11.91
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	591.86 ± 33.77	309.62 ± 24.20
R	1.32 ± 0.09	1.30 ± 0.07
$N_{\mathrm{data}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	1156.86 ± 77.47	1127.25 ± 63.03

Table B.11: Values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, using only positive *probe* muons.

Sample	Ratio	Relative Uncertainty	N_Z (data)	N_Z (simulation)
$Z^0 \to ee$	1.3050 ± 0.0194	1.48%	17505 ± 211	13413 ± 116
$Z^0 \to \mu\mu$	1.3416 ± 0.0156	1.16%	27916 ± 267	20808 ± 135

Table B.12: Fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective ratios using only positive *probe* muons.

B.2.2   Negative *probe* muons only

The background scale factors were calculated using only negative *probe* muons. Table B.13 shows the values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, Table B.14 shows the values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, and Table B.15 shows the fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective data/MC ratios.

tag-probe	e-μ	μ-μ
$N_{MC}^{\mathrm{SS,prompt}}$	354.86 ± 5.87	115.04 ± 1.90
$N_{\mathrm{data}}^{\mathrm{SS},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	280.00 ± 16.73	109.00 ± 10.44
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	12.18 ± 1.27	12.65 ± 1.30
$N_{MC}^{\mathrm{SS,prompt},p_{\mathrm{T}}^{\mu}>30\mathrm{GeV}}$	220.96 ± 4.65	69.42 ± 1.41
S	1.21 ± 0.08	1.39 ± 0.15
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	430.12 ± 29.30	159.66 ± 17.93

Table B.13: Values used for estimation of prompt scale factor, with the resulting scale factor and scaled number of events, using only negative *probe* muons.

tag-probe	e-μ	μ-μ
$N_{\mathrm{data}}^{\mathrm{SS}}$	1304.00 ± 36.11	1008.00 ± 31.75
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{SS}}$	598.41 ± 11.70	590.61 ± 9.29
$N_{\mathrm{MC}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	890.50 ± 12.60	870.48 ± 11.90
$N_{\mathrm{MC}}^{\mathrm{prompt,SS,scaled}}$	591.86 ± 33.77	309.62 ± 24.20
R	1.46 ± 0.08	1.44 ± 0.07
$N_{\mathrm{data}}^{\mu_{\mathrm{had}},\mathrm{OS}}$	1300.42 ± 75.99	1250.33 ± 59.72

Table B.14: Values used for estimation of the muons from hadron decays background scale factors, along with the calculated scale factors and scaled number of events, for only negative *probe* muons.

Sample	Ratio	Relative Uncertainty	N_Z (data)	N_Z (simulation)
$Z^0 \to ee$	1.3105 ± 0.0193	1.47%	17595 ± 210	13427 ± 116
$Z^0 \to \mu\mu$	1.3793 ± 0.0159	1.16%	28601 ± 272	20736 ± 135

Table B.15: Fitted numbers of $Z^0(\to ee) + (b\text{-})$jets and $Z^0(\to \mu\mu) + (b\text{-})$jets in data and simulation, along with their respective ratios using only negative *probe* muons.

Appendix C

Background scale factor systematic uncertainty estimations

As described in Section 4.6, the scale factors for muons from hadron decays have systematic uncertainties due to top quark modelling and the prompt-background subtraction. In order to estimate these uncertainties, the scale factors were recalculated using different configurations and compared to the nominal configuration. This appendix contains the input values and calculated values with each of the different configurations. Appendix C.1 contains these values for different top modelling configurations, and Appendix C.2 contains these values for the different prompt-background subtraction configurations.

C.1 Uncertainty due to top modelling

This appendix contains the input values and calculated values for the different top modelling configurations. Since some of the alternative configurations are generated using a faster simulation programme, the fast simulation version of the nominal configuration is used for comparison. This avoids folding in differences between the different simulation programmes into the comparison. A comparison of the fast simulation and full simulation nominal configuration (Powheg+Pythia8) is shown in Tables C.1 and C.2 for the e-μ channel and μ-μ channel respectively. The following configurations were compared against the nominal:

- Powheg+Herwig uses the Herwig v7.04 [104,105], H7UE tune [105], `MMHT2014LO` PDF set [153], which is an alternative generator to Pythia8. This varies the parton shower and hadronisation model. This comparison is shown in Tables C.3 and C.4 for the e-μ channel and μ-μ channel respectively.

