Measurement of the $t\bar{t}$ cross-section at 7 TeV with 36 pb^{-1} of data in the electron+jets decay channel using the CMS detector

Matthew Chadwick

A thesis submitted for the degree of
Doctor of Philosophy

School of Engineering and Design
Brunel University, Uxbridge

June 13, 2012

Abstract

A measurement of the top-pair production cross-section at a centre-of-mass energy of 7 TeV using proton-proton collisions with 36 pb^{-1} of data collected by CMS at the Large Hadron Collider is presented. The analysis is performed using the final state that consists of one isolated electron with jets, one of which is required to be identified as being consistent with including the decay of a B hadron. The measured cross-section with three or more selected jets is $169 \pm 13(\text{stat.})^{+37}_{-32}(\text{sys.})^{+8}_{-7}(\text{lumi.})$ pb and $197 \pm 17(\text{stat.})^{+38}_{-35}(\text{sys.})^{+9}_{-8}(\text{lumi.})$ pb for four or more jets. The results are consistent with NLO and approximate NNLO theoretical predictions.

Declaration

This thesis presents a measurement performed in the CMS collaboration, specifically within the Top Physics Analysis Group. As part of the CMS collaboration the author has contributed service work to CMS. In particular, the author has contributed to the software for monitoring part of the ECAL end-caps and is the database liaison and maintainer for the B-Tagging and Vertexing Physics Object Group.

The measurement was performed as part of the *BCBR* group which includes the author as well as Matthew Barrett[1], Freya Blekman[4], Joanne Cole[1], Vesna Cuplov[3], Karl Ecklund[3], Rebecca Gonzalez Suarez[4], William Martin[1], Julia Thorn[2], Yao Weng[2] and James Zabel[3]. The author is solely responsible for the measurement of the $t\bar{t}$ and W+jets b-tagging event selection efficiencies (Section 6.4 and Section 7.2.2), their systematic studies (Section 8.1), the systematic studies on the PDF uncertainties (Section 8.6), and the final calculation of the cross-section measurement (Section 8.11, Section 9.1 and Appendix D). The author was also responsible for maintaining the core software for the group.

[1]Brunel University, London, UK
[2]Cornell University, Ithaca, USA
[3]Rice University, Houston, USA
[4]Vrije University, Brussels, Belgium

Acknowledgements

I would like to thank all those who have given their very valuable time to give guidance and support for my work. In particular my supervisors, Peter Hobson and Akram Khan who were always available for advice when needed. I would also like to thank all those experienced members of the CMS collaboration who showed great patience and helped when they did not have to. This includes but is not limited to Sam Harper who is a good friend as well as an expert on electrons and triggers, Freya Blekman who taught me almost everything about the practicalities of performing an analysis and much more besides, and Matthew Barrett who was always available to discuss any issues with. Freya in particular remains an inspirational figure. I of course would like to thank everyone in the BCBR group who made working with them such a pleasure and more than anything convinced me that this is something I wish to continue doing.

Additionally I must acknowledge the support from my friends and family without which none of this would be possible. Especially the UK community at CERN whilst I was there, which included such a vibrant range of wonderful people. I would also like to thank the Science and Technologies Funding Council for supporting me financially during my studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The top quark was first discovered at the Tevatron by the CDF and DØ collaborations [1, 2]. Since the initial discovery in 1995, the properties of the top quark have been studied extensively at the Tevatron. A review of top quark physics can be found elsewhere [3]. The production of top quarks is a process of wide ranging importance at the LHC beyond mere academic curiosity. The top quark has the highest mass of any elementary particle discovered so far ($172.0 \pm 0.9 \pm 1.3$ GeV/c^2 [3]) with a very short lifetime of $\approx 0.5 \times 10^{-24}$ s. In particular, this lifetime is shorter than the hadronisation timescale meaning that the top is the only quark that can be studied in an unbound state through its decay products.

The topology of top decays (see Section 1.3) and their high production rate at the LHC makes $t\bar{t}$ production a very useful "standard candle" which can be exploited for a number of practical purposes. For detector calibration, top decays feature almost every physics object, which allows for $t\bar{t}$ events to be used in their calibration. In particular, $t\bar{t}$ events are useful for calibrating b-tagging efficiencies, the jet energy scale and the missing transverse energy in topologies quite different from the events in which these are typically measured.

In terms of pure Physics applications, the invariant mass of $t\bar{t}$ pairs can be used to

search for Z'-like bosons[4] and properties such as the charge asymmetry in $t\bar{t}$ events are sensitive to the presence of new physics beyond the Standard Model [5]. For other physics searches such as in the SUSY phase space, top quarks often present a large irreducible background so the precise classification of $t\bar{t}$ can improve the precision of these searches.

1.1 The Standard Model in brief

The Standard Model (SM) of particle physics is a theory that encompasses all known fundamental particles and their interactions with the electromagnetic, weak and strong nuclear forces. The SM incorporates three generations of quarks and leptons as well as four gauge bosons which mediate the interactions with the forces [3].

The electromagnetic interaction occurs between charged particles and is mediated by the photon. Within the SM it is described by Quantum Electrodynamics (QED) which incorporates electrodynamics within a relativistic quantum field theory. The weak interaction occurs between particles with weak isospin, and is mediated by the massive $W^{\pm}$ and Z^0 bosons. The fact that the exchange bosons of the weak interaction are massive leads to the property of a weak force (in comparison to the electromagnetic or strong forces) with a very short range interaction. An additional property of the weak interaction is that the $W^{\pm}$-boson only couples to left-handed (right-handed) particles (anti-particles) and as such is the only parity violating interaction. The weak interaction is also the only flavour changing interaction within the SM. At higher energies of the order of the Z^0 boson mass, the electromagnetic and weak interactions are described by the combined electroweak interaction.

The strong nuclear force within the SM is described by Quantum Chromodynamics (QCD) and is mediated by the gluon. Analogous to the electromagnetic force, it acts between particles which have "colour" charge, a property that only quarks and gluons possess. At lower energies the colour charge property of gluons leads to an unusual

aspect of QCD interactions known as "colour confinement" whereby attempting to separate two quarks in a bound state within a hadron increases the strength of the colour field between them. The practical effect of this property is that no free quarks can be observed as they instead form hadrons on very short time scales. However, a second feature of quarks within QCD is "asymptotic freedom" where by at high enough energy scales it becomes energetically favourable for a quark to form a new hadron with combinations of other quarks. The practical implication of this effect is that the strength of the coupling constant in QCD, α_s, becomes dependent on the energy scale of the process. This effect also gives rise to collimated showers of particles in which quarks hadronise. These showers of particles are commonly referred to as jets.

The final piece of the SM is the as yet unobserved Higgs boson. The original theory of the electroweak interaction does not include a description of the mass of the $W^{\pm}$ and Z^0 bosons within the Lagrangian. This clearly does not reflect reality. A number of theories have been developed to address the issue, within the SM this is addressed by the Higgs mechanism which leads to electroweak symmetry breaking[6]. A consequence of this mechanism is that it results in an additional, massive gauge boson referred to as the Higgs boson. Typically, this description is then extended to give mass to the quarks and leptons as well. The search for the Higgs boson forms a large part of the ongoing physics programmes at the Tevatron and the LHC. At the time of writing the mass of the SM Higgs boson is excluded for $m_H < 114.4$ GeV/c^2 by LEP[7] and in the region of $156 < m_H < 177$ GeV/c^2 by the Tevatron[8] at a 95% confidence level.

1.2 $t\bar{t}$ production at the LHC

The leading order production diagrams for $t\bar{t}$ pairs at the LHC are shown in Figure 1.1. The dominant production modes at the LHC are the gluon fusion processes which accounts for approximately 75% of the leading order production cross-section [9]. The cross-section for $t\bar{t}$ production ($\sigma_{t\bar{t}}$) is expressed mathematically as in Reference [9],

in equation 1.1,

$$d_{\sigma_{t\bar{t}}} = \sum_{i,j} \int d_{x_i} d_{x_j} \phi_i(x_i, \mu_F) \phi_j(x_j, \mu_F) d_{\sigma_{ij}^{\wedge}}(x_i, x_j, \mu_F, \mu_R, \alpha_s(\mu_R)) \tag{1.1}$$

where the summation is over the integrals for the partons i and j with momentum fractions of x_i and x_j. The functions $\phi(x, \mu_F)$ are the parton distribution functions which represent the probability density to find a parton of type i (j) with a momentum fraction of x_i (x_j). The parameters $d_{\sigma_{ij}^{\wedge}}$ are the parton interaction cross-sections for particles i and j. The constants μ_F, μ_R and α_s are the factorisation and renormalisation scales and the strong coupling constant, respectively. Some calculations of the $t\bar{t}$ production cross-section at next-to-leading order (NLO) and approximate next-to-next-to-leading order (NNLO) from Reference[10] are given in Table 1.1.

Table 1.1: Some theoretical calculations of the $t\bar{t}$ production cross-section at the LHC.

	$\sigma_{t\bar{t}}$	Notes
$\sigma_{t\bar{t}}^{NLO}$(MCFM)	158^{+23}_{-24} pb	Using MCFM [11, 12] with PDF4LHC prescription [13]
$\sigma_{t\bar{t}}^{approx.NNLO}$(Kidonakis)	163^{+11}_{-10} pb	Kidonakis with MSTW2008 NNLO PDF [14]
$\sigma_{t\bar{t}}^{NLO}$(HATHOR)	164^{+10}_{-13} pb	Using HATHOR [15, 16] with MSTW2008 NLO PDF
$\sigma_{t\bar{t}}^{approx.NNLO}$(HATHOR)	160^{+22}_{-23} pb	Using HATHOR with MSTW2008 NNLO PDF

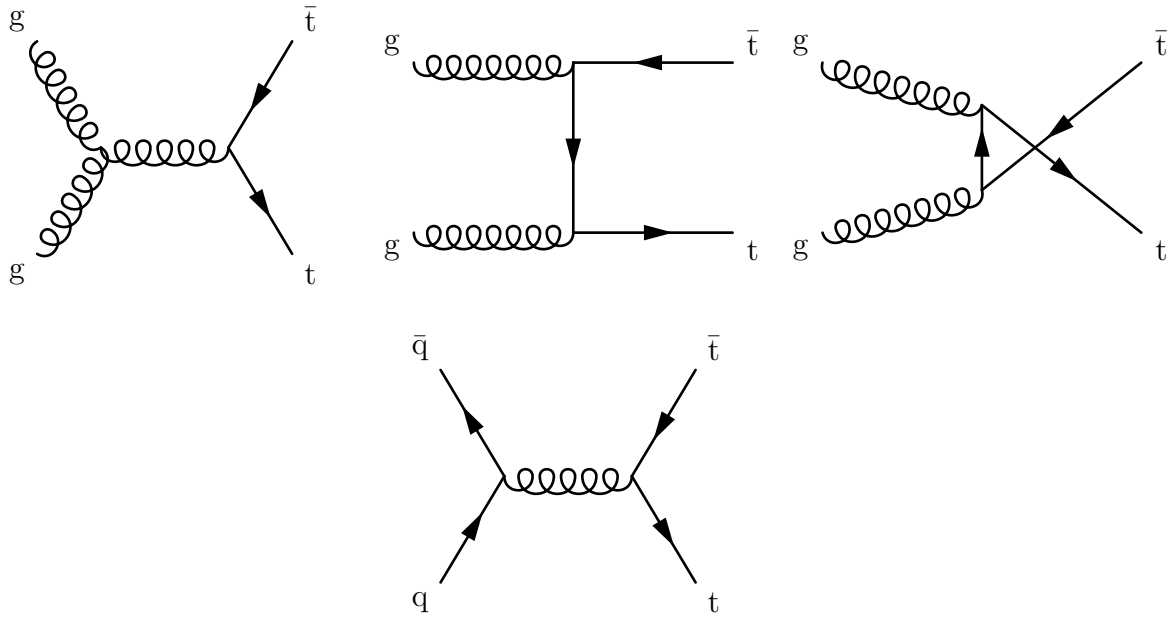


Figure 1.1: The leading order production diagrams for $t\bar{t}$ at the LHC.

1.3 Top quark decay topology

An important aspect of $t\bar{t}$ physics is that the lifetime of the top quark is so short that it decays before hadronisation can occur. This is in contrast to the other flavours of quarks, which hadronise with other quarks to form a wide variety of baryons and mesons with differing physical properties and decay topologies. As such, the decay of top quarks can only proceed by the electroweak interaction.

The coupling between quark flavours in charged interactions with $W^{\pm}$ bosons is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [3], which is shown in equation 1.2.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.2}$$

The relative proportion of decays that proceed by the process $t \to Wx$ are given by the square of the amplitude of the relevant CKM matrix element, $|V_{tx}|^2$. As the CKM matrix is approximately unitary, this suggests that the majority of top decays produce a $W^{\pm}$ boson and a b quark. The value of $|V_{tb}|$ has been directly measured experimentally at the Tevatron by the CDF and DØ collaborations from single-top production [17, 18, 19]:

$$|V_{tb}| = 0.88 \pm 0.07 \text{ (stat + syst)} \tag{1.3}$$

The best estimate of the current parameters of the CKM matrix is provided by a global fit to the individual measurements of the parameters. An overview of the numerous measurements and experiments involved can be found elsewhere [3]. The global fit yields the results shown in equation 1.4.

$$V_{CKM} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.00347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{pmatrix} \tag{1.4}$$

The $t\bar{t}$ events are then classified by the subsequent decay of the $W^{\pm}$ boson. The $W^{\pm}$ boson can decay either leptonically to an electron and a neutrino or hadronically to a quark and an anti-quark of differing flavour. The di-lepton or fully leptonic channel refers to the case where both $W^{\pm}$ bosons from the t and $\bar{t}$ quark decay leptonically. The semi-leptonic or lepton+jets channel refers to the case where one $W^{\pm}$ boson decays hadronically and one $W^{\pm}$ boson decays leptonically. The final channel is the fully hadronic channel where both $W^{\pm}$ bosons decay hadronically. The relative proportion of these decays is 10.3%, 43.5% and 46.2% respectively which includes the channels with tau-leptons [3].

Specifically for this thesis, the cross-section measurement is performed in the electron+jets channel. The final topology is categorised by two b-quarks, one electron, one neutrino and two additional quarks. A discussion of the LHC machine, the CMS detector and the reconstruction will be presented in Chapter 2, Chapter 3 and Chapter 4. The analysis event selection, application of b-tagging, background estimations and systematic studies will be presented in Chapter 5, Chapter 6, Chapter 7 and Chapter 8. The thesis concludes with the cross-section measurement and discussion in Chapter 9.

Chapter 2

The CMS Detector

The Compact Muon Solenoid (CMS) detector is housed at the the European Centre for Nuclear Research (CERN) as part of the Large Hadron Collider (LHC) experiment. The LHC is designed to deliver proton-proton collisions at a centre-of-mass energy of 14 TeV with a peak luminosity of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [20]. This peak luminosity corresponds to 2808 bunches of protons per beam with a total of $1.1 \cdot 10^{11}$ protons per bunch. The bunches are stored within a circular ring with a 27 km long beam pipe. The spacing between the bunches at design luminosity is 25 ns which provides a maximum collision frequency of 40 MHz.

The CMS detector makes use of a right-handed coordinate system. The origin is defined as the nominal interaction point, with the x-axis pointing into the ring, the y-axis pointing upwards and the z-axis pointing in the anti-clockwise direction. The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, ϕ, is defined as the angle in the $x-y$ plane. It is often useful to express sections or locations in the θ plane as pseudorapidities (η) which is defined in terms of θ in equation 2.1.

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

The spacial separation of objects are frequently compared using the quantity ΔR, which is defined as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

The CMS detector is optimised to take advantage of the high luminosity conditions at the LHC. In particular, the electronics and triggering system have to be fast enough to take advantage of the high collision frequency. The high luminosity does come with an associated cost in terms of the hostile radiation environment, particularly for the sub-detectors near to the beam pipe such as the tracking system or for those at higher pseudorapidities such as the endcap calorimeters and forward ($|\eta| > 3$) sub-detectors. This chapter will focus on sub-detectors essential to the performance of this analysis, a more complete description of the detector can be found elsewhere [21].

Throughout the 2010 data taking phase the centre-of-mass energy was set to 7 TeV. The delivered and recorded integrated luminosity at the CMS detector is shown in Figure 2.1. The CMS experiment itself is situated at Point 5 on the LHC ring.

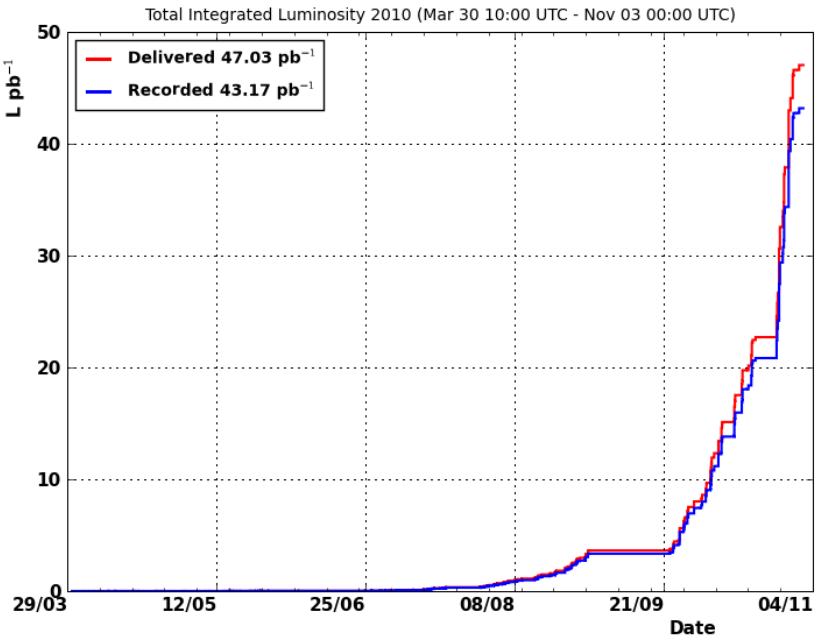


Figure 2.1: The delivered, and recorded integrated luminosity at the CMS experiment. The delivered luminosity corresponds to that delivered by the LHC. The recorded luminosity corresponds to when CMS was recording data.

2.1 Overview

The CMS detector is 21.6 metres long and 14.6 metres in diameter with a weight of 14000 tons, the layout of the sub-detectors is illustrated in Figure 2.2. The detector has a superconducting solenoid 13 m long and 6 m in diameter which provides an axial magnetic field of 3.8 T. Inside the central core of the magnet around the beam-pipe sits a silicon tracking system measuring 5.8 m in length and 2.6 m in diameter. The tracker provides good coverage of pseudorapidities up to $|\eta| < 2.5$. Around the tracker sits the electromagnetic calorimeter (ECAL). The calorimeter as a whole covers pseudorapidities up to $|\eta| < 3.0$ making use of lead tungstate ($PbWO_4$) scintillating crystals with the light being detected by silicon avalanche photo-diodes (APDs) in the barrel and vacuum photo-triodes (VPTs) in the endcap region. The hadronic calorimeter (HCAL) fills the remaining space between the ECAL and the superconducting magnet. This is a brass-scintillator sampling calorimeter using a plastic scintillator which covers pseudorapidities up to $|\eta| < 3.0$. Outside of the magnet are the muon detectors and additional forward calorimeters which extend the hadronic calorimeter coverage up to a pseudorapidity of $|\eta| < 5.0$.

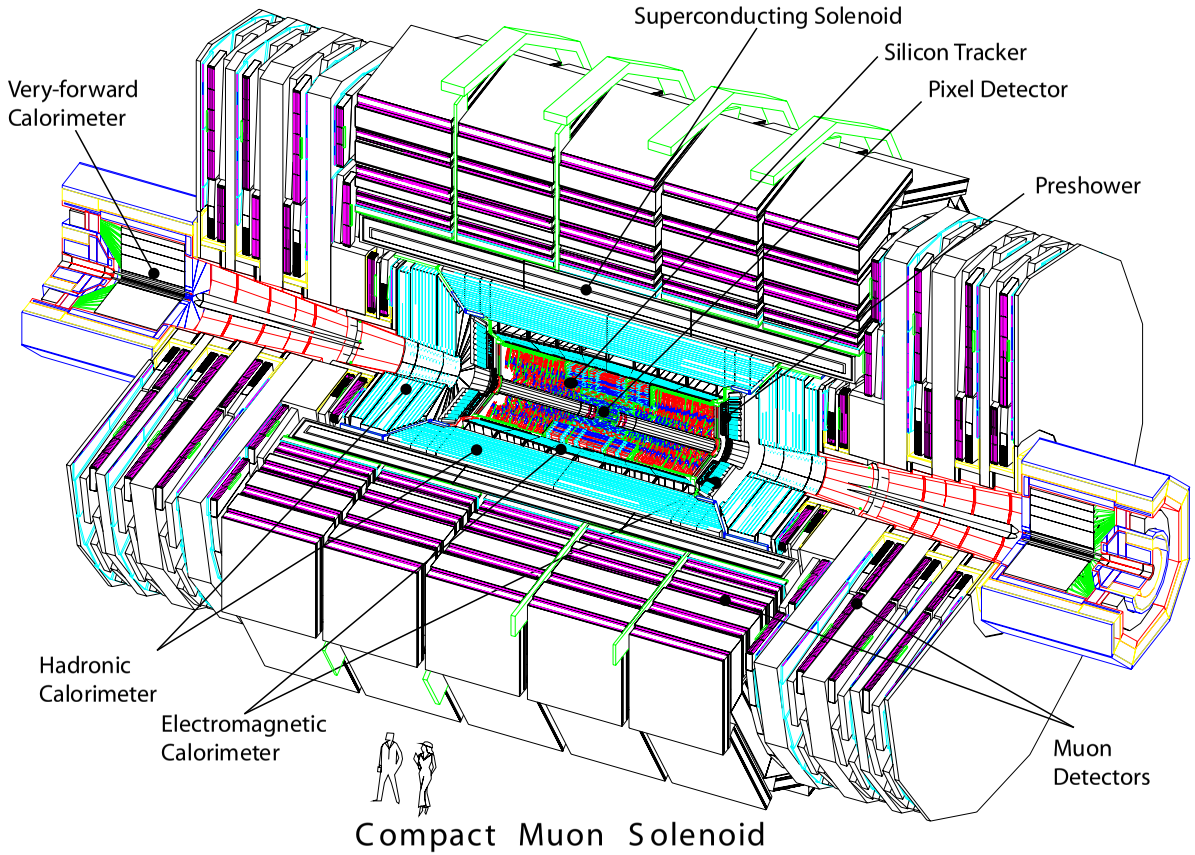


Figure 2.2: Cut away diagram of the CMS detector illustrating the location and relative size of the sub-detectors.