- Powheg+Pythia8-h_{damp} is the configuration wherein the Powheg h_{damp} parameter is varied from $1.5m_{\mathrm{top}}$ to $3m_{\mathrm{top}}$, which varies the resummation scale. This

comparison is shown in Tables C.5 and C.6 for the e-μ channel and μ-μ channel respectively.

- Final State Radiation down (FSR down) varies the renormalisation scale for QCD emission in the FSR by a factor of 0.5, to vary the FSR tuning. This comparison is shown in Tables C.7 and C.8 for the e-μ channel and μ-μ channel respectively.

- Final State Radiation up (FSR up) varies the renormalisation scale for QCD emission in the FSR by a factor of 2.0, to vary the FSR tuning. This comparison is shown in Tables C.9 and C.10 for the e-μ channel and μ-μ channel respectively.

- Initial State Radiation down (ISR down) uses the down variation of the A14 eigen-tune variation [126] of the strong coupling constant (α_s), to vary the ISR tuning. This comparison is shown in Tables C.11 and C.12 for the e-μ channel and μ-μ channel respectively.

- Initial State Radiation up (ISR up) uses the up variation of the A14 eigen-tune variation [126] of the strong coupling constant (α_s), to vary the ISR tuning. This comparison is shown in Tables C.13 and C.14 for the e-μ channel and μ-μ channel respectively.

- $\mu_{R,F}$ down (muRF down) varies the factorisation and renormalisation scales down by a factor of 0.5, in order to simulate changes to the amount of parton radiation and potential missing higher-order corrections. This comparison is shown in Tables C.15 and C.16 for the e-μ channel and μ-μ channel respectively.

- $\mu_{R,F}$ up (muRF up) varies the factorisation and renormalisation scales up by a factor of 2, in order to simulate changes to the amount of parton radiation and potential missing higher-order corrections. This comparison shown in Tables C.17 and C.18 for the e-μ channel and μ-μ channel respectively.

The largest difference in the e-μ channel is seen with the FSR down configuration, and the difference is 7.66%. In the μ-μ channel the largest difference is seen with the POWHEG+HERWIG configuration, and the difference is 3.05%. It is not concerning that the largest differences are seen in different configurations, since the e-μ channel has contributions due to charge misidentified *tag* electrons, which are harder to model.

$e\mu$	Nominal-fast-sim	Nominal-full-sim	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	776.77 ± 8.41	-47.85(-6.56%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	19.09 ± 1.61	21.38 ± 1.67	-2.29(-12.01%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	463.91 ± 6.49	489.57 ± 6.71	-25.66(-5.53%)
S	1.39 ± 0.06	1.32 ± 0.06	0.08(5.58%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1022.79 ± 44.75	-6.32(-0.62%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	1207.47 ± 16.39	-61.05(-5.33%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1767.67 ± 17.67	-32.66(-1.88%)
R	1.47 ± 0.06	1.39 ± 0.06	0.08(5.41%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2453.87 ± 108.57	92.50(3.63%)

Table C.1: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and PP8 Full Sim in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	Nominal-full-sim	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	284.15 ± 2.80	-1.57(-0.55%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	24.05 ± 1.85	25.26 ± 1.88	-1.21(-5.05%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	169.63 ± 2.03	171.77 ± 2.06	-2.15(-1.27%)
S	1.68 ± 0.11	1.65 ± 0.10	0.03(1.67%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	469.37 ± 30.14	5.34(1.13%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	1217.88 ± 15.21	-58.37(-5.03%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1736.61 ± 16.83	-17.10(-0.99%)
R	1.43 ± 0.05	1.37 ± 0.05	0.06(4.49%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2373.64 ± 87.14	87.02(3.54%)

Table C.2: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and PP8 Full Sim in the μ-μ channel.

$e\mu$	Nominal-fast-sim	Powheg+Herwig	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	674.39 ± 5.04	$54.54(7.48\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	$0.00(0.00\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	19.09 ± 1.61	13.86 ± 1.39	$5.23(27.38\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	463.91 ± 6.49	429.39 ± 4.02	$34.53(7.44\%)$
S	1.39 ± 0.06	1.52 ± 0.06	$-0.12(-8.91\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1024.24 ± 42.41	$-7.77(-0.76\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	962.06 ± 15.39	$184.36(16.08\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1518.60 ± 16.78	$216.40(12.47\%)$
R	1.47 ± 0.06	1.74 ± 0.08	$-0.27(-18.61\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2643.60 ± 117.67	$-97.22(-3.82\%)$

Table C.3: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and Powheg+Herwig in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	Powheg+Herwig	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	276.60 ± 2.31	$5.99(2.12\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	$0.00(0.00\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	24.05 ± 1.85	15.13 ± 1.46	$8.92(37.09\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu} > 30\,\mathrm{GeV}}$	169.63 ± 2.03	171.68 ± 1.81	$-2.06(-1.21\%)$
S	1.68 ± 0.11	1.71 ± 0.10	$-0.03(-1.89\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	473.46 ± 29.12	$1.25(0.26\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	981.31 ± 13.99	$178.20(15.37\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1498.50 ± 15.67	$221.01(12.85\%)$
R	1.43 ± 0.05	1.69 ± 0.06	$-0.26(-18.25\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2535.72 ± 94.67	$-75.06(-3.05\%)$

Table C.4: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and Powheg+Herwig in the μ-μ channel.

$e\mu$	Nominal-fast-sim	POWHEG+PYTHIA8-h_{damp}	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	721.69 ± 5.31	$7.23(0.99\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	$0.00(0.00\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	19.09 ± 1.61	21.28 ± 1.63	$-2.20(-11.51\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	463.91 ± 6.49	456.58 ± 4.18	$7.33(1.58\%)$
S	1.39 ± 0.06	1.41 ± 0.06	$-0.02(-1.26\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1019.07 ± 42.59	$-2.60(-0.26\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	1167.57 ± 15.89	$-21.16(-1.85\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1739.14 ± 17.22	$-4.13(-0.24\%)$
R	1.47 ± 0.06	1.44 ± 0.06	$0.03(1.96\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2502.31 ± 108.56	$44.06(1.73\%)$

Table C.5: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and PP8rad in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	POWHEG+PYTHIA8-h_{damp}	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	281.16 ± 2.62	$1.43(0.50\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	$0.00(0.00\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	24.05 ± 1.85	19.93 ± 1.59	$4.11(17.11\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	169.63 ± 2.03	167.87 ± 1.90	$1.76(1.04\%)$
S	1.68 ± 0.11	1.72 ± 0.11	$-0.04(-2.51\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	484.15 ± 30.40	$-9.44(-1.99\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	1161.48 ± 14.40	$-1.97(-0.17\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1732.13 ± 16.17	$-12.62(-0.73\%)$
R	1.43 ± 0.05	1.42 ± 0.05	$0.01(0.74\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2460.46 ± 90.88	$0.20(0.01\%)$

Table C.6: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and PP8rad in the μ-μ channel.

$e\mu$	Nominal-fast-sim	FSR down	difference
$N_{\text{MC}}^{\text{prompt,SS}}$	728.92 ± 8.09	708.70 ± 13.63	$20.23(2.77\%)$
$[N_{\text{data}}^{\text{SS}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	$0.00(0.00\%)$
$[N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	19.09 ± 1.61	15.44 ± 2.05	$3.65(19.11\%)$
$[N_{\text{MC}}^{\text{SS,prompt}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	463.91 ± 6.49	446.69 ± 10.77	$17.22(3.71\%)$
S	1.39 ± 0.06	1.46 ± 0.07	$-0.06(-4.44\%)$
$N_{\text{MC}}^{\text{SS,prompt,scaled}}$	1016.47 ± 44.50	1032.14 ± 51.96	$-15.67(-1.54\%)$
$N_{\text{data}}^{\text{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}$	1146.42 ± 16.27	1128.69 ± 23.84	$17.72(1.55\%)$
$N_{\text{MC}}^{\text{OS},\mu_{\text{had}}}$	1735.01 ± 17.79	1592.18 ± 26.10	$142.83(8.23\%)$
R	1.47 ± 0.06	1.48 ± 0.07	$-0.01(-0.62\%)$
$N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$	2546.37 ± 112.71	2351.33 ± 121.22	$195.04(7.66\%)$