2.2 Tracker

2.2.1 Overview

The tracker is designed to provide precise spacial measurements of the path of charged particles that pass through the tracking volume. As such it is an important component in the reconstruction of these charged particles in combination with the calorimeters and muon chambers. In particular for this analysis the tracker is designed to provide precise reconstruction of secondary vertices observed in heavy flavour decays as well as providing track information for electron and muon identification. An additional concern as the instantaneous luminosity increases, is the reconstruction of extra primary vertices caused by pileup. The tracker is designed to achieve high precision tracking of the relatively high momentum tracks originating from collisions at the TeV scale. As such it is composed entirely of silicon, making use of a silicon pixel tracker in the inner core close to the beam line with silicon micro strips at higher radii. The design of the tracker is a trade-off between a high density of silicon pixels and micro-strips which provide a high spatial resolution with the cost of these devices and also the amount of cooling and support structures that such a device needs to operate. It is obviously desirable to keep the amount of dead material to a minimum in the detector, but it also contributes significantly to the number of radiation lengths before the electromagnetic calorimeter (see Section 3.4). The tracker must also be able to survive the high radiation environment at the LHC [22, 23].

The layout of the tracker and its various sub-systems are detailed in Figure 2.3. As previously mentioned, closest to the beam line lies the pixel detector, located at radii of 4.4, 7.3 and 10.2 cm, arranged in cylindrical layers. There are also two disks of silicon pixels at each end, extending the pseudorapidity coverage. Each individual pixel covers an area of 100x150 μm^2 in the $r-\phi$ and z coordinates, respectively. In total there are 66 million pixels which cover a total surface area of 1 m^2.

Around the pixel detector is the silicon strip detector covering the radial region between

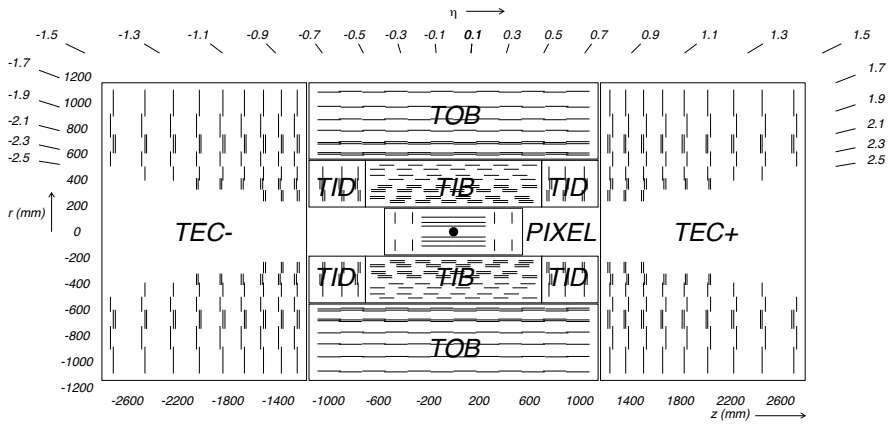


Figure 2.3: A schematic diagram of the layout of the silicon tracker in the CMS detector.

20 and 116 cm. The silicon strip detector is separated into multiple subsystems with the tracker inner barrel and disks (TIB/TID) extending up to a radius of 55 cm. The inner section is composed of four barrel layers and supplemented by three disks at each end. Additionally, the inner two layers of the TIB and the first two disks of the TID have extra silicon strips mounted on the back at small stereo angles. This enables measurements of the z co-ordinate in the barrel and r in the disks. The tracker outer barrel (TOB) sits around the TIB and TID covering the radius up to 116 cm. It consists of six barrel layers of silicon micro-strips. This is complemented by nine disks either side of the TOB known as the tracker endcap (TEC+ and TEC−, where + or − refers to the location in the z direction). These sub-systems again make use of additional strips mounted on the back for the inner two layers of the TOB as well as disks one, two, and five in the endcaps. The typical cell size of a micro-strip in the inner part of the tracker is 10 cm x 80 μm with a thickness of 320 μm. For the outer sub-detectors the thickness is increased to 500 μm and also the cell size is increased up to 10 cm x 180 μm. In total the CMS silicon strip tracker comprises 9.3 million strips with an active area of 198 m^2.

2.2.2 Performance

The performance of the silicon tracker has been extensively studied with collision events since data taking began. As the tracker performance is an integral part of the b-tagging performance it is presented in much greater detail in Chapter 3.

2.3 Electromagnetic calorimeter

2.3.1 Overview

The electromagnetic calorimeter (ECAL) is concerned primarily with the reconstruction of electrons and photons. The ECAL is separated into three distinct sections. The ECAL barrel and endcap sections make use of scintillating lead tungstate crystals ($PbWO_4$). The layout of the ECAL is shown in Figure 2.4. The ECAL barrel covers a pseudorapidity range up to 1.479 with the crystals arranged in a radial configuration. In total there are 61200 crystals in the barrel. The front face of each crystal has a cross-section of 22x22 mm^2. The length of each crystal is 230 mm which corresponds to 25.8 radiation lengths. The intensity of the scintillating light produced in the crystals is measured with avalanche photo-diodes (APD). The crystals in the barrel are grouped into 36 supermodules, each of which covers a 20° region in ϕ and one half of the full length of the barrel. The supermodules themselves are composed of four modules of crystals which contain between 400 and 500 crystals each.

The ECAL endcaps extend the pseudorapidity coverage to $1.479 < |\eta| < 3.0$. Each endcap is divided into two halves known as "dees"; each dee contains 3662 crystals. The crystals are arranged to a focus 1300 mm beyond the interaction point. The front face of each crystal in the endcap has a cross-section of 29x29 mm^2. The length of each crystal is 220 mm which corresponds to 24.7 radiation lengths. In the endcaps the intensity of the scintillating light is measured with vacuum photo-triodes (VPT).

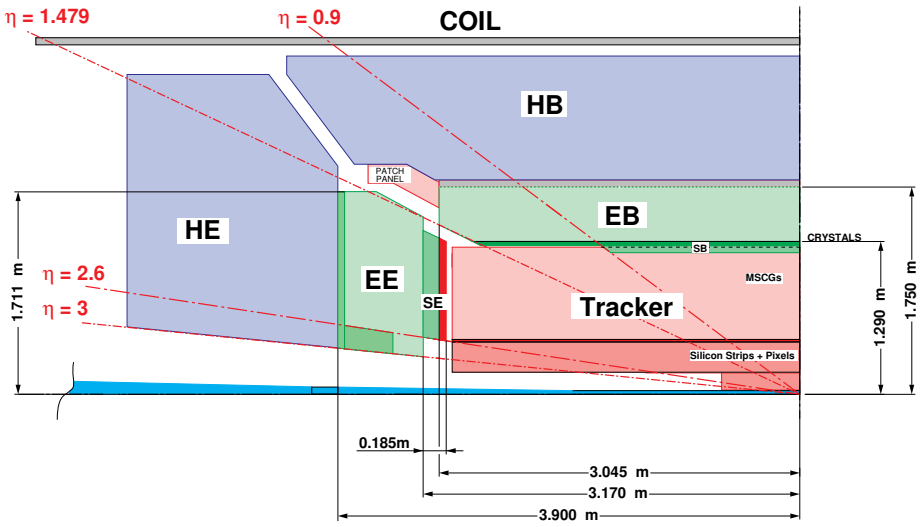


Figure 2.4: A schematic diagram of the electromagnetic calorimeter showing the position of the barrel (EB), endcaps (EE) and pre-shower (SE and SB) sub-detectors in relation to the tracker and hadron calorimeter.

Additional to the ECAL barrel and endcaps, a pre-shower detector is located between the ECAL endcap and the tracker. The pre-shower detector covers the pseudorapidity region $1.653 < |\eta| < 2.6$. The goal of the pre-shower detector is to identify neutral pions, help distinguish between electrons and minimum ionising particles, and to improve the position resolution of electrons and photons in the endcap region. The pre-shower is a sampling calorimeter composed of lead plates backed by silicon strip detectors. The first lead plate corresponds to two radiation lengths, whilst the second corresponds to one radiation length.

2.3.2 Performance

The calibration of the ECAL with collision data is an ongoing process. The estimated particle energy, E, measured by the ECAL is derived from the number of uncalibrated

reconstructed hits (ADC count) in the APDs and VPTs with equation 2.2.

$$E = F \cdot \sum_{\text{cluster crystals}} G(\text{GeV/ADC}) \cdot \text{C}_\text{i} \cdot \text{A}_\text{i} \tag{2.2}$$

where G is the energy scale in GeV as a function of ADC count, C_i is the inter-calibration constant of the crystal, and A_i is the the ADC count. The parameter F is an additional energy scale correction which is dependent on other factors such as the particle type, and its pseudorapidity. This is discussed in greater detail elsewhere [24]. The summation is performed over clusters of crystals chosen algorithmically. So far collision events have been used to measure the ECAL energy scale and the inter-calibration constants [25].

The inter-calibration constants have been measured using two separate techniques. The first technique exploits the ϕ symmetry in minimum bias events to measure the inter-calibration in sections of pseudorapidity. The results for a section are shown in Figure 2.5 compared to simulation. As this technique can only calibrate crystals relative to each other in regions of constant pseudorapidity, a separate technique is used to relate regions of different pseudorapidity. Due to the limited amount of data available for this study, it is performed by reconstructing the invariant mass of di-photon decays from π^0 and η mesons. An iterative procedure is then performed to extract the inter-calibration constants. The reconstructed invariant mass for π^0 decays is shown in Figure 2.6 for the ECAL barrel region.

The measurement of the ECAL energy scale can also be derived from the reconstruction of the invariant mass of the π^0 and η mesons. The measured invariant mass is compared to simulation for the each of the ECAL endcaps and the barrel separately. The shift in the mass is measured to be $-0.7\% \pm 0.02\%(\text{stat.}) \pm 0.9\%(\text{syst.})$ in the barrel and $+2.5\% \pm 0.2\%(\text{stat.}) \pm 2.2\%(\text{syst.})$ on average in the endcaps.

As the size of the current data-set is insufficient for a measurement of the energy resolution, this has so far only been studied with test beam data [26]. A complete supermodule from the barrel was tested using electrons with momenta between 20

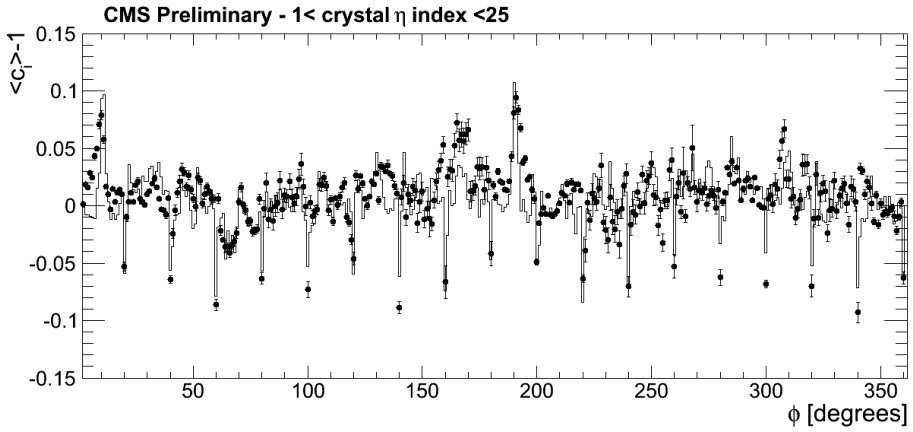
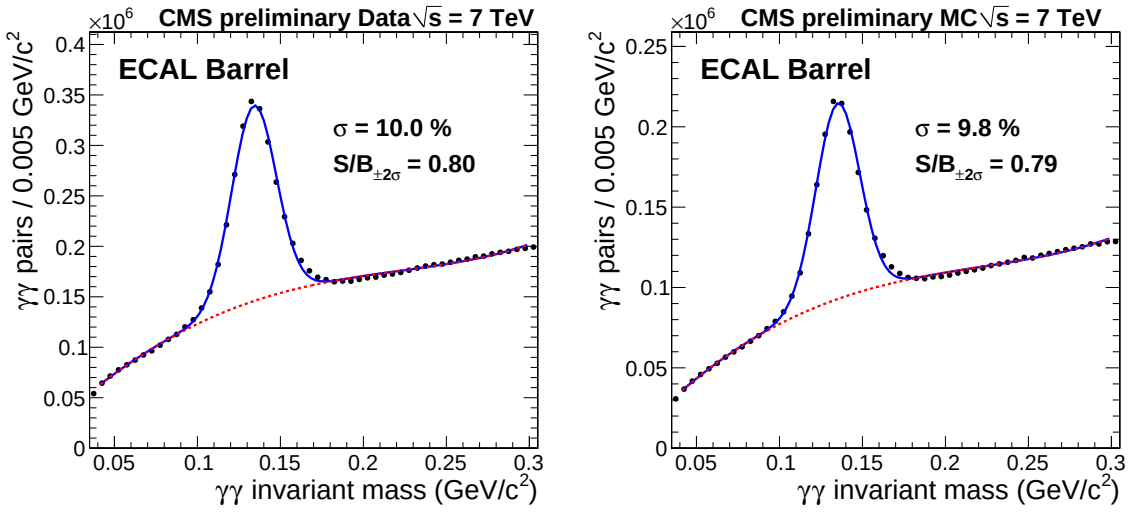


Figure 2.5: The average difference from unity of the inter-calibration constants measured in data (solid circles) and from simulation (histogram). The crystal index corresponds to a pseudorapidity range of $0 < \eta < 0.5$.

and 250 GeV/c in the H4 test beam at CERN. The energy resolution is parameterised as in equation 2.3, where S is the stochastic term, N is the noise term, and C is the constant term.

$$(\frac{\sigma}{E})^2 = (\frac{S}{\sqrt{E}})^2 + (\frac{N}{E})^2 + C^2 \tag{2.3}$$

The energy resolution as measured by the test beam data is shown in Figure 2.7.



(a) (b)

Figure 2.6: The reconstructed π^0 invariant mass for data (a) and simulation (b).

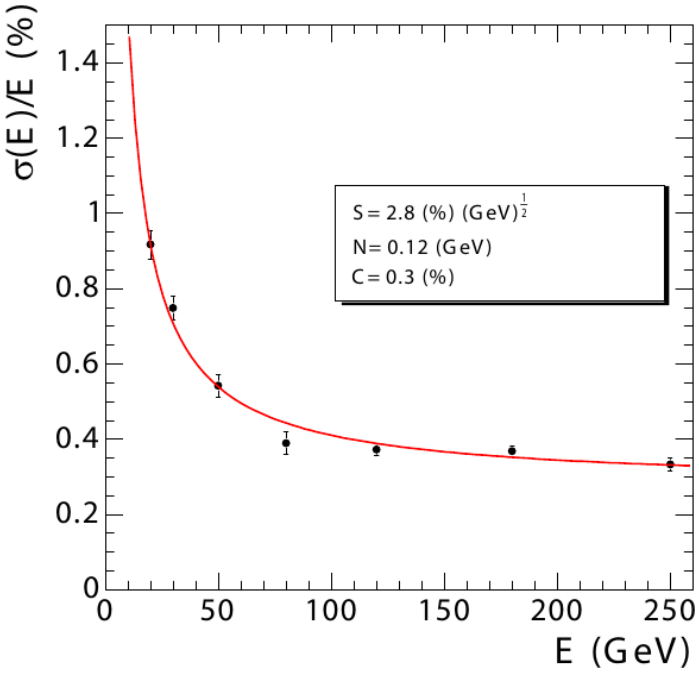


Figure 2.7: The energy resolution for electrons with momenta between 20 and 250 GeV/c measured in the H4 test beam.

2.4 Hadronic calorimeter

2.4.1 Overview

The hadronic calorimeter (HCAL) is a sampling calorimeter separated into four distinct sub-detectors. The barrel (HB) and endcap (HE) calorimeters sit inside the solenoid magnet. Outside of the solenoid are the outer (HO) and forward (HF) calorimeters. The location of the various sub-detectors that compose the HCAL is illustrated in Figure 2.8.

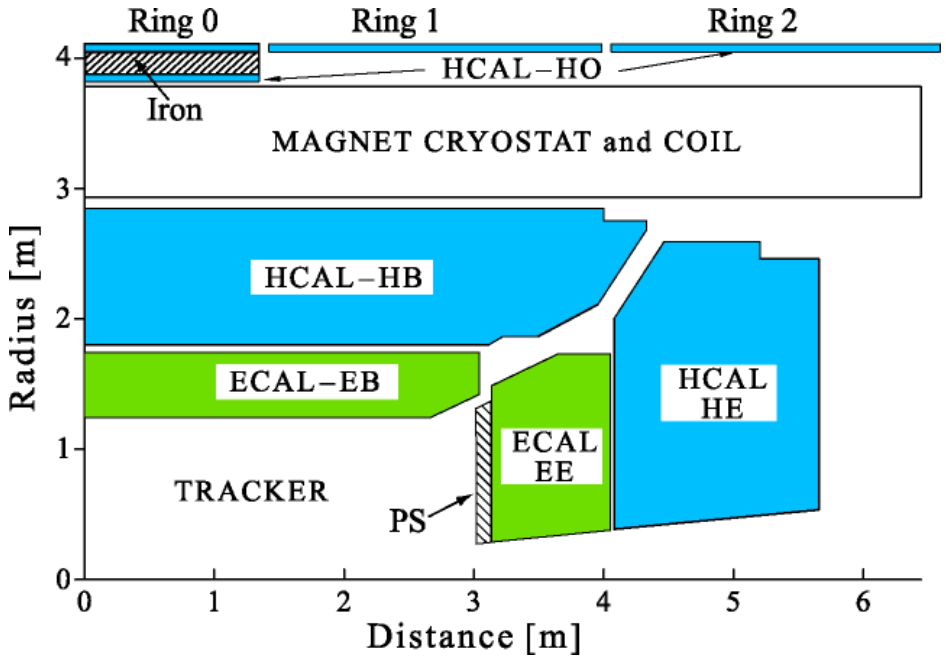


Figure 2.8: A schematic diagram of the hadron calorimeter showing the location of the hadron barrel, outer, and endcap calorimeters.

The hadron barrel calorimeter is separated into two half-barrels that cover pseudorapidities up to $|\eta| < 1.3$. Each half-barrel is segmented into eighteen 20° wedges in ϕ referred to as a tower. The towers are constructed from brass absorber plates interspersed with a plastic scintillator. The innermost and outermost absorber plates are constructed from steel to improve structural integrity. There is an additional block

of plastic scintillator situated directly behind the electromagnetic calorimeter to measure early showers, giving a total of seventeen active scintillator layers in a tower. Longitudinally, the scintillator blocks are divided into 16 η regions. The individual scintillator tile size is $\Delta\phi\times\Delta\eta = 0.087\times0.087$. The light from each plastic scintillator is extracted using wavelength shifting fibre before being directed down clear fibre along the length of the barrel. The light from each block in a tower is added optically and then measured with a hybrid photo-diode. The total absorber thickness corresponds to 5.82 interaction lengths at $\eta = 0$ which increases at higher pseudorapidities to a maximum of 10.6 interaction lengths. The electromagnetic calorimeter adds another 1.1 interaction lengths.

The hadron barrel calorimeter is complemented by the outer calorimeter which covers the same pseudorapidity region but lies outside of the solenoid magnet. The structure of the outer calorimeter is constrained to match that of the muon system (see Section 2.5) and as such is separated into 5 rings numbered as -2, -1, 0, +1, and +2. Each of these rings is instrumented with one or two blocks of plastic scintillator divided to match the segmentation in barrel calorimeter. The central region (ring 0) has the smallest number of interaction lengths so it makes use of two scintillator blocks which sandwich a block of iron absorber to increase the number of interaction lengths. The outer rings (±1 and ±2) make use of a single block of plastic scintillator.

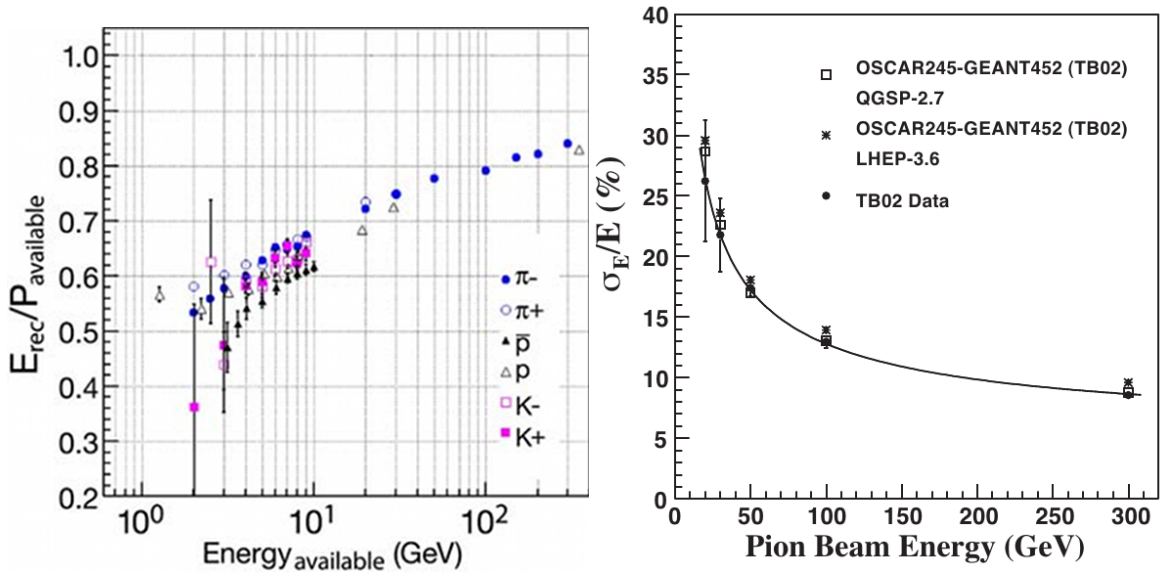
The hadronic endcap calorimeters cover the pseudorapidity region of $1.3 < |\eta| < 3.0$. The design uses the same technology as that of the barrel calorimeter with a segmentation in ϕ which also matches that of the barrel. For each of the towers in the endcap there are nineteen active layers of plastic scintillator. For the pseudorapidity range of $|\eta| < 1.74$, the plastic scintillator tile dimensions match those of the barrel in terms of $\Delta\phi\times\Delta\eta$. At higher pseudorapidities, the resolution is halved in ϕ to accommodate the bending of the fibres.

2.4.2 Performance

Studies of the performance of the hadron calorimeter have been conducted using test beam data, cosmic muons and also "beam splash" events in which the beam is directed to a collimator directly in front of CMS producing a shower of particles that reach the detector. Performance studies with the 2010 data-set are still on-going.

The initial calibration of each of the scintillating tiles in the HCAL is carried out by using a Co^{60} source [27]. This establishes the inter-calibration constants for each of the different tiles. The absolute energy scale has to be measured separately which was done using test beam data. The test beam used for this is the H2 beam line at CERN [28, 29]. The test beam delivers particles with a momentum range from 2 to 350 GeV/c incident on a quarter section of the HCAL barrel, endcap and outer calorimeters. In front of the section of the HCAL, a super-module from the ECAL is also included. A plot of the measured energy response for the combined sections of the ECAL and HCAL as a fraction of the available particle energy is shown in Figure 2.9 along with the fitted energy resolution for pions in comparison to several GEANT4 simulations. The fitted result to the test beam data gives an energy resolution of $(\sigma/E) = \frac{115.3}{\sqrt{E}} \oplus 5.5\%$.

The calibration of the HCAL modules was further studied during the 2008 Cosmic Run at Four Tesla (CRAFT) exercise using cosmic muons [30]. Muons will leave a small energy signature in HCAL towers as they pass through the detector, however as they are travelling through the detector vertically it is not possible to use these to calibrate those towers which are arranged horizontally at $\phi = 0°$ and $\phi = 180°$. The mean energy and RMS of the energy deposits from the cosmic muons is shown in Table 2.1 for two different ϕ sections for the two halves of the HCAL barrel before and after corrections with cosmic muons.



(a) (b)

Figure 2.9: The energy response for different particles (a), and the energy resolution of pions (b), reconstructed with the ECAL and HCAL modules at the H2 test beam.

Table 2.1: Mean and RMS of the muon energy distributions measured in four regions of the HCAL barrel before and after correcting the calibration constants using cosmic ray muon data.

Detector Region	Mean uncorrected (GeV)	RMS uncorrected (GeV)	Mean corrected (GeV)	RMS corrected (GeV)
HB-: $20° < \phi \leq 160°$	1.87 ± 0.03	0.12	1.78 ± 0.01	0.05
HB-: $190° < \phi \leq 370°$	1.83 ± 0.02	0.08	1.75 ± 0.02	0.07
HB+: $20° < \phi \leq 160°$	1.81 ± 0.02	0.08	1.78 ± 0.01	0.05
HB+: $190° < \phi \leq 370°$	1.70 ± 0.02	0.09	1.74 ± 0.01	0.04

2.5 Muon system

2.5.1 Overview

The muon system at CMS is comprised of three separate sub-detectors. The layout is illustrated in Figure 2.10.

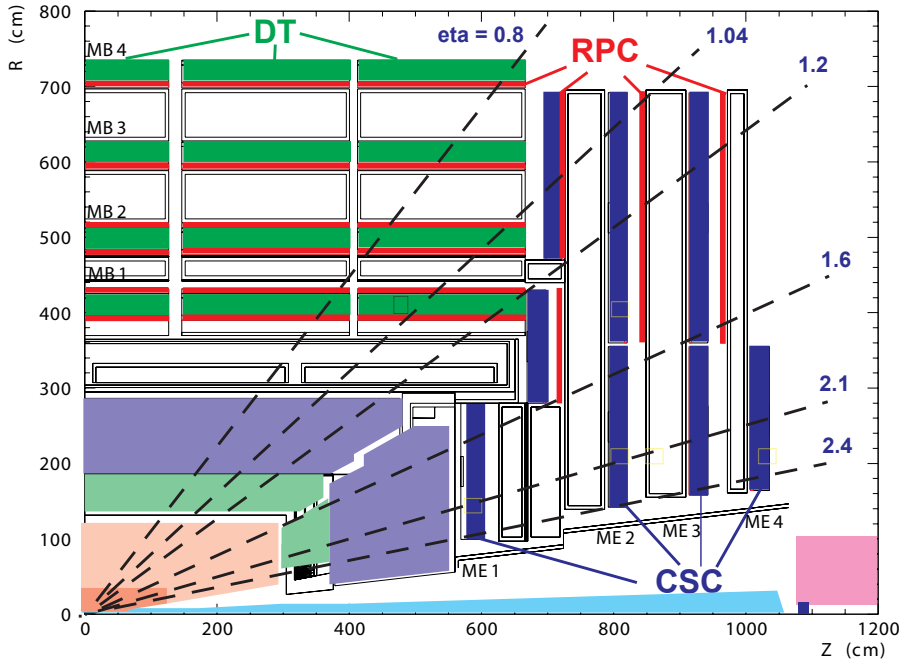


Figure 2.10: A schematic diagram of the muon system showing the drift tubes (DTs) in the barrel, the cathode strip chambers (CSCs) in the endcaps and the resistive plate chambers (RPCs).

The barrel muon system covers a pseudorapidity up to $|\eta| < 1.2$ and employs drift tubes to detect the presence of a passing muon. The drift tubes are grouped into four separate stations arranged as concentric cylinders around the beam-line. The inner three stations are composed of eight chambers separated into two groups of four for measuring the $r - \phi$ plane and four chambers for measuring the z coordinate. The two groups of $r - \phi$ measuring chambers are separated as far as possible to offer the

best angular resolution. The final outer station does not have the four chambers for measuring the z coordinate. In total, there are sixty drift chambers used in the inner three stations, and seventy used in the outer station.

The endcap muon system covers the pseudorapidity region $0.9 < |\eta| < 2.4$ overlapping with the barrel region. This section is instrumented with cathode strip chambers. The cathode strips are organised in a radial configuration giving measurements in the $r-\phi$ plane. The anode wires are orientated approximately perpendicular to the cathode strips providing a measurement of the polar angle. The wires and strips are organised in chambers perpendicular to the beam-line, with chambers being combined into a disk structure.

In addition to the drift tubes and cathode strip chambers, the muon system employs a third detector which is installed in the barrel region and partially covers the end cap region. The position of the resistive plate chambers (RPCs) in relation to the other muon systems is illustrated in Figure 2.10 with coverage extending up to pseudorapidities of $|\eta| < 1.6$. The RPCs are gaseous parallel plate chambers, which in particular, provide excellent timing resolution of the order of a nanosecond. A chamber consists of two parallel electrodes constructed from plastic coated with a layer of graphite paint. The primary role of the RPCs is for triggering, although they do provide adequate spacial resolution to be used in track reconstruction.

As the muon system is obviously concerned with the identification and reconstruction of a single physics object, the performance of the muon identification at CMS will be discussed with the other physics objects in Section 4.3. The details of the sub-detector performance itself have been extensively studied primarily with cosmic muons which can be found elsewhere [31, 32, 33].

2.6 Trigger system

The trigger system at CMS is required to reduce the high interaction rates of the LHC down to a rate at which selected events can be completely processed and stored. This is done typically by performing a fast but basic reconstruction to find physics objects that pass a certain minimum threshold. In this way events with only low energies that would be of little interest for analysis at the LHC are rejected. The trigger selection is also the first selection step for any analysis so it also plays a part in background rejection. This is a general discussion of the trigger system, the more specific example of the triggers used in this analysis can be found in Section 5.2.1.

2.6.1 Level-1 trigger

The trigger is separated into two steps; the Level-1 Trigger (L1) and the High-Level trigger (HLT). The event must first be accepted by the L1 trigger; an overview of the operating steps is given in Figure 2.11. This system reduces the rate of events to 30 kHz from a maximum collision rate of 40 MHz. This is accomplished largely by programmed electronics either on the detector itself or located nearby in the cavern area. As the L1 trigger has to process a very high rate of data it is not possible to use the full set of sub-detectors in CMS, as such it makes use of information from the calorimeters and the muon sub-systems only.

The calorimeters are divided into regions known as trigger towers for this purpose. A trigger tower consists of a readout from both the ECAL and the HCAL in a region of the detector in which the ECAL crystals are segmented to match the HCAL towers. The trigger towers are passed to the Regional Calorimeter Trigger which provides a basic reconstruction of electron and photon candidates. The candidates are passed to the Global Calorimeter trigger. At this stage jets are reconstructed and the missing transverse energy is calculated along with other global quantities. This information then passes to the Global Trigger.

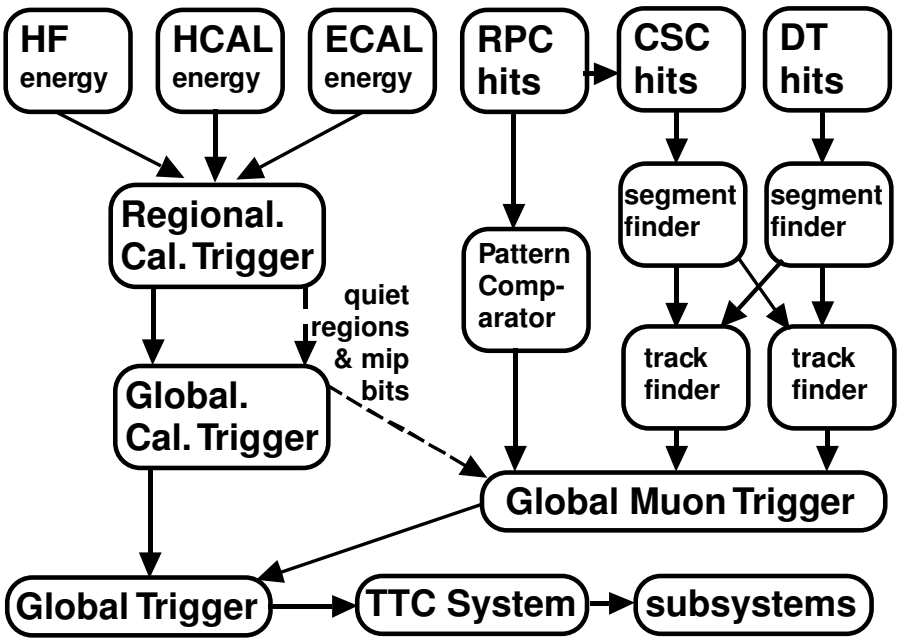


Figure 2.11: A flow chart illustrating the inputs and processing steps for the CMS Level-1 Trigger system.

The muon trigger makes use of all three of the muon sub-systems to reconstruct muon candidates. The coverage for triggering purposes extends up to pseudorapidities of $|\eta| < 2.1$. The sub-systems provide the spatial location of hits which are fitted to provide tracks for muon candidates. The resistive plate chambers fulfil a special role in triggering. They are designed specifically to offer excellent time resolution of the order of 1 ns, allowing for the precise identification of the bunch crossing which produced the collision. The muon candidates are then passed to the Global Trigger.

The final step for the L1 trigger is the Global Trigger. This uses all of the information passed from the individual trigger steps to make a final decision to pass the event to the HLT.

2.6.2 High level trigger

The High Level Trigger makes the final decision as to whether the event will be permanently read out and stored. The final goal is to reduce the rate down to 100 Hz. The HLT has access to all the information in the event as well as the information from the L1 trigger. At this stage a full reconstruction of the event is performed using the online software which allows for triggering on a large number of criteria.

Chapter 3

Silicon Tracker Performance

The commissioning of the CMS tracker is a continuous process which makes use of data collected from collisions as well as data collected from cosmic muons during periods without beam at the LHC. The initial commissioning phase with cosmic muons took place before the 2010 data taking period using the outer muon chambers to trigger the read-out of the sub-detectors [34, 35, 36].

3.1 Tracking efficiency

The efficiency to reconstruct the tracks of charged particles has been measured in data using a number of techniques [37].

The tracking efficiency for isolated muons and pions has been measured using a simulated track embedding technique. The isolated muons and pions are generated using CMS simulation software after which the tracks are reconstructed using the CMS standard reconstruction process [38]. The reconstructed simulated track is then embedded in a data event taken with a minimum bias trigger. The advantages of this method are that the track momenta can cover the full range and that the overall efficiency can be separated into the *acceptance* and the *reconstruction efficiency*. The acceptance ($\mathcal{A}$)

is defined as the probability that the charged particle will produce a sufficient number of hits in the tracker to reconstruct a track, and the reconstruction efficiency (ε) is defined as the probability that the reconstructed track matches the original simulated particle. The tracking efficiency is then measured using equation 3.1 from the number of generated particles (N_{gen}), the number of reconstructed tracks with ($N_{\text{reco,embed}}$) and without ($N_{\text{reco,iso}}$) being embedded in an event. The results of this study are shown in Figure 3.1.

$$\mathcal{P}_{\text{track}} = \mathcal{A} \cdot \varepsilon = \frac{N_{\text{reco,iso}}}{N_{\text{gen}}} \cdot \frac{N_{\text{reco,embed}}}{N_{\text{reco,iso}}} \tag{3.1}$$

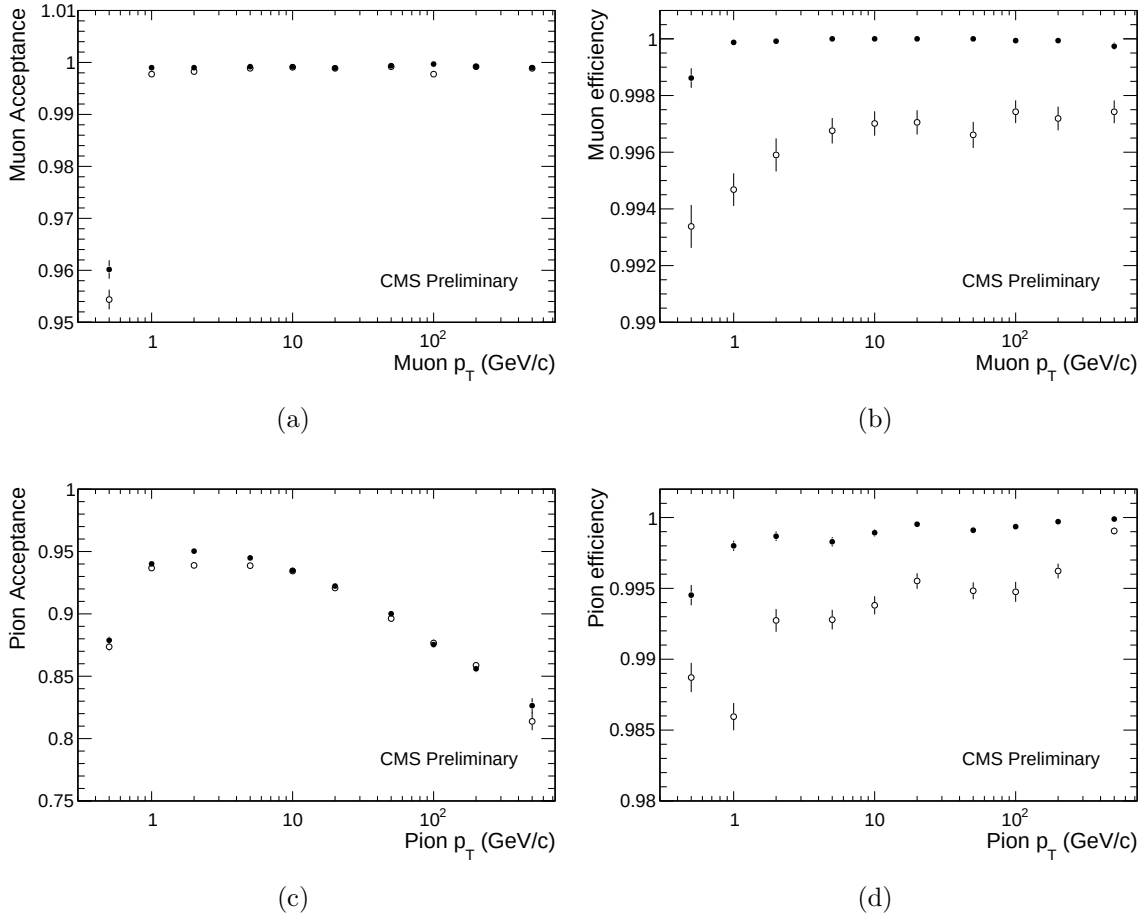


Figure 3.1: The acceptance and efficiency for isolated muons ((a), (b)) and pions ((c), (d)) against the transverse momentum of the simulated particle. Open circles show the results for simulated tracks embedded in data events and closed circles show the results from using only simulation.

The track reconstruction efficiency of isolated muons has also been studied using a "tag-and-probe" method utilising the J/ψ resonance. The tag muon is a high quality candidate reconstructed with a global fit to the muon and tracker systems. It is required to pass the muon trigger requirements with a momentum threshold of 3 GeV/c and have a minimum transverse momentum of 2.6 GeV/c as measured by the tracker. The separate cuts are imposed as the muon trigger does not make full use of the inner layers of the tracker. The probe muon is reconstructed with the muon system alone. The probe is considered to "pass" if it also matches a track in the tracker with the requirements of $|\Delta\eta| < 0.2$ and $\Delta R < 0.5$. The invariant mass of the tag and the probe muon is then reconstructed. The combined efficiency of the tracking and the matching to the muon system is then calculated from the measured yields of unbinned likelihood fits to the sub-samples of passing and failing probes. The results of the fits to the J/ψ mass are shown in Figure 3.2. From the combined efficiency ε, the tracking

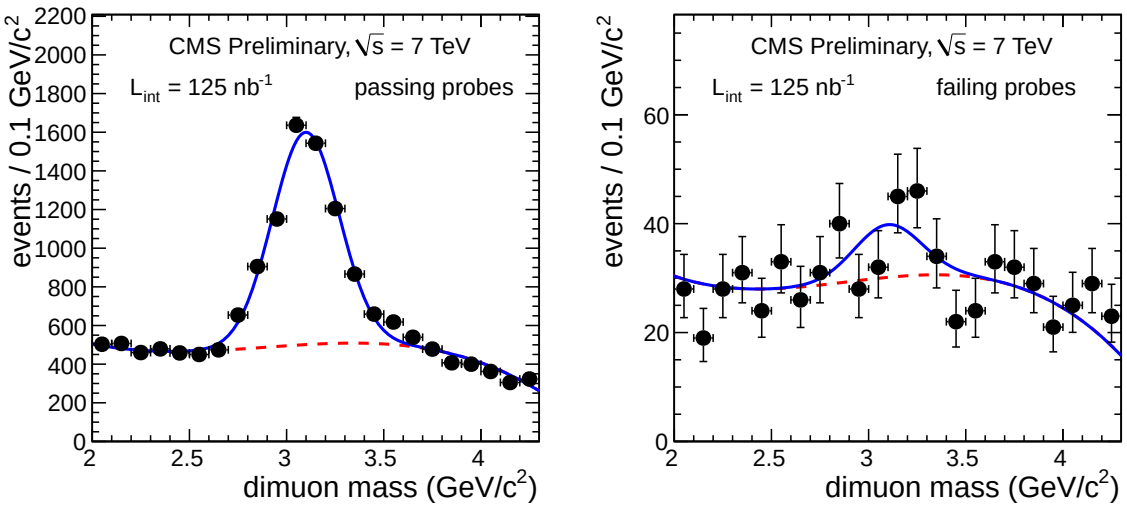


Figure 3.2: The reconstructed J/ψ invariant mass distribution with probes that pass (left) and probes that fail (right) the track matching requirement.

efficiency ε_{T} can be calculated with the efficiency to match fakes ε_{F} from equation 3.2. The muon matching efficiency ε_{M} is considered to be nearly 100% efficient and as such

can be neglected [37].

$$\varepsilon_{\mathrm{T}}\varepsilon_{\mathrm{M}} = \frac{\varepsilon - \varepsilon_{\mathrm{F}}}{1 - \varepsilon_{\mathrm{F}}} \tag{3.2}$$

To measure the rate of fake matches for the probe muons, events are altered before applying any track matching to the probe in which all tracks that reconstruct an invariant mass with the tag muon close to that of the J/ψ are removed. The rate at which probe muons match the remaining tracks in the collection gives a direct measurement of the fake matching efficiency. The measured tracking efficiency separated into $|\eta|$ bins is given in Table 3.1.

Table 3.1: The measured tracking efficiency of isolated muons using the tag-and-probe method for data and simulation.