Table C.7: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and FSR down in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	FSR down	difference
$N_{\text{MC}}^{\text{prompt,SS}}$	282.59 ± 2.77	275.52 ± 3.84	$7.07(2.50\%)$
$[N_{\text{data}}^{\text{SS}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	$0.00(0.00\%)$
$[N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	24.05 ± 1.85	18.62 ± 2.85	$5.43(22.57\%)$
$[N_{\text{MC}}^{\text{SS,prompt}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	169.63 ± 2.03	164.92 ± 2.80	$4.71(2.78\%)$
S	1.68 ± 0.11	1.76 ± 0.11	$-0.08(-4.81\%)$
$N_{\text{MC}}^{\text{SS,prompt,scaled}}$	474.71 ± 30.35	485.13 ± 31.61	$-10.41(-2.19\%)$
$N_{\text{data}}^{\text{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}$	1159.51 ± 14.87	1123.04 ± 22.23	$36.47(3.15\%)$
$N_{\text{MC}}^{\text{OS},\mu_{\text{had}}}$	1719.51 ± 16.67	1632.46 ± 25.88	$87.05(5.06\%)$
R	1.43 ± 0.05	1.47 ± 0.06	$-0.04(-2.60\%)$
$N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$	2460.66 ± 91.01	2396.82 ± 101.56	$63.84(2.59\%)$

Table C.8: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and FSR down in the μ-μ channel.

$e\mu$	Nominal-fast-sim	FSR up	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	750.64 ± 10.43	-21.71(-2.98%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	19.09 ± 1.61	20.67 ± 2.40	-1.58(-8.30%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	463.91 ± 6.49	479.13 ± 8.25	-15.22(-3.28%)
S	1.39 ± 0.06	1.35 ± 0.06	0.05(3.41%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1011.01 ± 46.36	5.45(0.54%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	1108.56 ± 18.66	37.86(3.30%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1718.13 ± 21.29	16.87(0.97%)
R	1.47 ± 0.06	1.52 ± 0.07	-0.06(-3.75%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2616.18 ± 120.98	-69.80(-2.74%)

Table C.9: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and FSR up in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	FSR up	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	288.71 ± 3.79	-6.13(-2.17%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	24.05 ± 1.85	28.40 ± 2.75	-4.35(-18.10%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	169.63 ± 2.03	174.00 ± 2.87	-4.38(-2.58%)
S	1.68 ± 0.11	1.61 ± 0.11	0.07(4.00%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	465.58 ± 31.11	9.13(1.92%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	1128.22 ± 17.37	31.29(2.70%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1708.80 ± 20.10	10.71(0.62%)
R	1.43 ± 0.05	1.48 ± 0.05	-0.05(-3.34%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2526.98 ± 97.53	-66.32(-2.70%)

Table C.10: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and FSR up in the μ-μ channel.

$e\mu$	Nominal-fast-sim	ISR down	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	735.99 ± 8.38	-7.07(-0.97%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	19.09 ± 1.61	19.19 ± 1.64	-0.10(-0.53%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	463.91 ± 6.49	466.98 ± 6.69	-3.07(-0.66%)
S	1.39 ± 0.06	1.39 ± 0.06	0.01(0.67%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1019.42 ± 44.82	-2.95(-0.29%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	1143.53 ± 16.43	2.88(0.25%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1734.60 ± 18.05	0.40(0.02%)
R	1.47 ± 0.06	1.47 ± 0.06	-0.00(-0.08%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2547.72 ± 113.47	-1.35(-0.05%)

Table C.11: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and ISR down in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	ISR down	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	281.84 ± 2.77	0.75(0.26%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	24.05 ± 1.85	23.74 ± 1.86	0.31(1.29%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	169.63 ± 2.03	168.63 ± 1.99	0.99(0.58%)
S	1.68 ± 0.11	1.69 ± 0.11	-0.01(-0.70%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	476.77 ± 30.44	-2.05(-0.43%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	1148.50 ± 14.97	11.01(0.95%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1713.91 ± 16.89	5.60(0.33%)
R	1.43 ± 0.05	1.44 ± 0.05	-0.01(-0.83%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2473.10 ± 91.91	-12.44(-0.51%)

Table C.12: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and ISR down in the μ-μ channel.