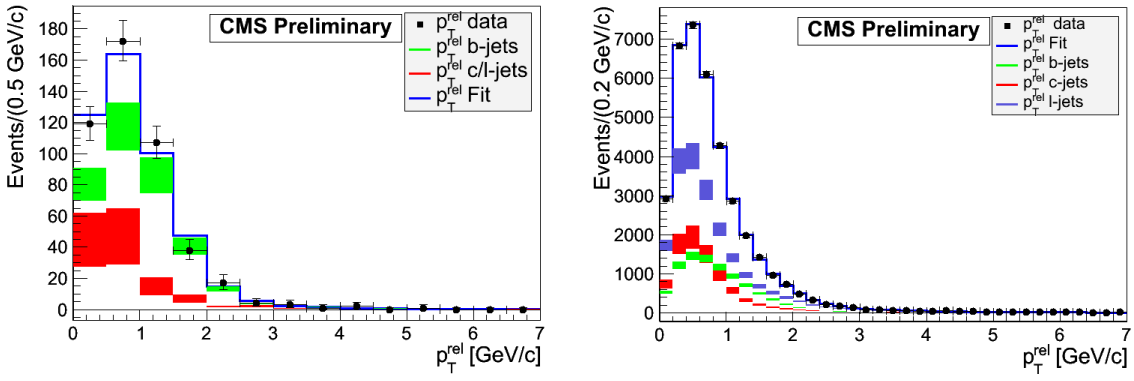
Region	Data Eff. (%)	Sim Eff. (%)	Data/Sim
$0.0 \leq \lvert\eta\rvert < 1.1$	$100.0^{+0.0}_{-0.3}$	$100.0^{+0.0}_{-0.3}$	$1.000^{+0.001}_{-0.003}$
$1.1 \leq \lvert\eta\rvert < 1.6$	$99.2^{+0.8}_{-1.0}$	$99.8.0^{+0.1}_{-0.1}$	$0.994^{+0.009}_{-0.010}$
$1.6 \leq \lvert\eta\rvert < 2.1$	$97.6^{+0.9}_{-1.0}$	$99.3^{+0.1}_{-0.1}$	$0.983^{+0.009}_{-0.010}$
$2.1 \leq \lvert\eta\rvert < 2.4$	$98.5^{+1.5}_{-1.6}$	$97.6^{+0.2}_{-0.2}$	$1.010^{+0.015}_{-0.016}$
Combined	$98.8^{+0.5}_{-0.5}$	$99.2^{+0.1}_{-0.1}$	$0.996^{+0.005}_{-0.005}$

The tracking efficiency has also been measured for non-isolated muons using heavy flavour decays and light-hadron decays in flight. This result is significant as it is a direct measure of the efficiency in environments of higher density such as jets which is important for b-tagging (see Section 6). This measurement uses a data sample consisting of muon candidates from the muon system and two jets. One of the jets (referred to as the *muon jet*) is required to lie within a cone of radius $\Delta R = 0.4$ with respect to the muon candidate. The second jet is required to pass a high-purity b-tagging algorithm and to also be separated from the muon-jet by at least $\Delta R = 1.5$. The purpose of this is to create a heavy flavour enriched di-jet sample.

The composition of the final sample still contains a significant amount of muons originating from light-hadron decays (estimated to be 20% from simulation). As the tracking efficiency for these muons differs from those originating from heavy flavour decays due to the difference in lifetime, they have to be measured separately. This is performed by measuring the heavy flavour fractions using a distribution which discriminates between the flavour of the jets known as P_t^{rel}. The variable P_t^{rel} is defined as the transverse momentum of the muon relative to the muon-jet axis. Using this, the fraction of b-jets, c-jets and light-jets are measured in a second sample in which the b-jet fails the b-tagging algorithm. For the heavy flavour enriched sample the number of events that remain are too small to fit the c-jet fraction separately so it is combined and fitted with the light-jet fraction. The separate c- and light-jet fractions are then calculated from their ratio in simulation. The results of these fits are shown in Figure 3.3 with the templates for the fits being taken from simulation.

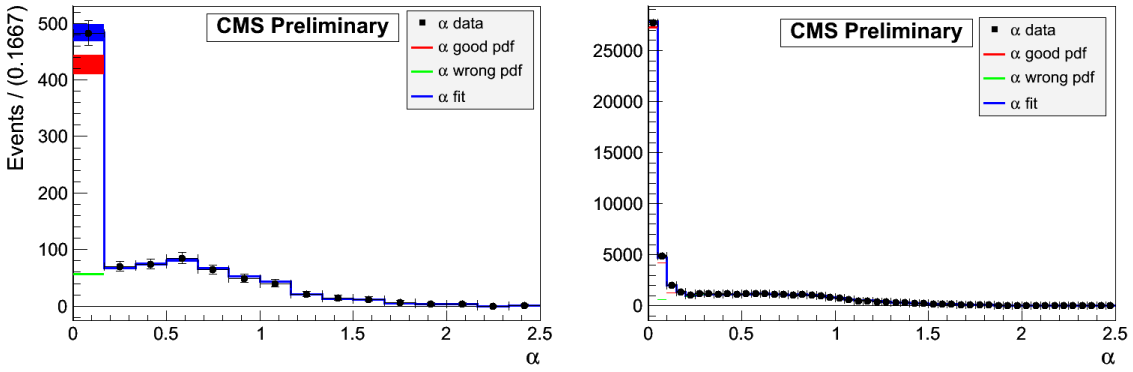
With the flavour fraction established, the uncorrected tracking efficiency is then measured with the muon candidates. This is performed by using template fits to a matching variable α, defined as the distance between the extrapolated track and position of the hit in the muon detector closest to the interaction region, divided by the distance from the muon hit to the centre of CMS. The templates of α for correctly matched muons are derived from a high purity sample of muons reconstructed with both the muon system and the silicon tracker. The templates for incorrectly matched muons

are derived from α distributions using sets of tracks that exclude the track of the muon candidate. These templates are then fitted to the distribution for both the tagged non-muon jet and un-tagged non-muon jet samples (Figure 3.4) to give the uncorrected tracking efficiencies in these samples. The final result is then calculated by solving a pair of equations simultaneously using the measured flavour fractions and uncorrected tracking efficiencies for both samples to give a heavy flavour tracking efficiency of $\varepsilon_{bc} = (93.2 \pm 5.3)\%$.



(a) (b)

Figure 3.3: The P_t^{rel} distribution for the b-tagged non-muon jet (a) and un-tagged non-muon jet (b). The distributions are fitted for the relative contribution of heavy and light flavour.



(a) (b)

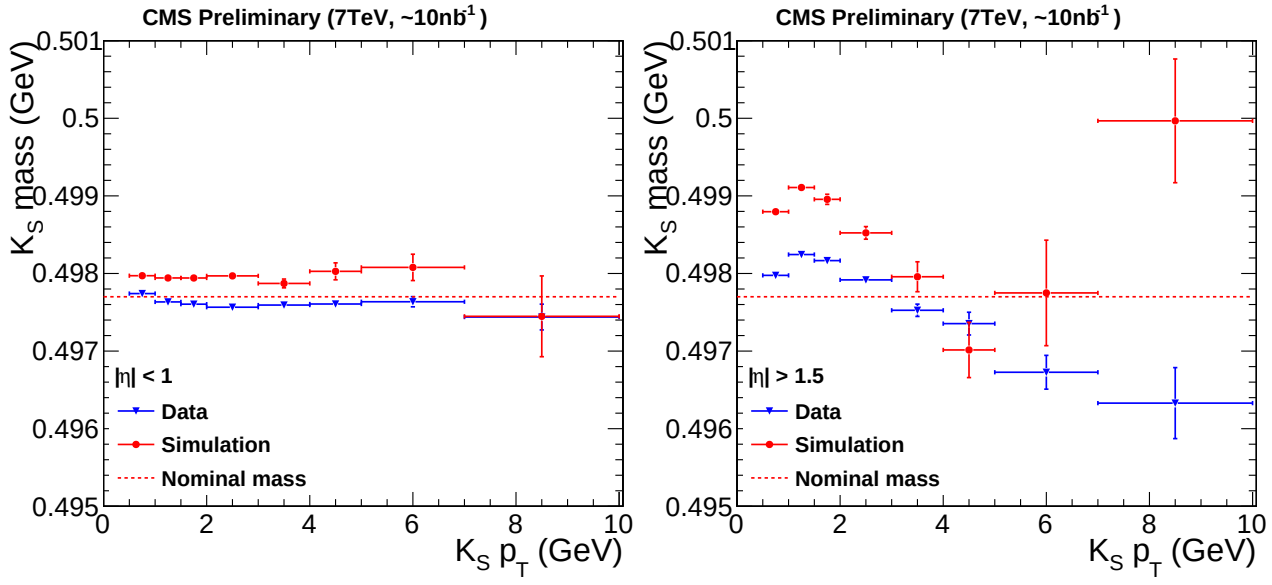
Figure 3.4: The α distribution for the b-tagged non-muon jet (a) and un-tagged non-muon jet (b). The "good" and "wrong" PDF (probability distribution function) refer to the α distribution for correctly matched muons and incorrectly matched muons, respectively.

3.2 Track momentum scale and resolution

The track momentum scale and resolution have been studied in data by reconstructing the $K_s \to \pi^+\pi^-$, $\phi \to K^+K^-$ and $J/\psi \to \mu^+\mu^-$ resonances [39]. In all cases the tracks must fulfil a number of identification criteria which are detailed in Ref. [39].

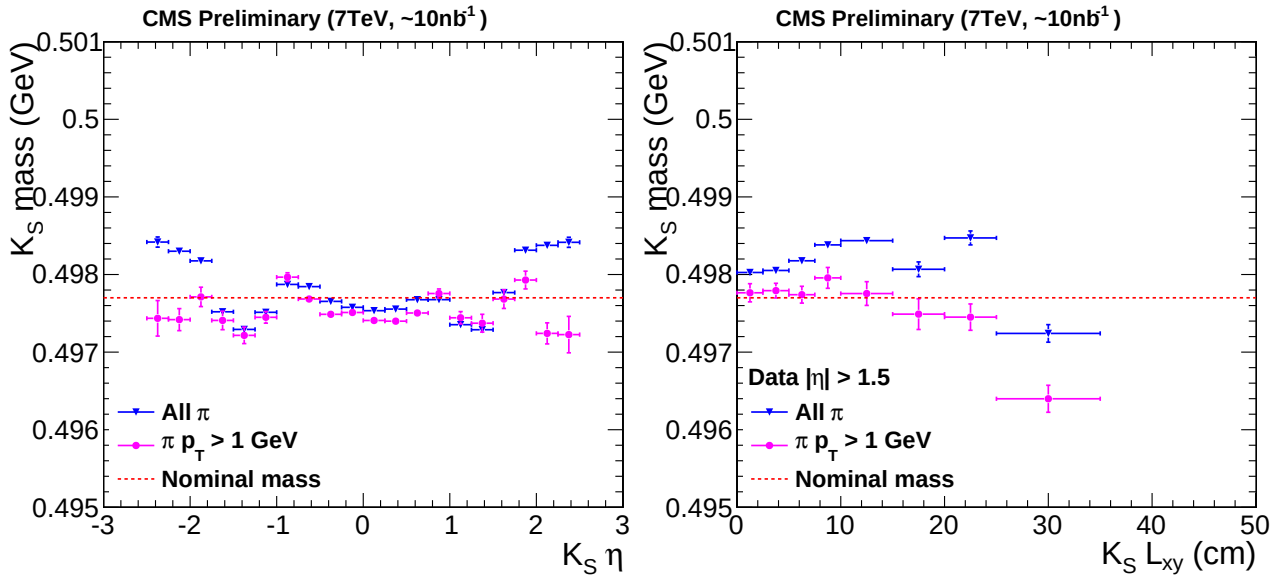
The K_s and ϕ resonances are primarily used to investigate biases arising in the reconstruction from effects such as an incorrect material budget description and misalignment. For both resonances, the invariant mass of the decay products is reconstructed in a number of different bins of transverse momentum. The K_s distribution is fitted with a Gaussian function and the background is modeled with a quadratic polynomial. The ϕ distribution is fitted with the convolution of a Gaussian and a relativistic Breit-Wigner function and the background is modeled with an arctangent. The results for the K_s resonance parameterised in the transverse momentum of the reconstructed K_s, pseudorapidity and the projection of the transverse decay length onto the momentum of the reconstructed K_s (L_{xy}) are shown in Figure 3.5 and Fig. 3.6. The discrepancies are attributed to the decays producing low momentum pions which are also more prevalent in the endcap region.

The results of the fit to the ϕ distribution against the transverse momentum are shown in Figure 3.7 along with the resulting resolution from the fit.



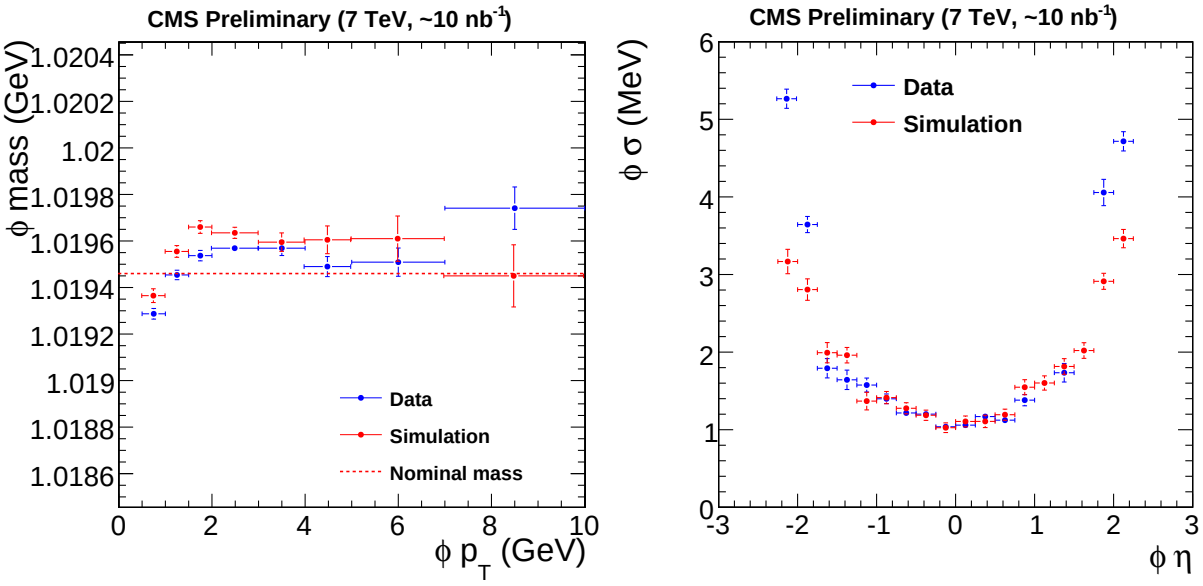
(a) (b)

Figure 3.5: The fitted value of the K_s mass in comparison to the nominal value versus transverse momentum in the barrel (a) and endcap(b).



(a) (b)

Figure 3.6: The fitted value of the K_s mass in comparison to the nominal value versus pseudorapidity (a) and L_{xy} (b). The distributions are shown with (magenta) and without (blue) a 1 GeV cut on the pion transverse momentum.



(a) (b)

Figure 3.7: The fitted value of the ϕ in comparison to the nominal result versus the transverse momentum (a). The resolution from the fit to the ϕ mass versus pseudorapidity (b).

The J/ψ analysis uses a more complex technique which aims to measure the momentum resolution and scale simultaneously as they are intrinsically linked. To do this, the observed mass is parameterised as a function of the momentum scale and momentum resolutions in transverse momentum, ϕ and $\cot\theta$. The parameters are then fitted to the invariant mass distribution using an unbinned maximum likelihood fit. This analysis is described in greater detail elsewhere [39]. The results for the momentum resolution measured using the J/ψ resonance are shown in Figure 3.8.

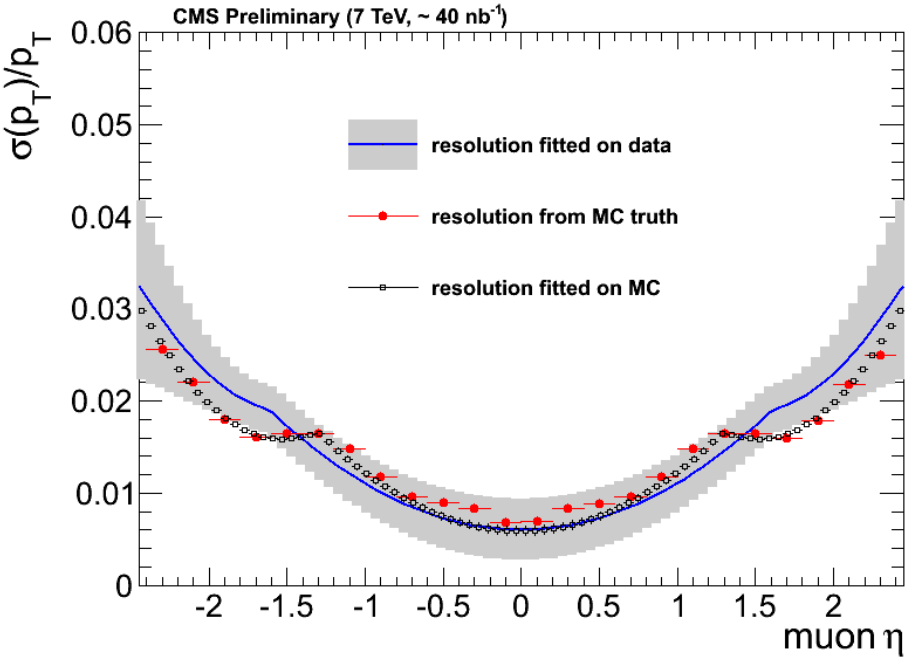


Figure 3.8: The measured transverse momentum resolution measured with the J/ψ resonance parameterised in bins of pseudorapidity of the muon decay products. The blue line is the measured result on data with the grey band being the uncertainty, the red points are the results for the resolution calculated with simulation information and the black points are the result of this analysis technique applied to simulated samples.

3.3 Primary vertex reconstruction

The characteristics of the primary vertex reconstruction were measured in the early stages of running at 7 TeV [40]. The primary vertex resolution and reconstruction efficiency are measured by separating the tracks into two sets for each event. For the resolution measurement, the tracks are split evenly and for the reconstruction efficiency a "tag" set is constructed with two thirds of the tracks and a "probe" set is given the remaining one third. In each case the tracks are ranked first by transverse momentum of the track before being distributed randomly into each set to ensure that, on average, the track momentum distribution of each set is similar. Each set of tracks is then passed independently to the adaptive vertex fitter to reconstruct a primary vertex.

The primary vertex resolution is found by fitting a Gaussian to the difference in position of the two reconstructed vertices with the resolution being taken from the width. The results for the x, y, and z coordinates as a function of the number of tracks in the event are shown in Figure 3.9.

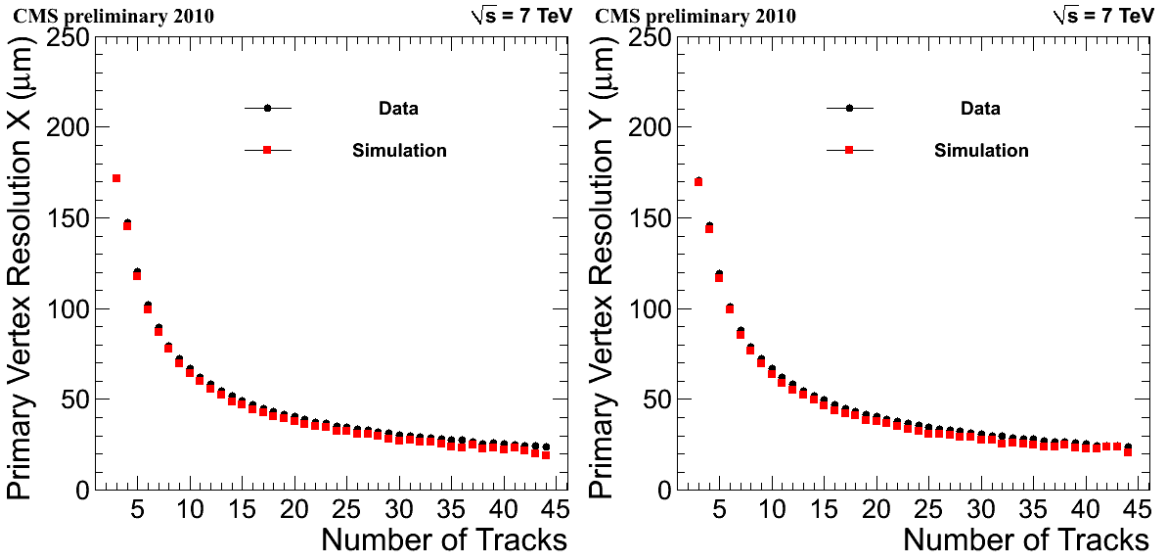
The primary vertex reconstruction efficiency is found from how often the "probe" vertex is considered consistent with the "tag" vertex after their independent reconstruction. The vertices are considered to be consistent if they lie within 5σ of each other in z, where σ is chosen to be the largest uncertainty in z returned by the adaptive vertex fitter for either of the two sets of tracks. The results for the primary vertex reconstruction efficiency are shown in Figure 3.10.

As well as the reconstruction efficiency and resolution of the primary vertex, the resolution of the track impact parameter has also been measured in data. This measurement is performed by selecting high quality tracks consistent with having being produced promptly in the collision event. The primary vertex reconstruction is then performed without including the selected track in the event reconstruction. The impact parameter is then calculated with these track-vertex pairs. The final distribution is smeared

by a combination of three effects:

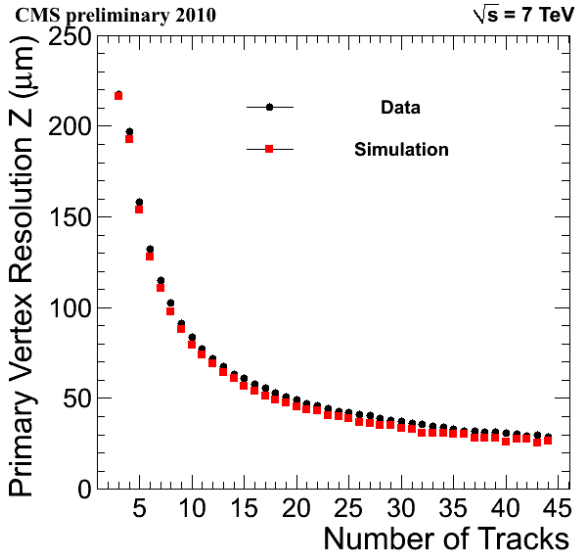
- The uncertainty on the track impact parameter.
- The uncertainty on the primary vertex position.
- The fraction of incorrectly selected tracks which do not correspond to prompt production (for example heavy flavour decays). This pollution of the prompt track sample corresponds to less than 2%(10%) for transverse momenta less than 4 GeV/c (20 GeV/c) according to simulation studies.

As the effect of the non-prompt track pollution is small, it is neglected. The distribution is fitted with the convolution of two Gaussian functions. The width of the first Gaussian is fixed to the primary vertex position uncertainty derived from simulation and the width of the second Gaussian corresponds to the impact parameter resolution which is the result of the fit. The results are shown for the transverse and longitudinal impact parameter resolutions against the pseudorapidity of the track in Figure 3.11.



(a)

(b)



(c)

Figure 3.9: The resolution of the primary vertex reconstruction as measured on data in x (a), y (b) and z (c).

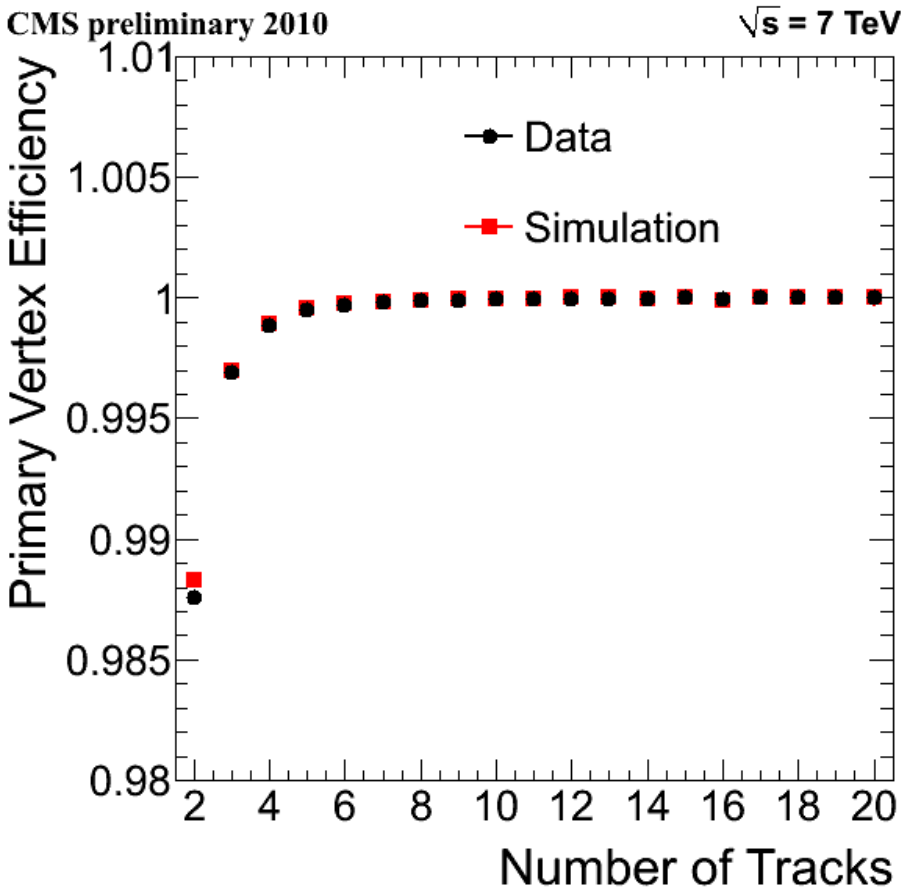
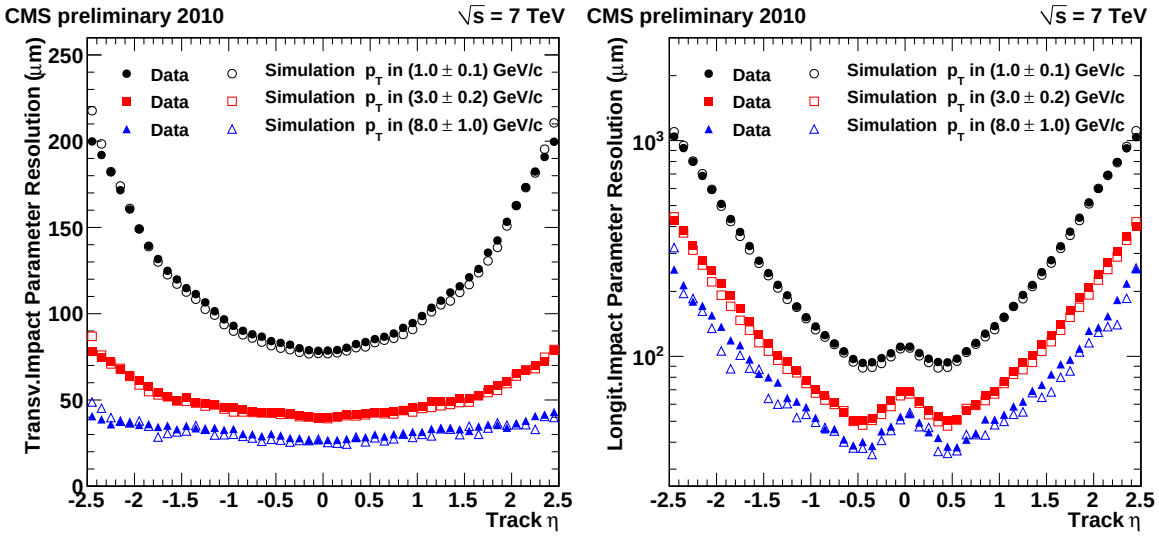


Figure 3.10: The reconstruction efficiency of the primary vertex as measured on data.



(a) (b)

Figure 3.11: The transverse (a) and longitudinal (b) impact parameter resolution for three different track transverse momentum ranges.

3.4 Tracker material budget

The high density of silicon and large tracking volume within CMS comes with a cost. There are a significant number of radiation lengths present before the electromagnetic calorimeter, leading to photon conversions being the primary cause of fake, prompt, isolated electrons. As such it is necessary to ensure that the material of the tracker is very precisely simulated through the measurement of the material in data.

The material in the tracker can effectively be mapped by looking for the characteristic signatures of photon conversions or nuclear interactions [41]. Photon conversions are identified by looking for pairs of oppositely charged tracks with a small separation and a small opening angle between them; this reconstruction relies solely on tracking information. Once a pair of tracks have been identified as a conversion candidate, the tracks are then passed to a 3D-constrained kinematic vertex fitter algorithm [42]. If the fit converges and returns a χ^2 probability $> 5\text{x}10^{-4}$ as determined in the fitting procedure, the candidate is accepted as a conversion. The location of the conversion vertices in the $x - y$ plane are shown in Figure 3.12 for $|z| < 26$ cm, in which the structure of the CMS tracker can clearly be seen.

The method used for the identification of nuclear interactions proceeds as follows. Tracks that are separated by a sufficiently small distance of closest approach are linked together in a "block". After being linked, the tracks in a block are all refitted together under the constraint of a displaced vertex. Nuclear interactions are required to have a relatively large vertex displacement ($r > 2.5$ cm). The distribution of the resulting nuclear interactions in pseudorapidity is shown in Figure 3.13.

Using the results for the number of photon conversions and nuclear interactions it is possible to derive information on the material budget. This process is detailed elsewhere [41]. The final results for the material budget are shown in Figure 3.14 measured with photon conversions, and in Figure 3.15 measured with nuclear interactions.

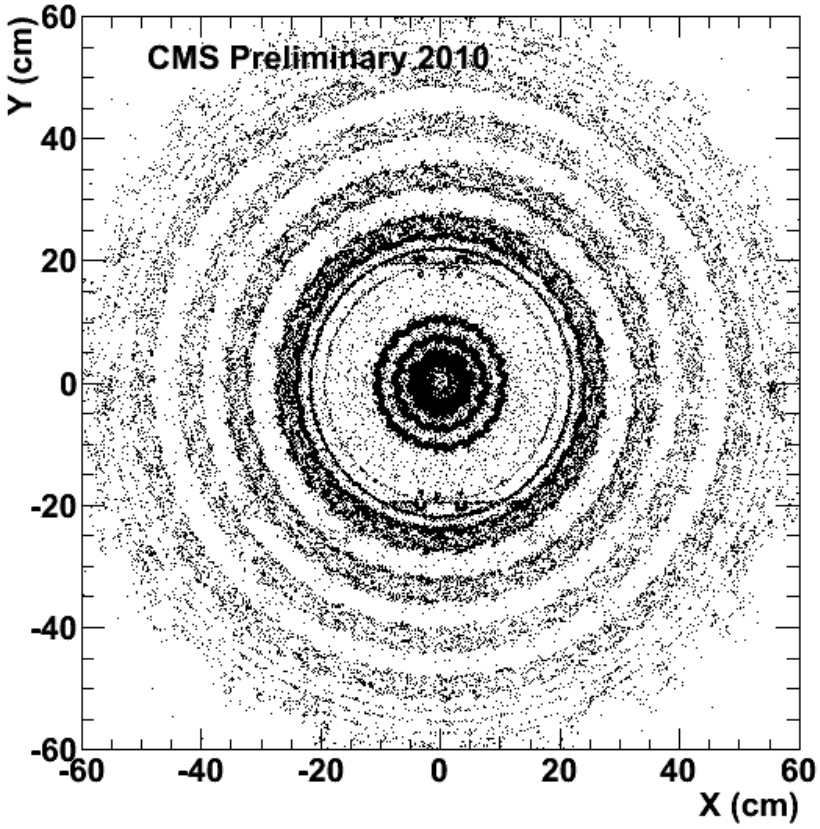


Figure 3.12: The location of the reconstructed conversion vertices within the CMS tracker in the $x - y$ plane for $|z| < 25$ cm.

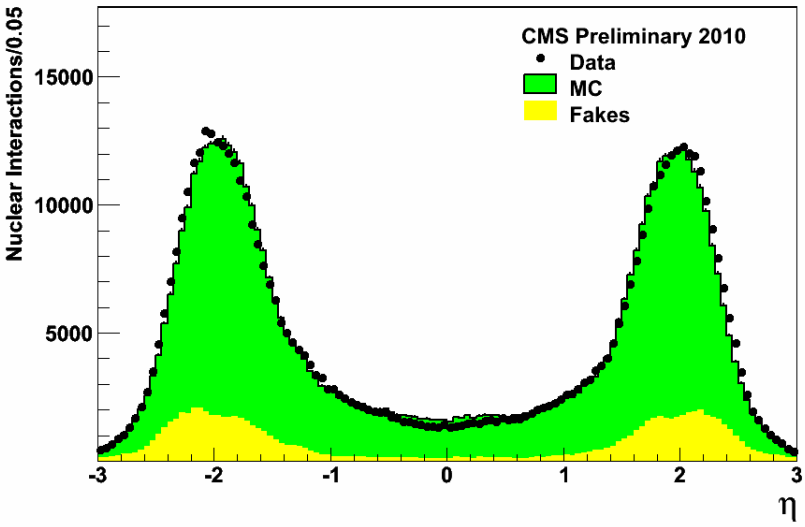


Figure 3.13: Number of reconstructed nuclear interactions against pseudorapidity in data and simulation. A fake nuclear conversion refers to a reconstructed displaced vertex which passes the identification criteria but is not a result of a nuclear conversion. The number of fake nuclear conversions shown in yellow, is determined from simulation.

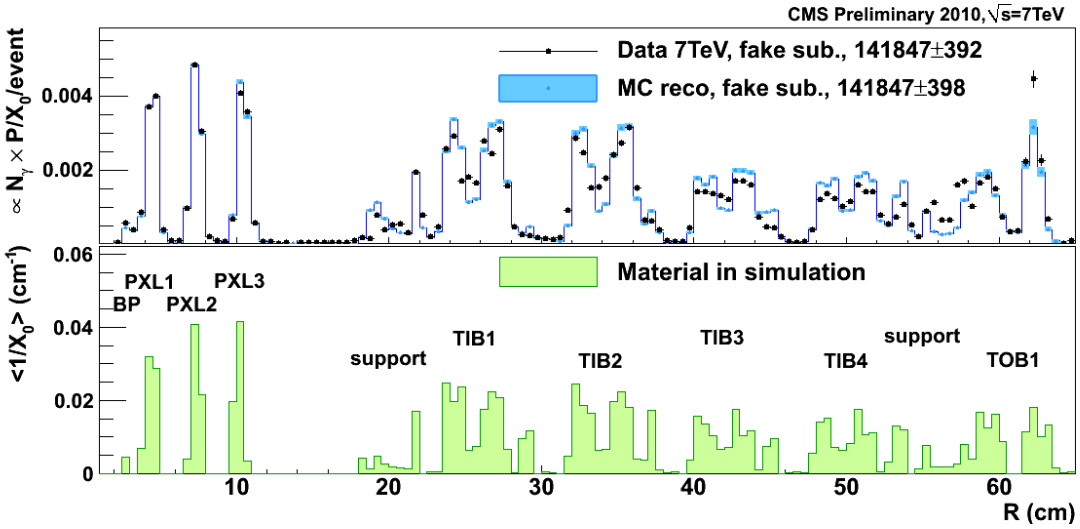


Figure 3.14: The material budget of the tracker as derived from photon conversions. The units are given as the inverse of the radiation length which is shown as proportional to the number of prompt photons per event (N_γ) multiplied by the conversion probability (P/X_0/event). The total number of photon conversions as measured in data and simulation (MC reco) with the fakes subtracted using simulation are shown in the upper right of each plot.

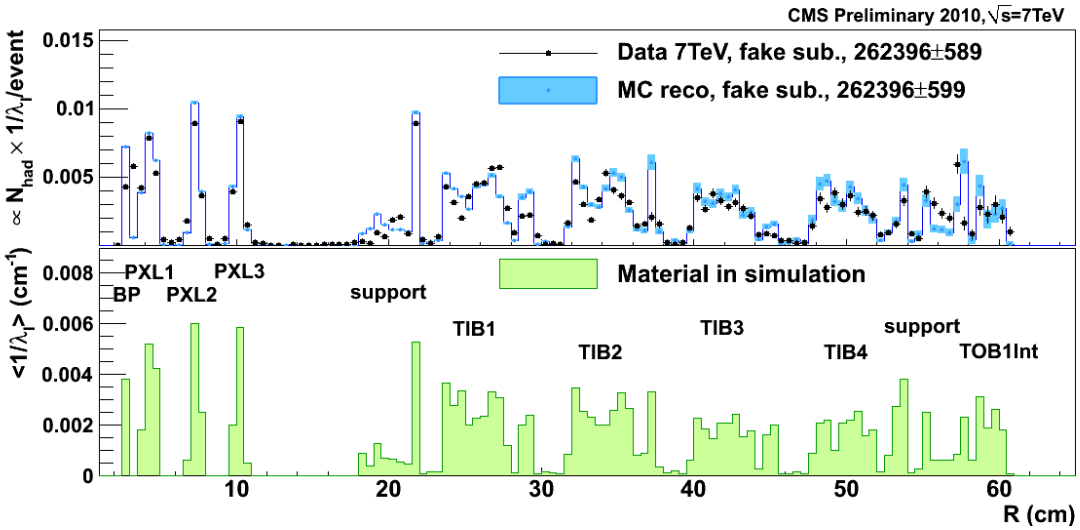


Figure 3.15: The material budget of the tracker as derived from nuclear interactions. The units are given as the inverse of the interaction length which is shown as proportional to the number of prompt hadrons per event (N_{had}) multiplied by the nuclear interaction probability per event ($1/\lambda_0$/event). The total number of nuclear interactions as measured in data and simulation (MC reco) with the fakes subtracted using simulation are shown in the upper right of each plot.

Chapter 4

Data Formats and Reconstruction

4.1 Simulated samples

The use of simulated samples is an integral part of the measurement both at the planning and development stage as well as in the final results. The production of samples for use in CMS data analysis are conducted centrally by the Generator Group. For this particular analysis, the majority of the simulated samples are produced with the MadGraph generator [43]. The MadGraph generator is used to produce the top pair, single-top, $W^{\pm}$ boson and Z^0 boson samples which are produced with up to four additional partons in the matrix-element calculation. The events generated by MadGraph are then passed to pythia 6.4 [44] which simulates the showering and hadronisation of the partons. The matching of the partons to the parton shower is performed using the Kt-MLM algorithm described elsewhere [45]. Some simulated samples were not available with the extra MadGraph step, most notably the multi-jet production through QCD processes is produced only using pythia 6.4. The parton distribution functions (PDFs) used are the CTEQ6.6 PDF sets [46].

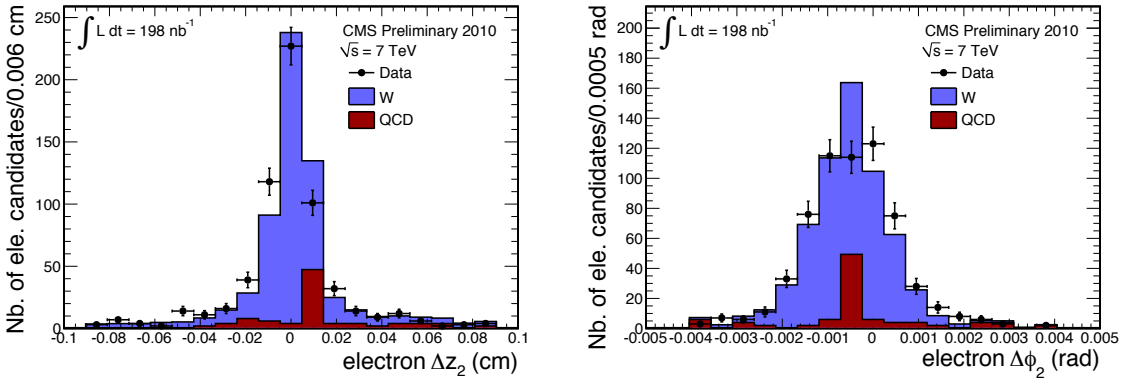
Having completed the generator phase, the resulting events are then passed to the detector simulation. A full reconstruction of the CMS detector has been created in

GEANT4 [47] for this purpose. The simulation phase models all the interactions of the detector materials with the generated events including producing signal deposits in the active elements of the detector. A full reconstruction is then performed using the CMS software with the simulated signals from the sub-detectors.

The cross-section for $t\bar{t}$ in simulation is $157.5^{+23.2}_{-24.4}$ pb calculated with MCFM [11] at NLO. The factorisation and renormalisation scales are set to $(2m_t)^2 + (\sum p_{\mathrm{T}}^{jet})^2$, with $m_t = 172.5$ GeV for this process.

4.2 Electron reconstruction

The reconstruction of electrons makes use of both the electromagnetic calorimeter and the silicon tracker sub-detectors [48]. Electron objects can be reconstructed up to pseudorapidities of $|\eta| < 2.5$. The reconstruction of electrons is performed initially by finding what is referred to as a "seed" which can either be found in the tracker or the ECAL. The tracker-driven seeds use techniques specific to the *particle flow* reconstruction algorithms discussed in Section 4.4. The ECAL driven seeds are found initially by looking for clusters of 5×5 crystals with a minimum energy deposit of 1 GeV. The algorithm then takes this information and looks for energy deposits in surrounding crystals. In particular, it is optimised for the characteristic shape of the shower produced by an electron which is collimated in $\Delta\eta$ but wide in $\Delta\phi$ due to the strong axial magnetic field. Tracking information is then used to attempt to match the seed from the ECAL with a track. The trajectory of the electron in the ECAL is determined from the energy averaged position. The tracking information is matched using a Gaussian Sum Filter (GSF) [49]. The precision of this matching has been measured in data using electrons from leptonically decaying $W^{\pm}$ events. Figure 4.1 shows the difference between the calculated trajectory from the ECAL-driven seed and the position of the track in the second layer of the pixel sub-detector for the ϕ and z coordinates.



(a) (b)

Figure 4.1: The difference in the z (a) and ϕ (b) coordinate of ECAL driven seeds and hits in the second pixel layer (denoted by the "2" subscript) for $W^{\pm}$ events decaying to an electron and a neutrino.

The electron reconstruction has been commissioned in data for electrons with a transverse energy of $E_T > 20\,\text{GeV}$. The primary method is to use Z^0 boson decays to two electrons in a "tag and probe" experiment. One set of tag candidates are required to pass the complete electron identification and reconstruction as well as a minimum E_T cut of 20 GeV. The second probe set of candidates must meet a minimum requirement of $E_T > 20\,\text{GeV}$ and to have an invariant mass when combined with the tagged candidate within the Z^0 mass window (between 85 and 95 GeV). The fraction of probe candidates which subsequently pass the full electron reconstruction is then the reconstruction efficiency. A second study that performs an unbinned maximum likelihood fit to the reconstructed transverse mass in $W^{\pm}$ events decaying to an electron and a neutrino, is detailed elsewhere [48]. The results from the first measurement give electron reconstruction efficiencies of $99.3 \pm 1.4\%$ and $96.8 \pm 3.4\%$ for the ECAL barrel and endcaps, respectively. The ratio of the measured efficiency to the predicted efficiency in simulation for both measurements are illustrated in Figure 4.2.