$e\mu$	Nominal-fast-sim	ISR up	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	739.63 ± 8.51	-10.71(-1.47%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	19.09 ± 1.61	19.41 ± 1.69	-0.32(-1.69%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	463.91 ± 6.49	472.60 ± 6.86	-8.68(-1.87%)
S	1.39 ± 0.06	1.37 ± 0.06	0.03(1.89%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1011.94 ± 44.60	4.52(0.44%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	1170.19 ± 16.75	-23.77(-2.07%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1763.29 ± 18.41	-28.28(-1.63%)
R	1.47 ± 0.06	1.44 ± 0.06	0.03(1.77%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2542.12 ± 112.58	4.25(0.17%)

Table C.13: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and ISR up in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	ISR up	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	284.10 ± 2.90	-1.52(-0.54%)
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	0.00(0.00%)
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	24.05 ± 1.85	24.85 ± 1.97	-0.80(-3.33%)
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	169.63 ± 2.03	171.15 ± 2.18	-1.52(-0.90%)
S	1.68 ± 0.11	1.66 ± 0.11	0.02(1.17%)
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	471.69 ± 30.36	3.02(0.64%)
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	0.00(0.00%)
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	1188.29 ± 15.39	-28.78(-2.48%)
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1745.75 ± 17.26	-26.24(-1.53%)
R	1.43 ± 0.05	1.40 ± 0.05	0.03(2.24%)
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2442.14 ± 90.43	18.52(0.75%)

Table C.14: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and ISR up in the μ-μ channel.

$e\mu$	Nominal-fast-sim	$\mu_{R,F}$ down	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	728.92 ± 8.09	677.69 ± 7.27	$51.23(7.03\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	$0.00(0.00\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	19.09 ± 1.61	17.06 ± 1.44	$2.02(10.61\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	463.91 ± 6.49	430.62 ± 5.83	$33.29(7.18\%)$
S	1.39 ± 0.06	1.51 ± 0.06	$-0.11(-8.07\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1016.47 ± 44.50	1021.27 ± 44.34	$-4.80(-0.47\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1146.42 ± 16.27	1029.91 ± 15.35	$116.51(10.16\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1735.01 ± 17.79	1560.33 ± 17.48	$174.68(10.07\%)$
R	1.47 ± 0.06	1.63 ± 0.07	$-0.16(-10.99\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2546.37 ± 112.71	2541.78 ± 113.81	$4.59(0.18\%)$

Table C.15: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and $\mu_{R,F}$ down in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	$\mu_{R,F}$ down	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	282.59 ± 2.77	277.75 ± 2.55	$4.83(1.71\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	$0.00(0.00\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	24.05 ± 1.85	21.17 ± 1.65	$2.88(11.97\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>30\,\mathrm{GeV}}$	169.63 ± 2.03	167.06 ± 1.86	$2.57(1.51\%)$
S	1.68 ± 0.11	1.72 ± 0.11	$-0.04(-2.56\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	474.71 ± 30.35	478.54 ± 30.16	$-3.83(-0.81\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1159.51 ± 14.87	1040.04 ± 13.92	$119.47(10.30\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1719.51 ± 16.67	1535.48 ± 15.42	$184.03(10.70\%)$
R	1.43 ± 0.05	1.59 ± 0.06	$-0.16(-11.23\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2460.66 ± 91.01	2444.06 ± 91.14	$16.60(0.67\%)$

Table C.16: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and $\mu_{R,F}$ down in the μ-μ channel.

$e\mu$	Nominal-fast-sim	$\mu_{R,F}$ up	difference
$N_{\text{MC}}^{\text{prompt,SS}}$	728.92 ± 8.09	801.08 ± 9.65	-72.15(-9.90%)
$[N_{\text{data}}^{\text{SS}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	666.00 ± 25.81	666.00 ± 25.81	0.00(0.00%)
$[N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	19.09 ± 1.61	21.84 ± 1.88	-2.75(-14.41%)
$[N_{\text{MC}}^{\text{SS,prompt}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	463.91 ± 6.49	509.98 ± 7.70	-46.06(-9.93%)
S	1.39 ± 0.06	1.26 ± 0.05	0.13(9.42%)
$N_{\text{MC}}^{\text{SS,prompt,scaled}}$	1016.47 ± 44.50	1011.86 ± 45.10	4.61(0.45%)
$N_{\text{data}}^{\text{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	0.00(0.00%)
$N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}$	1146.42 ± 16.27	1292.32 ± 17.76	-145.91(-12.73%)
$N_{\text{MC}}^{\text{OS},\mu_{\text{had}}}$	1735.01 ± 17.79	1942.02 ± 23.22	-207.02(-11.93%)
R	1.47 ± 0.06	1.31 ± 0.06	0.16(11.05%)
$N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$	2546.37 ± 112.71	2535.33 ± 113.23	11.04(0.43%)

Table C.17: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and $\mu_{R,F}$ up in the e-μ channel.

$\mu\mu$	Nominal-fast-sim	$\mu_{R,F}$ up	difference
$N_{\text{MC}}^{\text{prompt,SS}}$	282.59 ± 2.77	288.22 ± 3.12	-5.63(-1.99%)
$[N_{\text{data}}^{\text{SS}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	309.00 ± 17.58	309.00 ± 17.58	0.00(0.00%)
$[N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	24.05 ± 1.85	27.65 ± 2.17	-3.60(-14.98%)
$[N_{\text{MC}}^{\text{SS,prompt}}]_{p_{\text{T}}^{\mu}>30\,\text{GeV}}$	169.63 ± 2.03	172.67 ± 2.28	-3.05(-1.80%)
S	1.68 ± 0.11	1.63 ± 0.10	0.05(3.01%)
$N_{\text{MC}}^{\text{SS,prompt,scaled}}$	474.71 ± 30.35	469.62 ± 30.63	5.09(1.07%)
$N_{\text{data}}^{\text{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	0.00(0.00%)
$N_{\text{MC}}^{\text{SS},\mu_{\text{had}}}$	1159.51 ± 14.87	1303.81 ± 16.54	-144.30(-12.45%)
$N_{\text{MC}}^{\text{OS},\mu_{\text{had}}}$	1719.51 ± 16.67	1934.29 ± 18.93	-214.78(-12.49%)
R	1.43 ± 0.05	1.28 ± 0.05	0.15(10.79%)
$N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$	2460.66 ± 91.01	2469.21 ± 91.25	-8.55(-0.35%)

Table C.18: Comparison of inputs and calculated values for the the muons from hadron decays background scale factor between nominal-fast-sim and $\mu_{R,F}$ up in the μ-μ channel.

C.2 Uncertainty due to prompt-background subtraction

The systematic uncertainty due to the modelling of the prompt subtraction (eq. (4.5)), as well as the value of the cut, was estimated by repeating the study at four different cut values: $p_T^\mu < 20$ GeV (Table C.19 (C.20) for e-μ (μ-μ)), $p_T^\mu < 30$ GeV (nominal), $p_T^\mu < 40$ GeV (Table C.21 (C.22) for e-μ (μ-μ)) and $p_T^\mu < 50$GeV (Table C.23 (C.24) for e-μ (μ-μ)). The estimated rate of muons from hadron decays in data, $N_{\text{data}}^{\text{OS},\mu_{\text{had}}}$, varied by at most $|0.96\%|$ ($|1.32\%|$) in the e-μ (μ-μ) channel, when using a cut of $p_T^\mu < 50$ GeV, as opposed to the nominal cut of $p_T^\mu > 30$ GeV for the scaling of the prompt background subtraction. This is therefore the estimated uncertainty due to value of the p_T^μ cut using this approach.

$e\mu$	$p_{\mathrm{T}}^{\mu} > 30\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu} > 20\mathrm{GeV}$	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	776.77 ± 8.41	776.77 ± 8.41	$0.00(0.00\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	666.00 ± 25.81	818.00 ± 28.60	$-152.00(-22.82\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	21.38 ± 1.67	48.47 ± 2.54	$-27.09(-126.74\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	489.57 ± 6.71	593.28 ± 7.39	$-103.71(-21.18\%)$
S	1.32 ± 0.06	1.30 ± 0.05	$0.02(1.49\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1022.79 ± 44.75	1007.54 ± 41.11	$15.25(1.49\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1207.47 ± 16.39	1207.47 ± 16.39	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1767.67 ± 17.67	1767.67 ± 17.67	$0.00(0.00\%)$
R	1.39 ± 0.06	1.40 ± 0.06	$-0.01(-0.91\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2453.87 ± 108.57	2476.20 ± 105.58	$-22.32(-0.91\%)$

Table C.19: Comparison of values used in the estimation of the μ_{had} scale factor for p_{T}^{μ} cuts of 30 GeV and 20 GeV, for the e-μ channel.