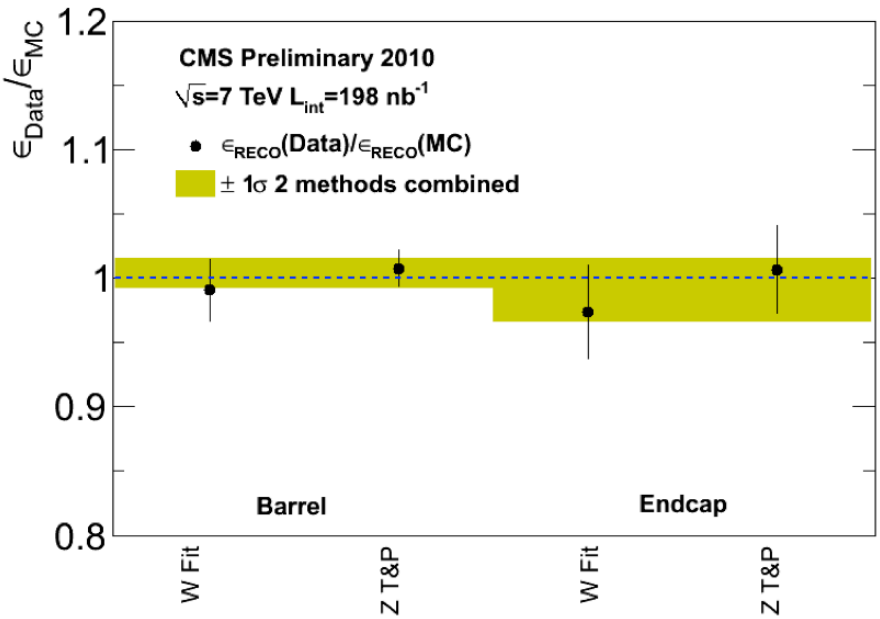


Figure 4.2: The ratio of the electron reconstruction efficiency in data to simulation (MC) as measured in the Z^0 decaying to two electrons, tag and probe study and with the unbinned maximum likelihood fit to the transverse mass distribution in $W^{\pm}$ events decaying to an electron and a neutrino.

4.3 Muon reconstruction

The reconstruction of muons in the CMS detector primarily uses the muon system and the silicon tracker [50]. The reconstruction proceeds initially by reconstructing candidate muons in the muon system referred to as *stand-alone muons*. The muon candidates can then be matched to tracks by starting from the muon system and extrapolating to a track in the tracker, making what are known as *global muons*. A second method can also be used which considers all tracks to be muon candidates and then extrapolates this to a match in the muon system which produces *tracker muons*. The algorithm used for this analysis will be the global muon reconstruction.

The efficiency to reconstruct a muon with the global muon algorithm has been measured extensively with cosmic muons during the Cosmic Run At Four Tesla [51]. So far only the low momentum range has been probed using the $J/\psi \to \mu^+\mu^-$ resonance [50]. The advantage of using cosmic muons for calibration is that they are a highly pure signal and cover a very wide range of momenta. The disadvantage is they are only useful for the barrel section of the muon system ($|\eta| < 0.8$) due to their approximately vertical trajectories.

To measure the efficiency in the barrel section with cosmic muons, the barrel is separated into top and bottom halves. One half is required to reconstruct a good-quality global muon. The other half is required to reconstruct a global muon which matches this. The matching requires that the reconstructed trajectory lies within $|\Delta\phi| < 0.3$ and $|\Delta\eta| < 0.3$ of the original candidate. Additionally, the distance of closest approach of both muon candidates are required to be within $r < 4$ cm (the radius of the beam pipe) and within $|z| < 10$ cm of the nominal interaction region. This is to ensure that they accurately emulate reconstructions for real collisions. The efficiency was measured to be $(97.1 \pm 0.6)\%$ for global muons in the barrel region.

4.4 Particle flow reconstruction

The particle flow reconstruction algorithm is an attempt to use all the detector information to reconstruct all of the relatively long-lived particles in the event [52]. As such it includes a complete description of all of the physics objects previously discussed. However, due to its more complex nature, it was not certain that it would be commissioned in time for this analysis and most of the initial studies were conducted with more simplistic algorithms to take a conservative approach. Since particle flow reconstruction has been commissioned in time, it has been possible to replace some objects such as the jets (Section 4.5) and the missing transverse energy (Section 4.6) with this improved method.

The basis of particle flow reconstruction is to exploit the excellent spatial resolution and track reconstruction efficiency of the tracker to identify individual particles. The first two steps of the algorithm are to identify muon and electron candidates. The particle flow muon reconstruction proceeds in a very similar manner as already discussed, except there is an additional requirement for the global muon to have a momentum compatible with that measured only by the tracker within three standard deviations. This additional requirement is due to the reliance of the particle flow algorithm on the track momentum for the identification of charged and neutral hadrons. The electron reconstruction differs in that it is seeded by the tracker. The particle flow electron reconstruction exploits the tendency for electrons to produce short tracks and to radiate photons in the tracker material due to Bremsstrahlung. As before, the trajectory of the electron is corrected using the Gaussian-Sum Fitter algorithm and then matched to deposits in the electromagnetic calorimeter.

Having identified muon and electron candidates, their tracks in the tracker and energy deposits in the calorimeters are removed from further consideration. The reconstruction then proceeds by trying to identify charged and neutral hadrons in the event. This involves linking tracks to clusters of energy in both calorimeters to identify charged hadrons. Neutral hadrons are then identified as excess energy in the hadronic calorime-

ter clusters once the charged components have been subtracted. The details of how the clusters in both calorimeters are linked to tracks is quite involved and is detailed elsewhere [52]. In rare cases additional muons may also be identified at this stage if the remaining track momentum greatly exceeds that of the energy deposits in the calorimeters. Finally, the remaining clusters in the electromagnetic calorimeter that are not linked with a track are assigned as particle flow photons.

4.5 Jet reconstruction

Jets are clustered from a list of particles provided by the particle flow algorithm and referred to as *PFJets*. They may also be produced from generator level information in simulation based studies, referred to as *GenJets*. Jets are a useful construct to identify the hadronisation of the initial partons in the event as they have a tendency to produce a collimated shower of relatively high energy particles within the detector. The jet reconstruction algorithm used as standard at CMS is the *anti-k_t* algorithm. The anti-k_t algorithm uses successive combination to produce jets from a list of object positions and transverse momenta [53, 54].

The algorithm operates using equation 4.1 and equation 4.2,

$$d_{ij} = \min(\frac{1}{k_{t,i}^2}, \frac{1}{k_{t,j}^2})\frac{\Delta_{i,j}^2}{R^2} \tag{4.1}$$

$$d_{iB} = \frac{1}{k_{t,i}^2} \tag{4.2}$$

where parameters k_t refer to the transverse momentum of the particle(s), $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ (y being the rapidity) and R is the size parameter which is 0.5 for standard CMS jets. The algorithm proceeds by calculating d_{ij} for all combinations of two objects and by calculating d_{iB} for individual objects. If the resulting smallest value is a combination of objects using equation 4.1 then those two are merged into one object referred to as a *pseudo jet* and the resulting transverse momentum and

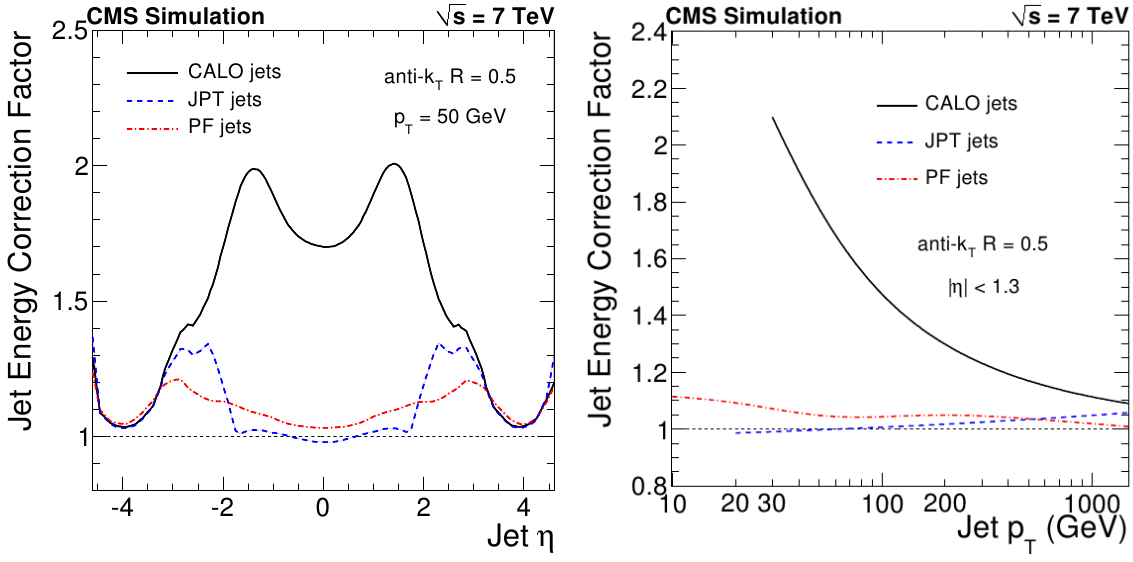
positions are recalculated according to equations 4.3.

$$\begin{aligned} k_T &= k_{T,i} + k_{T,j} \\ y &= [k_{T,i} \cdot y_i + k_{T,j} \cdot y_j]/k_T \\ \phi &= [k_{T,i} \cdot \phi_i + k_{T,j} \cdot \phi_j]/k_T \end{aligned} \tag{4.3}$$

If the smallest value corresponds to an individual object as in equation 4.2, this object is now labelled a jet and removed from further consideration. This process repeats until all objects are clustered into jets. It is important to note that this will result in isolated electrons and muons generating jets on their own and these have to be removed in the analysis event selection.

The jets that are produced in the event have additional corrections to the energy applied to account for effects such as calorimeter response. The strategy for jet energy corrections at CMS proceeds in two steps [55]. The first is a comparison between the PFJet reconstruction and GenJets using simulated samples. As the scale factor is dependent on the generator used, other generators are also studied and contribute to the systematic uncertainty on this measurement. Individual PFJets are matched to GenJets with the requirement $\Delta R < 0.3$. The average ratio of the transverse momentum for the PFJet and associated GenJet is then the calibration factor which is parameterised in uncorrected p_t and η. The calibration factor for PFJets is illustrated in Figure 4.3.

The second stage of the jet energy corrections are the in-situ calibrations which currently include two correction factors. The relative jet energy scale is measured using di-jet events. One jet is constrained to lie within the barrel region ($|\eta| < 1.3$) known as the barrel jet, the second jet (probe jet) can be in any region provided that the barrel and probe jets are separated in ϕ by $\Delta\phi > 2.7$. Additionally, no other jets in the event may have a transverse momentum greater than 20% of the average transverse momentum of the two jets. The correction factor is measured by calculating the



(a) (b)

Figure 4.3: The simulation calibration factor for PFJets for (a) η at a fixed jet transverse momentum of 50 GeV/c and (b) p_T for $|\eta| < 1.3$. Also shown are two other jet types, calorimeter (CALO) and jet plus track (JPT). The details of these two algorithms can be found elsewhere [56].

quantity known as "balance" defined in equation 4.5.

$$p_T^{\text{ave}} = (p_T^{\text{probe}} + p_T^{\text{barrel}})/2 \tag{4.4}$$

$$B = \frac{p_T^{\text{probe}} - p_T^{\text{barrel}}}{p_T^{\text{ave}}} \tag{4.5}$$

The average value of B is measured in a number of $|\eta^{\text{probe}}|$ and p_T^{ave} bins from which the relative response, R, is calculated for each bin using equation 4.6.

$$R(\eta^{\text{probe}}, p_T^{\text{ave}}) = \frac{2+ < B >}{2- < B >} \tag{4.6}$$

At this stage an additional residual correction has to be introduced. The residual correction is caused by a systematic bias arising from the differing energy resolutions of the barrel and probe jets which is discussed in more detail elsewhere [55]. This effect is measured in simulation.

The second set of corrections determine the absolute jet energy scale using photon+jet and Z^0+jet events with a technique known as missing transverse energy projection fraction (MPF). The event selection is similar to the di-jet balance measurement with a reconstructed photon or Z^0 candidate and a recoiling jet in the barrel region ($|\eta| < 1.3$). For events with a Z^0 boson decaying to electrons (muons) the recoiling jet is also required to be separated in ϕ by $\Delta\phi > 2.7$ ($\Delta\phi > 2.8$). For events with a photon converting to an electron-positron pair, the recoiling jet is required to be separated in ϕ by $\Delta\phi > 2.7$. The photon candidates are required to be isolated and have a minimum transverse momentum of $p_T^{\gamma} > 15$ GeV/c. Events are vetoed if there are additional jets with transverse momenta greater than 20% of that of the photon or Z^0, that are outside a cone of $\Delta R > 0.25$ around the direction of the reference particles.

In a perfect system, the transverse momenta of the photon or the Z^0 boson and the recoiling jet would balance. However, as they are subject to detector resolution effects, a certain amount of missing transverse energy will be measured incorrectly in the event. As such, the detector response to the jet measured with the MPF method (R_{MPF}) can be calculated from the response to photons or Z^0 bosons($R_{\gamma,Z}$), the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) and the transverse momentum of the photon or Z^0 boson ($p_T^{\gamma,Z}$) by equation 4.7.

$$R_{MPF} = R_{\gamma,Z} + \frac{E_{\mathrm{T}}^{\mathrm{miss}} \cdot p_T^{\gamma,Z}}{(p_T^{\gamma,Z})^2} \tag{4.7}$$

The value of R_{MPF} is subject to a small bias which results from the presence of secondary jets that still pass the veto described above. To compensate for this effect, the measurement is repeated whilst the maximum transverse momentum fraction of the secondary jets is decreased. The resulting data points are then extrapolated to a maximum secondary jet p_{T} of 0 to give the absolute jet energy scale.

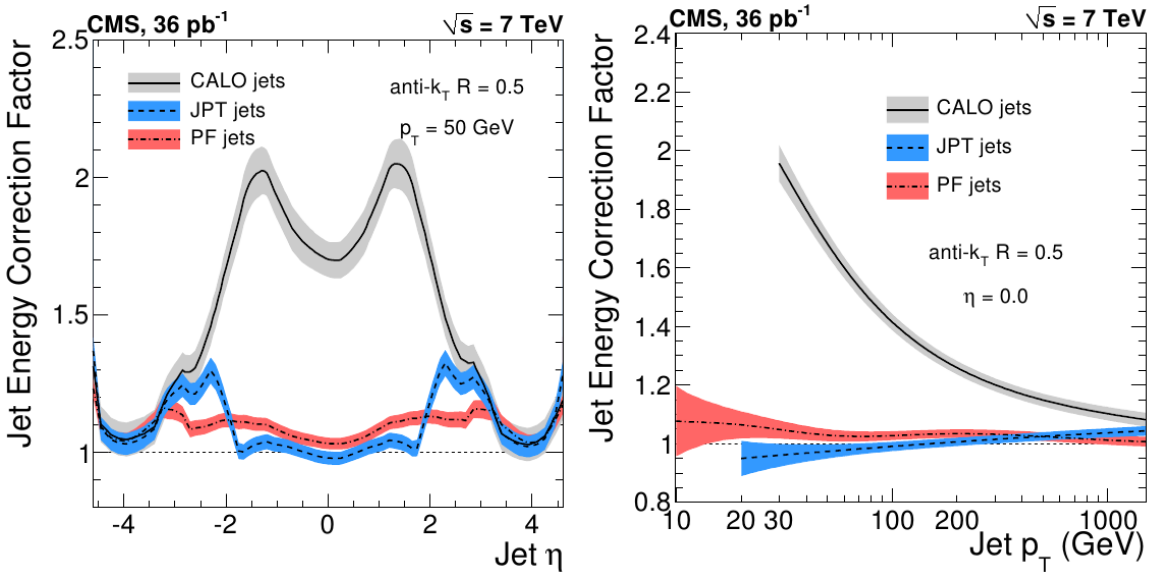
The full jet energy correction is calculated using equation 4.8, where C_{MC} is the correction factor from simulation, C_{rel} is the relative energy scale factor, C_{abs} is the absolute energy scale factor, p_T^{raw} is the uncorrected transverse momentum, and p_T' is the transverse momentum after applying the simulation and relative energy scale factors. The

four momentum of the jet is recalculated using the scale factor with equation 4.9.

$$C = C_{\mathrm{MC}}(p_T^{\mathrm{raw}}, \eta) \cdot C_{\mathrm{rel}}(\eta) \cdot C_{\mathrm{abs}}(p'_T, \eta) \tag{4.8}$$

$$p_\mu^{\mathrm{corr}} = C \cdot p_\mu^{\mathrm{raw}} \tag{4.9}$$

The full jet energy corrections are shown in Figure 4.4.



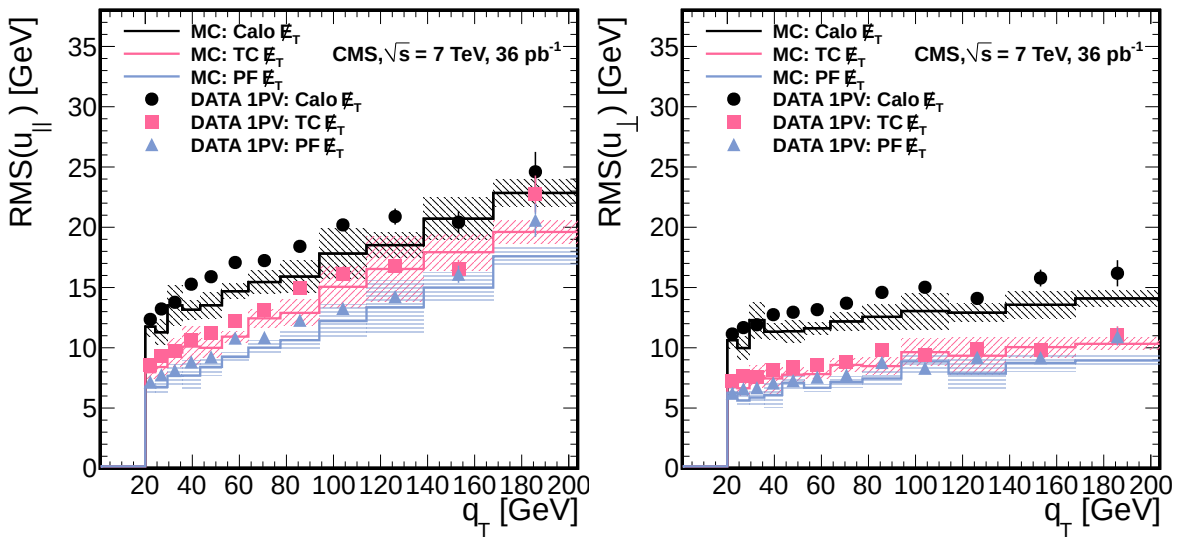
(a) (b)

Figure 4.4: The full jet energy correction factor for PFJets for: (a) η at a fixed jet transverse momentum of 50 GeV/c and (b) p_T for $\eta = 0.0$. Also shown are two other jet types, calorimeter and jet plus track. The coloured bands indicates uncertainty on the measurements.

4.6 Missing transverse energy

The missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) in an event is a useful indicator of the presence of very weakly interacting particles such as neutrinos and as such can be a useful tool in identifying events which contain real $\mathrm{W}^{\pm}$ bosons decaying leptonically. For this analysis $E_{\mathrm{T}}^{\mathrm{miss}}$ is used for one background study but is not a feature of the analysis selection. The missing transverse energy in the event is simply determined by a vector sum of the transverse momentum over all the reconstructed objects in the event. For the list of objects, particle flow reconstruction is used as described in Section 4.4.

The performance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ using particle flow is discussed in detail elsewhere [57]. To study the $E_{\mathrm{T}}^{\mathrm{miss}}$ energy scale and resolution events with reconstructed $\mathrm{W}^{\pm}$, Z^0 and γ bosons are used. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is described by the equation, $\vec{E_{\mathrm{T}}^{\mathrm{miss}}} = -\vec{q_T} - \vec{u_T}$, where q_T is the transverse component of the boson momentum and u_T is the vectorially summed transverse momentum of everything else. Under the assumption that the momentum of the boson is well measured, the resolution of $\vec{E_{\mathrm{T}}^{\mathrm{miss}}}$ is dependent only on the $\vec{u_T}$ term. The $\vec{u_T}$ term is decomposed into a parallel ($u_{||}$) and longitudinal component ($u_{\perp}$) with respect to the reconstructed direction of the boson. The measured resolution of these components is shown in Figure 4.5 for three different $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithms.



(a) (b)

Figure 4.5: Resolution of the hadronic recoil u_T for the component parallel to the boson (a), and the component perpendicular to the boson (b). Results are shown for three $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms, calorimeter (Calo), track corrected (TC) and particle flow (PF). The Calo and TC $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithms are summarised elsewhere [57]. Shaded regions indicates the statistical uncertainty on simulation samples.

4.7 Primary vertex reconstruction

The primary vertex reconstruction makes use of the reconstructed charged particle tracks in the event. The tracks are selected according to their compatibility with the LHC beam-line. This selection is performed with requirements on the transverse impact parameter significance with respect to the track and the beam-line. They are also required to pass a minimum number of strip and pixel hits in the tracker and pass a cut on the normalised χ^2. From the remaining tracks, vertex candidates are then constructed by clustering tracks according to their z co-ordinate. The maximum separation between the track and its nearest neighbour is $z_{sep} = 2$ mm for clustering into a common vertex. The vertex candidates are then passed to the adaptive vertex fitter which performs a three dimensional fit to try and reconstruct as many vertices as possible from the candidates [58]. Reconstructed vertices are then ranked according to the scalar sum of the transverse momenta squared of all the tracks in the vertex.