$\mu\mu$	$p_{\mathrm{T}}^{\mu} > 30\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu} > 20\mathrm{GeV}$	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	284.15 ± 2.80	284.15 ± 2.80	$0.00(0.00\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	309.00 ± 17.58	406.00 ± 20.15	$-97.00(-31.39\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	25.26 ± 1.88	52.19 ± 2.67	$-26.93(-106.62\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	171.77 ± 2.06	205.98 ± 2.25	$-34.21(-19.91\%)$
S	1.65 ± 0.10	1.72 ± 0.10	$-0.07(-3.99\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	469.37 ± 30.14	488.08 ± 28.94	$-18.71(-3.99\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1217.88 ± 15.21	1217.88 ± 15.21	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1736.61 ± 16.83	1736.61 ± 16.83	$0.00(0.00\%)$
R	1.37 ± 0.05	1.35 ± 0.05	$0.02(1.12\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2373.64 ± 87.14	2346.96 ± 86.14	$26.68(1.12\%)$

Table C.20: Comparison of values used in the estimation of the μ_{had} scale factor for p_{T}^{μ} cuts of 30 GeV and 20 GeV, for the μ-μ channel.

$e\mu$	$p_{\mathrm{T}}^{\mu} > 30\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu} > 40\mathrm{GeV}$	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	776.77 ± 8.41	776.77 ± 8.41	$0.00(0.00\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	666.00 ± 25.81	529.00 ± 23.00	$137.00(20.57\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	21.38 ± 1.67	11.90 ± 1.25	$9.48(44.35\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	489.57 ± 6.71	390.47 ± 5.97	$99.10(20.24\%)$
S	1.32 ± 0.06	1.32 ± 0.06	$-0.01(-0.58\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	1022.79 ± 44.75	1028.69 ± 49.71	$-5.90(-0.58\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1207.47 ± 16.39	1207.47 ± 16.39	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1767.67 ± 17.67	1767.67 ± 17.67	$0.00(0.00\%)$
R	1.39 ± 0.06	1.38 ± 0.06	$0.00(0.35\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2453.87 ± 108.57	2445.24 ± 113.04	$8.63(0.35\%)$

Table C.21: Comparison of values used in the estimation of the μ_{had} scale factor for p_{T}^{μ} cuts of 30 GeV and 40 GeV, for the e-μ channel.

$\mu\mu$	$p_{\mathrm{T}}^{\mu} > 30\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu} > 40\mathrm{GeV}$	difference
$N_{\mathrm{MC}}^{\mathrm{prompt,SS}}$	284.15 ± 2.80	284.15 ± 2.80	$0.00(0.00\%)$
$[N_{\mathrm{data}}^{\mathrm{SS}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	309.00 ± 17.58	239.00 ± 15.46	$70.00(22.65\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	25.26 ± 1.88	13.44 ± 1.37	$11.82(46.79\%)$
$[N_{\mathrm{MC}}^{\mathrm{SS,prompt}}]_{p_{\mathrm{T}}^{\mu}>p_{\mathrm{T}}^{\mu}\,\mathrm{GeV}}$	171.77 ± 2.06	135.98 ± 1.73	$35.80(20.84\%)$
S	1.65 ± 0.10	1.66 ± 0.12	$-0.01(-0.42\%)$
$N_{\mathrm{MC}}^{\mathrm{SS,prompt,scaled}}$	469.37 ± 30.14	471.36 ± 33.31	$-1.99(-0.42\%)$
$N_{\mathrm{data}}^{\mathrm{SS}}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{SS},\mu_{\mathrm{had}}}$	1217.88 ± 15.21	1217.88 ± 15.21	$0.00(0.00\%)$
$N_{\mathrm{MC}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	1736.61 ± 16.83	1736.61 ± 16.83	$0.00(0.00\%)$
R	1.37 ± 0.05	1.37 ± 0.05	$0.00(0.12\%)$
$N_{\mathrm{data}}^{\mathrm{OS},\mu_{\mathrm{had}}}$	2373.64 ± 87.14	2370.81 ± 89.44	$2.83(0.12\%)$

Table C.22: Comparison of values used in the estimation of the μ_{had} scale factor for p_{T}^{μ} cuts of 30 GeV and 40 GeV, for the μ-μ channel.