The performance of the vertex reconstruction is intrinsically linked to that of the tracker and was discussed previously in Section 3.3.

Chapter 5

Analysis Overview

5.1 Introduction

The analysis proceeds as a "counting experiment". A tight initial selection is applied as discussed in Section 5.2 to attempt to separate the $t\bar{t}$ signal from the background processes. Following the selection, the number of background events that remain are then measured separately as discussed in Chapter 7. The cross-section for $t\bar{t}$ production is then calculated using equation 5.1,

$$\sigma_{t\bar{t}} = \frac{N_{\text{obs}} - N_{\text{bkg}}}{A \cdot \epsilon^{\text{tag}} \cdot \mathcal{L}} \tag{5.1}$$

where N_{obs} is the total number of observed events, N_{bkg} is the estimated number of background events, A is the acceptance, ϵ^{tag} is the selection efficiency specifically for b-tagging and $\mathcal{L}$ is the integrated luminosity.

The acceptance is the efficiency with which $t\bar{t}$ events pass the event selection excluding the b-tagging requirement. This is primarily calculated using $t\bar{t}$ simulation, although corrections for the trigger efficiency and electron identification are applied from data. The efficiency to pass the b-tagging selection, ϵ^{tag}, is also measured in simulation with the application of scale factors measured in data (see Section 6).

The primary backgrounds to $t\bar{t}$ production in the electron+jets channel listed in order of their final contribution are as follows:

- Leptonically decaying $W^{\pm}$ bosons with additional jets (W+jets).
- QCD (see below).
- Single-top production.
- Leptonically decaying Z^0 bosons with additional jets (Z+jets).
- Di-Boson production.

The category of QCD includes all other processes not included as part of the other named backgrounds. These events are primarily distinguished from the others as they have no real sources of single isolated electrons. This includes QCD multi-jet production and also photon production in association with jets (photon+jets).

5.2 Event selection

The initial event selection is a common event selection for the Top Physics Analysis Group at CMS to provide a reference point for all of the analyses working within each of the channels. The basic principle is to select the characteristics of a top pair decay in the electron channel whilst reducing the background and also the contamination from the di-lepton and muon+jets channels. The selection is a result of extensive studies on simulation before the LHC start up in 2010 [59].

5.2.1 Trigger

The technical aspects of the trigger system was discussed previously in Section 2.6. The triggers used by this analysis are single electron triggers without isolation. Due to the rapid increase in the collision rate during the 2010 data taking period, a number

of triggers were used to keep the transverse energy threshold of the electrons as low as possible whilst avoiding any trigger pre-scaling (where the rate of the trigger is artificially limited). A secondary effect of this is that none of the triggers were available in simulation and as such are not part of the event selection for simulation studies. However, as the offline electron selection is significantly more stringent than what is applied by the trigger the difference in selection efficiency is negligible.

The Level 1 trigger object is based solely on calorimeter deposits. The algorithm starts with the highest energy trigger tower in the ECAL and then adds the energy from the highest energy deposit of its four neighbours on the broad-side of the crystal. The candidate must then pass a requirement based on the spread of the shower shape in the magnetic field and also a requirement on the fraction of energy deposited in the HCAL in this region. The final component is a transverse energy cut which is configurable and varied between 5 and 8 GeV for 2010 data taking.

The HLT trigger makes use of the full event information. In particular, the triggers used include varying degrees of electron identification with a match to a pixel hit and possibly a track in a small (SW) or large window (LW) which corresponds to the size of the $\Delta\phi$ cut between the extrapolated track and the energy deposits. Specifically, the electron identification makes use of the HCAL energy fraction, $\Delta\phi$, $\Delta\eta$ and $\sigma_{i\eta i\eta}$ variables which are discussed as part of the electron selection in Section 5.2.3. There is also a transverse energy cut which varies between 10 and 22 GeV for this analysis. As the collision rate increased these criteria had to be tightened to avoid having to pre-scale the trigger. The triggers used later on in the data taking period have more stringent electron identification and a higher transverse energy cut. The high level triggers used and their characteristics are listed in Appendix A.

The efficiency for events to pass the trigger requirement is measured in data with the full electron-triggered 2010 data-set. The efficiency is calculated from the rate at which high quality isolated electron candidates, as defined in the electron selection below, are matched to the electron trigger objects described above. The matching of trigger

objects to electron candidates is performed using a $\Delta R < 0.2$ cut. The efficiency of the trigger was measured to be $\epsilon_{\text{trigger}} = 0.982 \pm 0.001(\text{stat})$.

5.2.2 Primary vertex selection

The primary vertex is required to be consistent with the beam spot (the location of the beam crossing point within CMS). The consistency with the beam spot is determined by requiring that $|\Delta z| < 24.0$ cm and $\Delta\rho = \sqrt{x^2 + y^2} < 2.0$ cm between the primary vertex and the beam spot. Additionally, the vertex is required to have more than four degrees of freedom. For events where multiple primary vertices are reconstructed, the vertex with the highest sum of the squared track transverse is used.

5.2.3 Electron selection

The events are required to have a single isolated electron which passes the following requirements:

- The electron is required to be consistent with the primary vertex with a $|\Delta z| <$ 1.0 cm.
- The two dimensional impact parameter between the electron track and the beam spot is required to be $|d_0| < 0.02$ cm.
- The transverse energy of the electron must be greater than 30 GeV.
- The electron must be located within the tracker acceptance with pseudorapidities of $|\eta| < 2.5$. In addition, electrons lying within the transition region between the ECAL barrel and endcaps are excluded with a cut on the pseudorapidity of the ECAL supercluster of $1.4442 < |\eta_{sc}| < 1.5660$.
- The electron must pass the *simple cut based electron ID* W70 (nominally a working point with 70% efficiency). This is discussed in detail below.

The electron identification uses three different shower shape variables ($\sigma_{i\eta i\eta}$, $\Delta\eta$ and $\Delta\phi$) as well as the fraction of energy deposited in the HCAL compared to the ECAL (H/E) to attempt to identify real electrons. The variables $\Delta\eta$ and $\Delta\phi$ measure the spatial matching between the electron supercluster and the track, and the variable $\sigma_{i\eta i\eta}$ is a measure of the supercluster width in pseudorapidity [48]. The selection for the W70 working point is given in Table 5.1.

The isolation requirement is evaluated with equation 5.2, referred to as RelIso, where $tkIso$, $ecalIso$ and $hcalIso$ are the total track momentum, ECAL transverse energy, and HCAL transverse energy summed in a cone of $\Delta R < 0.3$ around the electron, respectively. The contribution from the electron candidate is subtracted from these quantities. The RelIso is required to be less than 0.1.

$$\text{RelIso} = \frac{\text{tkIso} + \text{ecalIso} + \text{hcalIso}}{\text{p}_\text{T}(\text{electron})} \tag{5.2}$$

The efficiency to identify electrons in data is corrected using a scale factor measured with Z^0 boson decays. The measurement is performed with a "tag and probe" experiment in which one electron candidate is tagged by requiring the selection above and compared with other loose candidates as defined by the loose electron selection in Section 5.2.5. The other loose candidates are required to reconstruct an invariant mass within 15 GeV/c^2 of the Z^0 boson mass to be identified as the probe. The efficiency with which the probe electron then passes the full tight identification and isolation requirements above is taken to be the identification efficiency in data. To apply this in the measurement, it is used as a scale factor defined as $SF_{\text{ele}} = \varepsilon^{\text{data}}/\varepsilon^{\text{sim.}}$, where $\varepsilon^{\text{data}}$ is the efficiency measured in data and ε^{sim} is the efficiency calculated in $\text{t}\bar{\text{t}}$ simulation. Using the tag and probe technique the scale factor was measured to be $SF_{\text{ele}} = 0.950 \pm 0.015(\text{stat.}) \pm 0.029(\text{sys.})$.

Table 5.1: The simple cut based electron identification for the W70 working point for electrons in the ECAL barrel and endcaps.

Variable	ECAL Barrel	ECAL Endcaps
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi$	< 0.03	< 0.02
$\Delta\eta$	< 0.004	< 0.005
H/E	< 0.025	< 0.025

5.2.4 Muon veto

To reject top pair decays in the muon+jets and di-lepton channels a loose muon veto is employed. Events are rejected if they contain a global muon candidate (see Section 4.3) with a transverse momentum greater than 10 GeV/c, pseudorapidity of $|\eta| < 2.5$ and a RelIso < 0.2, where RelIso is calculated as in equation 5.2 using the transverse momentum of the muon instead of the electron.

5.2.5 Z^0 boson veto

The background from Z+jets events is reduced by vetoing events in which a second, more loosely defined, electron reconstructs an invariant mass in combination with the single isolated electron that is consistent with the Z^0 boson mass. The second loose electron is selected with a minimum transverse energy of 20 GeV, the same pseudorapidity selection as the tight electron candidate and a RelIso of less than 1.0. The electron identification uses the W95 working point (nominally corresponding to a 95% selection efficiency) with the selection defined in Table 5.2.

The event is rejected if the invariant mass of any loose electron candidates in combination with the isolated electron is within the window $76 < m_{ll} < 106$ GeV/c^2.

Table 5.2: The simple cut based electron identification for the loose electron at the W95 working point for electrons in the ECAL barrel and endcaps.

Variable	ECAL Barrel	ECAL Endcaps
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi$	< 0.8	< 0.7
$\Delta\eta$	< 0.007	< 0.010
H/E	< 0.15	< 0.07

5.2.6 Photon conversion veto

Electrons originating from photon conversions represent a significant source of fake isolated electrons. Events are rejected if the electron is identified as having originated from a photon conversion. Identifying such electrons is performed using two techniques, only one of which is required to identify a conversion for the event to be rejected.

The first technique is to identify conversions with missing hits from the electron track in the inner part of the tracker. If there are missing hits the electron is identified as a photon conversion. The second technique attempts to identify a partner track to the electron track which would be consistent with having originated from a conversion that also produced the electron candidate. The event is vetoed if a partner track is identified with an absolute distance between the two tracks of less than 0.02 cm and a difference in the cotangent of the polar angle of each track of less than 0.02.

5.2.7 Jet selection

The cross-section is measured twice with the requirement of three or more and four or more jets. Each jet must fulfil the following identification criteria:

- A transverse momentum greater than 30 GeV/c.

- The jet must be located within the tracker acceptance with pseudorapidities of $|\eta| < 2.4$.
- The fraction of energy deposited in the ECAL from reconstructed charged hadrons must be less than 0.99.
- The fraction of energy deposited in the ECAL from reconstructed neutral hadrons must be less than 0.99.
- The fraction of energy deposited in the HCAL from reconstructed neutral hadrons must be less than 0.99.
- The fraction of the energy of the jet from reconstructed charged hadrons must be greater than zero.
- The number of reconstructed charged hadrons within the jet must be greater than zero.

In addition, as isolated electrons can be reconstructed as jets, any jet within a cone of $\Delta R < 0.3$ of the electron candidate is removed.

5.2.8 Selection of b-jet candidates

At least one of the remaining jets is required to be identified (tagged) as a b-jet using a life-time based b-tagger. The selection of b-jet candidates is discussed in detail in Section 6.

5.3 Selection summary

The event selection is applied as follows:

1. Pass high level trigger (data samples only).
2. Pass primary vertex selection.

3. A single isolated high quality electron candidate must be present.
4. No isolated loose muon candidates should be present.
5. No Z^0 boson candidates within the $76 < m_{ll} < 106$ GeV/c^2 mass window should be present.
6. Electron candidate must not be consistent with a photon conversion.
7. At least three (four) good jets must be present.
8. At least one b-tagged jet must be present.

The selection efficiencies for $t\bar{t}$ events as measured in simulation are shown in Table 5.3. The number of events normalised to the integrated luminosity for the signal and background simulation samples are shown in Table 5.4. The number of events in data are shown in Table 5.5. The jet multiplicity distribution after event selection including the requirement of one or more b-tagged jets is shown in Figure 5.1.

Table 5.3: Number of events selected and selection efficiency in the $t\bar{t}$ simulation sample. Uncertainties are only due to simulation statistics. The inclusive efficiency is the combined effect of all proceeding cuts, where as the cut efficiency refers only to the effect of that cut on the remaining events before it.

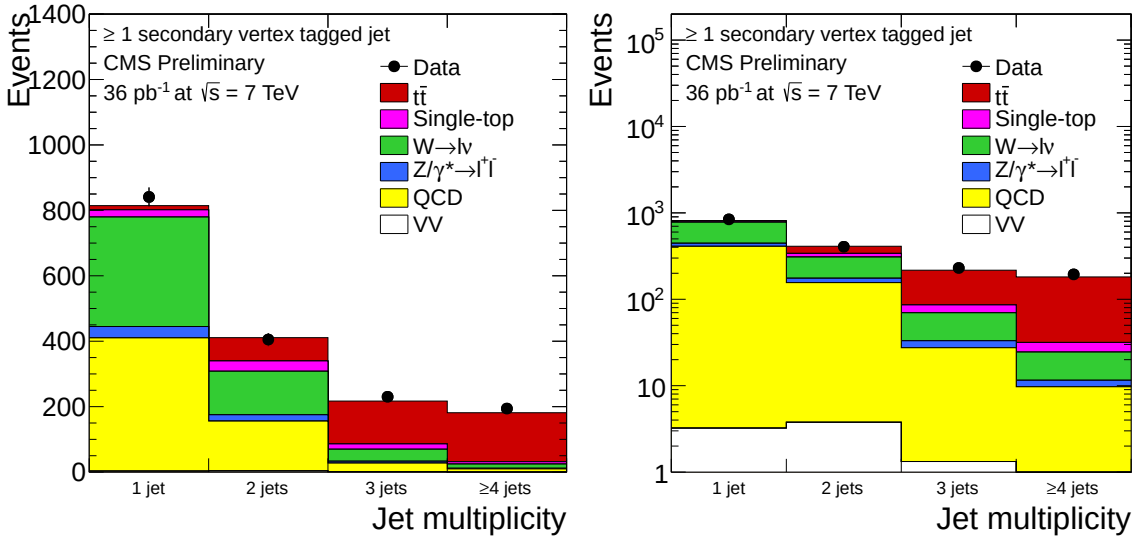
Step	Events left	Inclusive efficiency	Cut efficiency
Initial events	1286182 ± 1134	1	1
1 electron	137574 ± 371	0.1070 ± 0.0003	0.1070 ± 0.0003
μ veto	121783 ± 349	0.0947 ± 0.0003	0.885 ± 0.003
Z^0 veto	119937± 346	0.0933 ± 0.0003	0.985 ± 0.003
Conv. rej.	112951 ± 336	0.0878 ± 0.0003	0.942 ± 0.003
$\geq$ 3 jets	81351 ± 285	0.0633 ± 0.0002	0.720 ± 0.003
$\geq$ 3 jets $\geq$ 1 b-tag	63451 ± 252	0.0493 ± 0.0002	0.780 ± 0.003
$\geq$ 4 jets	42015 ± 205	0.0327 ± 0.0002	0.372 ± 0.003
$\geq$ 4 jets $\geq$ 1 b-tag	33837 ± 184	0.0263 ± 0.0002	0.805 ± 0.004

Table 5.4: Summary of the expected number of events in 36 pb^{-1} measured in simulated samples. The uncertainties shown are statistical only. The QCD contribution is shown separately as QCD multi-jet and photon+jets.

Step	$t\bar{t}$	W+jets	Z+jets	QCD multi-jet	Single top	Photon+jets	Di-Boson
$\geq$ 3 jets	357 ± 1	503 ± 6	88 ± 2	305 ± 19	33 ± 0	72 ± 2	12 ± 0
$\geq$ 3 jets $\geq$ 1 b-tag	278 ± 1	50 ± 2	8±1	27 ± 5	23 ± 0	8 ± 1	2 ± 0
$\geq$ 4 jets	184 ± 1	101 ± 3	17 ± 1	47 ± 6	9 ± 0	9 ± 1	2 ± 0
$\geq$ 4 jets $\geq$ 1 b-tag	148 ± 1	13 ± 1	2 ± 0	8 ± 3	7 ± 0	1 ± 0	0 ± 0

Table 5.5: The number of events selected in 36 pb^{-1} of data. Uncertainties are statistical only.

Step	Events left	Inclusive eff. efficiency	Cut eff. efficiency
1 electron	175361 ± 419	1	1
μ veto	175164 ± 419	1.00	1.00
Z veto	170823 ± 413	0.97	0.98
Conv. rej.	128443 ± 358	0.73	0.75
≥ 3 jets	1611 ± 40	0.0092	0.013
≥ 3 jets ≥ 1 b-tag	424 ± 21	0.0024	0.26
≥ 4 jets	428 ± 21	0.0024	0.0033
≥ 4 jets ≥ 1 b-tag	194 ± 14	0.0011	0.45



(a) (b)

Figure 5.1: A comparison between data and simulation of the jet multiplicity distribution for selected events including one or more b-tagged jets shown on a linear (a) and a logarithmic scale (b).

5.4 Luminosity measurement

An important part of any cross-section measurement is the integrated luminosity which features directly in the calculation in equation 5.1 and is also used for the normalisation of simulated samples. At CMS, a prompt luminosity estimate is provided using two different methods which both use the activity in the forward calorimeter of the HCAL (HF) [60]. The first method infers the rate of interactions from the average number of towers in the HF calorimeter without any activity. The second method exploits the linear relationship between the average energy deposit in an HF calorimeter tower and the luminosity. The primary online measurement is the latter technique.

The offline measurements use a further two techniques. The first again makes use of the HF calorimeter, in this case a coincidence of transverse energy deposits with a minimum of $E_{\mathrm{T}} > 1$ GeV in the forward and backward HF calorimeters is used to estimate the rate. A second method using the reconstruction of a vertex is also used. These offline techniques are typically used as a cross-check against the prompt results.

To calibrate the rates measured with these techniques, an absolute calibration must be performed with Van der Meer scans. This particular technique is detailed elsewhere [61]. These measurements provide an integrated luminosity for the 2010 data-set of 35.8 pb^{-1} with an uncertainty of 4%.

Chapter 6

Identifying b-jets and Measuring the $t\bar{t}$ Event Efficiency

6.1 Introduction

The tagging of b-jets is a very useful tool for identifying $t\bar{t}$ events simply because the top decay to a W^+ boson and a b quark provides a high probability of reconstructing a b-jet that will pass the selection criteria. The dominant production mode for b-jets in the background samples (excluding single-top) is that of gluon splitting to a $b\bar{b}$ pair.

The flavour of a jet can be separated into three categories. The first two are often referred to collectively as "heavy flavour" which are b- and c-jets. The final category is "light flavour" which includes jets of flavours u, d, s and g and is often abbreviated to l-jets. For simulated samples, the jet flavour is defined using GenJets. For each generator jet, a cone of $\Delta R < 0.3$ is considered. If there is a b or c parton in the cone, the jet is labelled as a b or c-jet, with b partons taking precedence. Otherwise, the flavour is assigned by the parton with the highest momentum within the cone. This particular definition ensures that jets in which a gluon splits to a $b\bar{b}$ or a $c\bar{c}$ pair are labelled as heavy flavour which better reflects their performance with respect to b-jet

identification. Reconstructed PFJets are then associated to GenJets within a cone of $\Delta R < 0.3$ and inherit their flavour. In very rare cases no GenJet is associated to the PFJet and no flavour can be assigned. For the purposes of this analysis these are assumed to be light flavour.

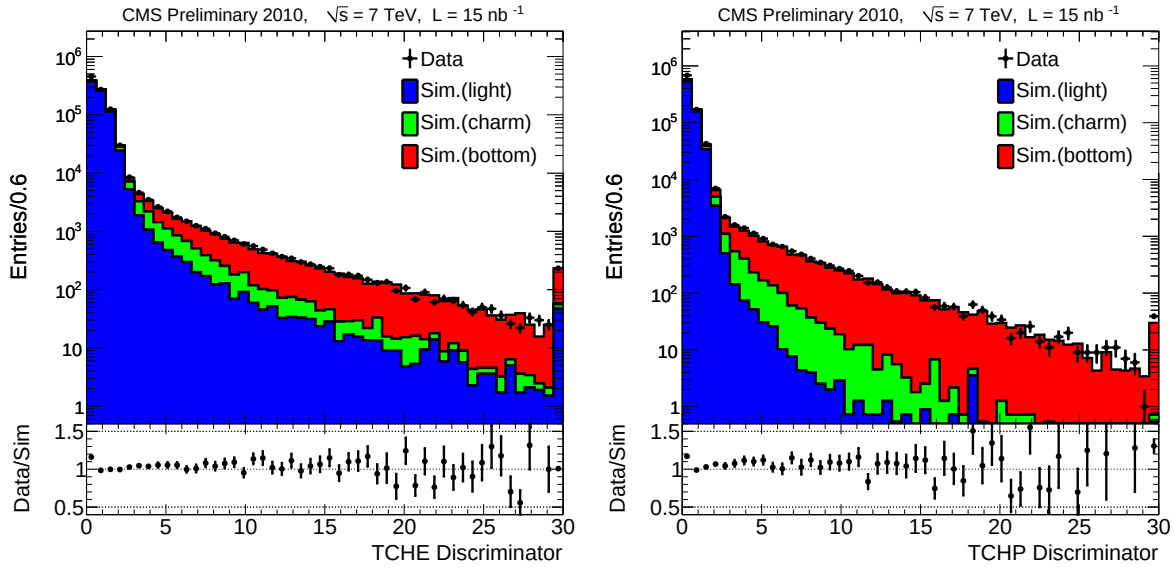
The techniques for tagging the flavour of jets in data, their efficiency measurements and the event selection efficiency are discussed in the following sections.