$e\mu$	$p_\mathrm{T}^\mu > 30\,\mathrm{GeV}$	$p_\mathrm{T}^\mu > 50\,\mathrm{GeV}$	difference
$N_\mathrm{MC}^\mathrm{prompt,SS}$	776.77 ± 8.41	776.77 ± 8.41	$0.00(0.00\%)$
$[N_\mathrm{data}^\mathrm{SS}]_{p_\mathrm{T}^\mu > p_\mathrm{T}^\mu\,\mathrm{GeV}}$	666.00 ± 25.81	390.00 ± 19.75	$276.00(41.44\%)$
$[N_\mathrm{MC}^{\mathrm{SS},\mu_\mathrm{had}}]_{p_\mathrm{T}^\mu > p_\mathrm{T}^\mu\,\mathrm{GeV}}$	21.38 ± 1.67	6.50 ± 0.91	$14.88(69.60\%)$
$[N_\mathrm{MC}^\mathrm{SS,prompt}]_{p_\mathrm{T}^\mu > p_\mathrm{T}^\mu\,\mathrm{GeV}}$	489.57 ± 6.71	295.89 ± 5.20	$193.67(39.56\%)$
S	1.32 ± 0.06	1.30 ± 0.07	$0.02(1.57\%)$
$N_\mathrm{MC}^\mathrm{SS,prompt,scaled}$	1022.79 ± 44.75	1006.76 ± 55.90	$16.03(1.57\%)$
$N_\mathrm{data}^\mathrm{SS}$	2699.00 ± 51.95	2699.00 ± 51.95	$0.00(0.00\%)$
$N_\mathrm{MC}^{\mathrm{SS},\mu_\mathrm{had}}$	1207.47 ± 16.39	1207.47 ± 16.39	$0.00(0.00\%)$
$N_\mathrm{MC}^{\mathrm{OS},\mu_\mathrm{had}}$	1767.67 ± 17.67	1767.67 ± 17.67	$0.00(0.00\%)$
R	1.39 ± 0.06	1.40 ± 0.07	$-0.01(-0.96\%)$
$N_\mathrm{data}^{\mathrm{OS},\mu_\mathrm{had}}$	2453.87 ± 108.57	2477.33 ± 119.27	$-23.46(-0.96\%)$

Table C.23: Comparison of values used in the estimation of the μ_had scale factor for p_T^μ cuts of 30 GeV and 50 GeV, for the e-μ channel.

$\mu\mu$	$p_\mathrm{T}^\mu > 30\,\mathrm{GeV}$	$p_\mathrm{T}^\mu > 50\,\mathrm{GeV}$	difference
$N_\mathrm{MC}^\mathrm{prompt,SS}$	284.15 ± 2.80	284.15 ± 2.80	$0.00(0.00\%)$
$[N_\mathrm{data}^\mathrm{SS}]_{p_\mathrm{T}^\mu > p_\mathrm{T}^\mu\,\mathrm{GeV}}$	309.00 ± 17.58	176.00 ± 13.27	$133.00(43.04\%)$
$[N_\mathrm{MC}^{\mathrm{SS},\mu_\mathrm{had}}]_{p_\mathrm{T}^\mu > p_\mathrm{T}^\mu\,\mathrm{GeV}}$	25.26 ± 1.88	9.29 ± 1.17	$15.97(63.23\%)$
$[N_\mathrm{MC}^\mathrm{SS,prompt}]_{p_\mathrm{T}^\mu > p_\mathrm{T}^\mu\,\mathrm{GeV}}$	171.77 ± 2.06	105.89 ± 1.51	$65.89(38.36\%)$
S	1.65 ± 0.10	1.57 ± 0.13	$0.08(4.68\%)$
$N_\mathrm{MC}^\mathrm{SS,prompt,scaled}$	469.37 ± 30.14	447.39 ± 36.57	$21.98(4.68\%)$
$N_\mathrm{data}^\mathrm{SS}$	2134.00 ± 46.20	2134.00 ± 46.20	$0.00(0.00\%)$
$N_\mathrm{MC}^{\mathrm{SS},\mu_\mathrm{had}}$	1217.88 ± 15.21	1217.88 ± 15.21	$0.00(0.00\%)$
$N_\mathrm{MC}^{\mathrm{OS},\mu_\mathrm{had}}$	1736.61 ± 16.83	1736.61 ± 16.83	$0.00(0.00\%)$
R	1.37 ± 0.05	1.38 ± 0.05	$-0.02(-1.32\%)$
$N_\mathrm{data}^{\mathrm{OS},\mu_\mathrm{had}}$	2373.64 ± 87.14	2404.99 ± 92.22	$-31.34(-1.32\%)$

Table C.24: Comparison of values used in the estimation of the μ_had scale factor for p_T^μ cuts of 30 GeV and 50 GeV, for the μ-μ channel.