6.2 Algorithms

To identify b-jets in data samples the algorithms attempt to exploit the characteristics of the decay of B hadrons. In particular, the relatively long lifetime of the B hadrons can give rise to tracks that are not consistent with prompt production and which can therefore be identified through the impact parameter of the tracks or reconstructed secondary vertices associated with the jet. The B hadron also decays more frequently to leptons with a high transverse momentum relative to the jet than their light flavour counterparts. The properties of these leptons can also be used to identify jets containing B hadrons. It is important to note that D hadrons also possess all of these properties but to a lesser degree.

There are two main candidate algorithms that were commissioned for the 2010 dataset [62, 63]. The first is the *Track Counting* algorithm which uses the signed impact parameter significance of tracks as a discriminant. The impact parameter (IP) is defined as the distance of closest approach between a track in the jet and the reconstructed primary vertex. The sign of the impact parameter is determined by the sign of the scalar product of the impact parameter and the reconstructed jet direction. The purpose of this is to identify tracks that are inconsistent with the jet direction. Jets that feature these tracks are typically produced by light flavour decays. Finally, the signed impact parameter significance is defined simply as IP/σ_{IP}, where σ_{IP} is the uncertainty on the impact parameter.

The tracks considered are required to pass the quality criteria detailed in Reference [63]. Tracks are assigned to each jet within a cone of $\Delta R < 0.5$ between the track and the jet axis. The tracks in a jet are then ranked in decreasing order of signed impact parameter significance. The *High Efficiency* version of the algorithm uses the impact parameter significance of the second ranked track, where as the *High Purity* version uses the third track. As the tracks are ranked by impact parameter, the high purity algorithm requires more tracks above the threshold which rejects relatively more light flavour decays and thus increases the purity. For each of the different versions, working points are defined by the B-Tagging Physics Object Group which correspond to a nominal efficiency for light jets of 10% (loose), 1% (medium) and 0.1% (tight). These working points are measured in simulation on a QCD multi-jet sample. The discriminant distributions for the Track Counting algorithms in data in comparison to simulated QCD samples are shown in Figure 6.1.



(a) (b)

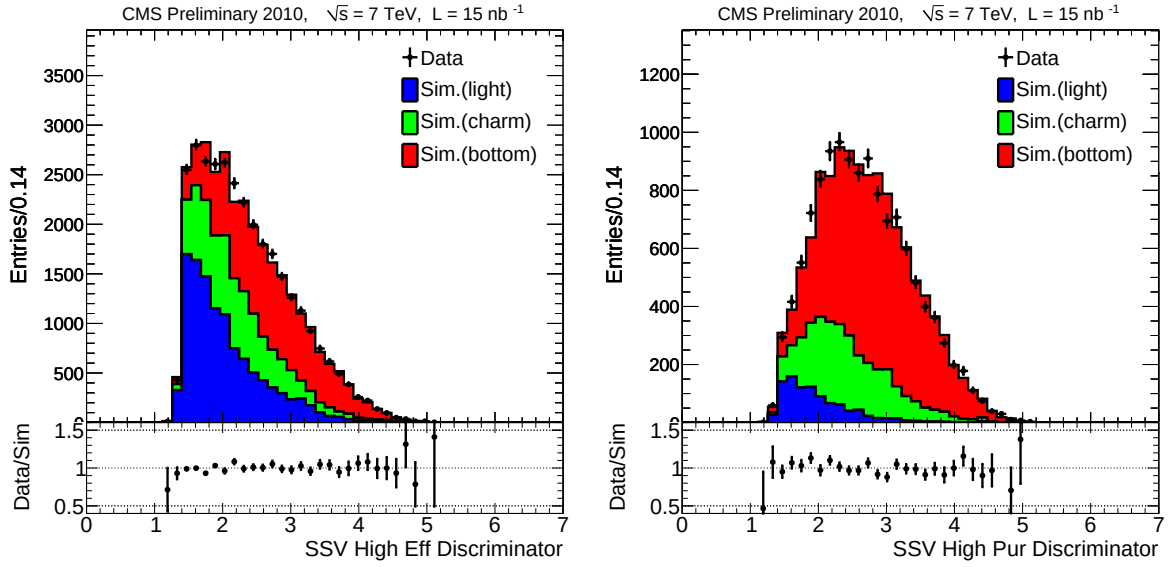
Figure 6.1: The discriminant of the Track Counting High Efficiency (TCHE) (a) and the Track Counting High Purity (TCHP) algorithms (b). The error bars on the data points indicate the statistical uncertainty.

The second candidate is the *Simple Secondary Vertex* algorithm. This algorithm uses a function of the signed, three dimensional flight distance significance of secondary vertices as a discriminant. The discriminant is calculated using equation 6.1, where D_{3D} is the signed three dimensional flight distance between the secondary vertex and the primary vertex, and $\sigma_{D_{3D}}$ is the uncertainty on this parameter.

$$SSV_{Disc} = \log\left(1 + \frac{D_{3D}}{\sigma_{D_{3D}}}\right) \tag{6.1}$$

To reconstruct the secondary vertices in an event the same track quality criteria and jet association is applied as in the Track Counting algorithm. The collection of tracks are then passed to the adaptive vertex fitter. Secondary vertices which share at least 65% of their tracks with the primary vertex are discarded. This vertex finding process does impose a maximum efficiency limit of 60 to 70% for B hadron decays. This is due to the efficiency for reconstructing a secondary vertex with a separation from the primary vertex typical of the B hadron flight distance. There are again two versions of the algorithm. The *High Efficiency* versions requires at least two tracks to reconstruct a vertex where as the *High Purity* version requires three or more. The working points are defined in the same manner as above, although in this case there is no loose working point due to the limited maximum efficiency of the vertex finding process. The discriminant distributions for the Simple Secondary Vertex algorithms in data compared to simulated QCD multi-jet samples are shown in Figure 6.2.

The algorithm chosen for this particular analysis is the Simple Secondary Vertex High Efficiency (SSVHE) at the medium working point. This corresponds to a reasonable trade off in b-jet identification efficiency and purity for the original target data luminosity of 20 pb^{-1}. The Track Counting High Efficiency algorithm does yield similar performance at the medium working point. However, indications from studies conducted in simulation of possible tracker misalignment and pixel detector failure scenarios in early data suggest that the Simple Secondary Vertex b-tagger is potentially more robust. This particular study is detailed elsewhere [64].



(a) (b)

Figure 6.2: The discriminant of the Simple Secondary Vertex High Efficiency (a) and the Simple Secondary Vertex High Purity algorithms (b). The error bars on the data points indicate the statistical uncertainty.

6.3 Jet tagging efficiency measurements with data

There are a number of possible techniques with which the efficiency to identify b-jets (ε_{b}) and the efficiency to mistag light-jets (ε_{l} or mistag rate) can be measured. There are two primary methods for measuring ε_{b} in early data at CMS: the p_{Trel} method [65] and the System 8 method [66, 67]. In each case it is important to note that both measure the efficiency for the specific case where the B hadron decays leptonically. The correction to inclusive decays is expected to be small and is determined through simulation.

The Mistag method as the name suggests measures the mistag rate for light-jets [68].

6.3.1 p_{Trel} method

The p_{Trel} method makes use of di-jet events in which one jet is identified as having a reconstructed muon, referred to as the muon jet. Muons are associated with a jet within a cone of $\Delta R < 0.4$ between the muon and the jet axis. Events where multiple muons fulfil this criteria are rejected. The second jet in the event (the away jet) is required to pass the Track Counting High Purity lifetime tagger at the medium working point to enhance the b purity of the sample. The transverse momentum of the muon calculated with respect to the muon jet axis is referred to as p_{Trel}. The p_{Trel} of muons originating from B hadron decays has a tendency to be larger than the other flavours due to the larger mass of the hadron. The distribution of the muon p_{Trel} from simulation and data is given in Figure 6.3.

Templates of the p_{Trel} distributions separated into b and combined c and light flavours are produced from simulation. A binned likelihood fit is then performed with the templates to measure the relative fraction of b- and c+light-jets in the event. The b-tagging algorithm which is being studied is then applied to the di-jet sample and the fraction measurements are repeated allowing for the number of b-jets to be counted before and after the application of the tagger and thus deriving the b-tagging efficiency.

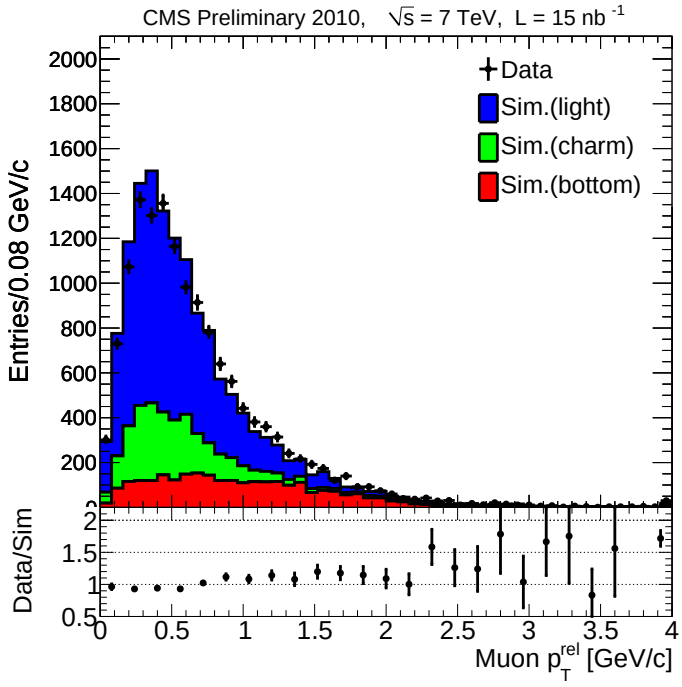


Figure 6.3: The muon p_{Trel} distribution from simulation and data separated into b, c and light flavours. The error bars on the data points indicate the statistical uncertainty.

6.3.2 System 8 method

The System 8 method can also be used to measure the b-tagging efficiency in muon-jet data. To do this it uses di-jet data from which a number of sub-samples are constructed with varying degrees of b-jet purity. Initially, one sub-sample is produced from the original sample by requiring that the non-muon jet (away jet) pass a lifetime-based tagger. The original sample is referred to as n and the sub-sample as p. The following sub-samples are then generated separately for both the n sample and p sub-sample:

- A sub-sample in which the muon-jet passes the lifetime tagger being studied.
- A sub-sample in which the muon-jet passes a p_{Trel} cut.
- A sub-sample in which the muon-jet passes both the lifetime tagger and the p_{Trel} cut.

The sample and the various sub-samples are then related by a system of equations which separates out the number of b-jets (labelled with the b subscript) and the number of c and light-jets (labelled with $c\ell$) by incorporating the selection efficiencies of the cuts (ε^{tag} and $\varepsilon^{p_{Trel}}$) and the correlations between the different cuts and the away-jet b-tagger. The equations are as follows:

$$\begin{aligned}
n &= n_b + n_{c\ell} \\
p &= p_b + p_{c\ell} \\
n^{tag} &= \varepsilon_b^{tag} n_b + \varepsilon_{c\ell}^{tag} n_{c\ell} \\
p^{tag} &= \beta_{12}\, \varepsilon_b^{tag} p_b + \alpha_{12}\, \varepsilon_{c\ell}^{tag} p_{c\ell} \\
n^{p_{Trel}} &= \varepsilon_b^{p_{Trel}} n_b + \varepsilon_{c\ell}^{p_{Trel}} n_{c\ell} \\
p^{p_{Trel}} &= \beta_{23}\, \varepsilon_b^{p_{Trel}} p_b + \alpha_{23}\, \varepsilon_{c\ell}^{p_{Trel}} p_{c\ell} \\
n^{tag,p_{Trel}} &= \beta_{13}\, \varepsilon_b^{tag} \varepsilon_b^{p_{Trel}} n_b + \alpha_{13}\, \varepsilon_{c\ell}^{tag} \varepsilon_{c\ell}^{p_{Trel}} n_{c\ell} \\
p^{tag,p_{Trel}} &= \beta_{123}\, \varepsilon_b^{tag} \varepsilon_b^{p_{Trel}} p_b + \alpha_{123}\, \varepsilon_{c\ell}^{tag} \varepsilon_{c\ell}^{p_{Trel}} p_{c\ell}\ .
\end{aligned} \tag{6.2}$$

The correlation parameters for b-jets are the β parameters and the α parameters are the corresponding parameters for the combined c and light-jets. The correlation factors are measured using simulated samples. The equations are then solved numerically on the data sample to extract the b-tagging efficiency.

6.3.3 Mistag method

The mistag rate or light-jet tagging efficiency is measured using the "negative tags" of the signed discriminant. The negative tags for the Simple Secondary Vertex algorithms are defined as those with negative three dimensional decay lengths that have a discriminant greater in absolute magnitude than the working point cut. Similarly, the negative tags for the Track Counting algorithms rank the tracks by increasing signed impact parameter significance (as opposed to decreasing). As before with the Track Counting algorithm, the second ranked track is used for the High Efficiency negative

tag and the third ranked for the High Purity negative tag. The positive and negative discriminant for the Track Counting and Simple Secondary Vertex algorithms are shown in Figure 6.4. These negative tags are a result of misreconstruction and as such are highly enriched in light flavour jets as they have a tendency to be closest to the primary vertex.

The mistag rate in data is measured from the negative tag rate by defining the ratio $R_{\text{light}} = \varepsilon^{\text{mistag}}/\varepsilon^{-}$, where $\varepsilon^{\text{mistag}}$ is the mistag rate and ε^{-} is the negative tag rate. This ratio is measured in simulation. The systematic uncertainty on R_{light} takes into account a number of discrepancies observed between data and simulation. In particular, the gluon content was found to be larger than predicted as well as the number of relatively long lived light hadrons such as K_s and Λ particles. The consideration of these effects is discussed in more detail elsewhere [63].

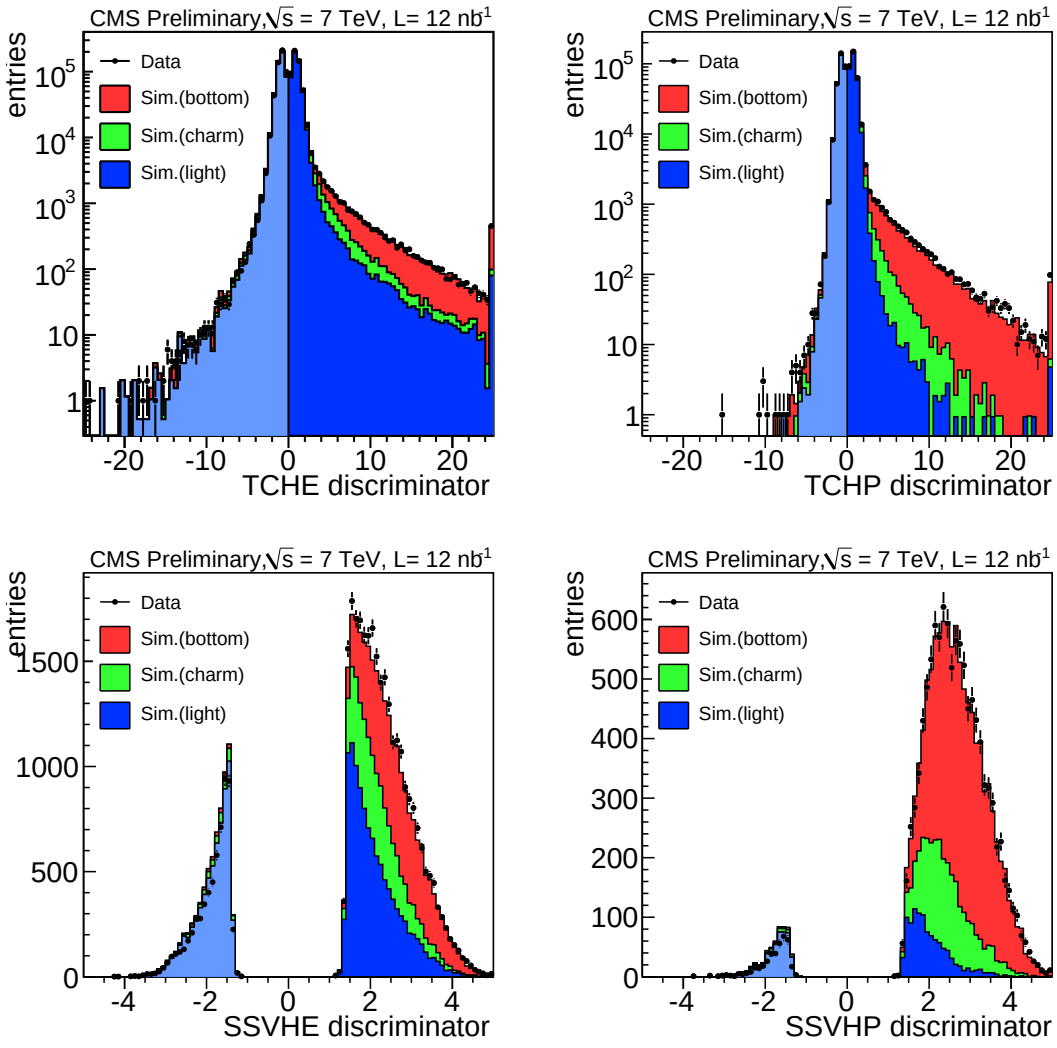


Figure 6.4: The positive and negative discriminants of the Track Counting (TC) and Simple Secondary Vertex (SSV) algorithms in data and simulation. Shown for both high efficiency (HE) and high purity (HP) version of each algorithm. The error bars on the data points indicate the statistical uncertainty. The lighter shade of blue indicates the light-jets with a negative discriminant

6.3.4 Results summary

The results for the b-jet tagging efficiency and light-jet mistag rates are provided as scale factors, SF_{b} and SF_{l}, respectively. The scale factor is defined in equation 6.3, where $\varepsilon^{\mathrm{Data}}$ is the measured efficiency on data and $\varepsilon^{\mathrm{MC}}$ is the measured efficiency in simulated QCD multi-jet samples:

$$SF = \frac{\varepsilon^{\mathrm{Data}}}{\varepsilon^{\mathrm{MC}}} \tag{6.3}$$

The advantage of using a scale factor instead of raw efficiencies is that the scale factor can be used to correct the efficiencies measured in simulation for samples with differing topologies in contrast to the di-jet samples for the b-tagging efficiencies and the QCD multi-jet samples for light-jet mistag rate. The obvious and relevant example for this analysis would be top pair decays in the lepton+jets channel, which have, on average, more jets reconstructed per event.

The scale factor for the b-tagging efficiency for 2010 data provided by the B-Tagging Physics Object Group is $SF_{\mathrm{b}} = 0.9$ with a total uncertainty of 15%. This is based on the range of the preliminary results provided by the p_{Trel} and System 8 methods, discussed previously, as well as analysis of tag multiplicities in di-jet events and track reconstruction efficiencies. The scale factor for the light-jet mistag rate is provided by the Mistag method and is parameterised in the absolute value of pseudorapidity and transverse momentum of the jet. The scale factor for the mistag rate is shown in Figure 6.5.

The efficiency to tag c-jets must also be corrected with an appropriate scale factor. However, no c-jet tagging efficiency measurements are currently available. An estimate of the c-jet efficiency can be provided by applying the scale factor for b-jets to the efficiency to tag c-jets measured in simulation. The justification for this strategy is that both B and D hadrons produce real displaced vertices with measurable decay lengths and as such their detector response should be comparable to a certain degree. This is difficult to prove at this stage without any dedicated measurements of the c-jet

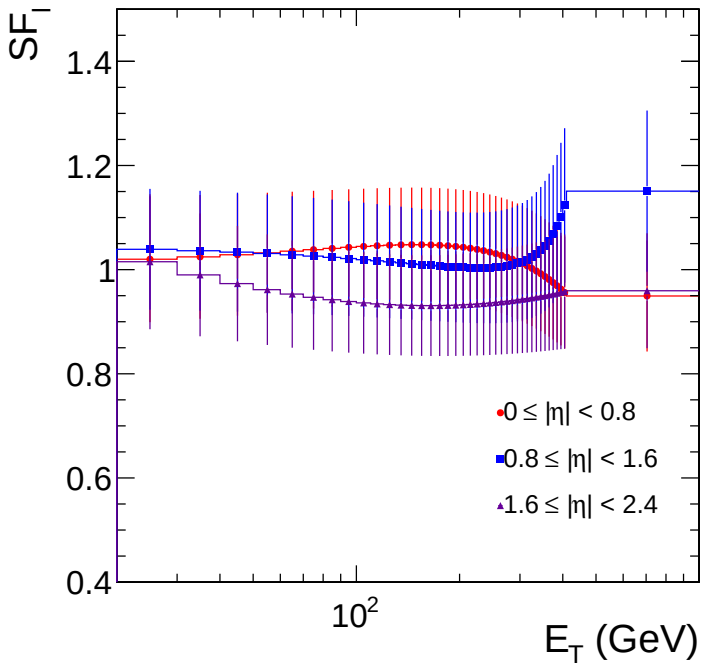


Figure 6.5: The scale factor for correcting the light-jet mistag rate on data. Error bars show the total uncertainty on each measurement.

measurements. However, the excellent performance of the tracker (Section 3) and the comparison of the SSV discriminants with simulation in Figure 6.2 suggest that it is reasonable to expect that D hadrons produce a measurable secondary vertex in CMS, albeit with a lower reconstruction efficiency. The uncertainty on this scale factor when applied to c-jets is doubled to account for this assumption.

More recently measurements for the 2011 data-set have been produced which can be found elsewhere [69]. These results are still numerically consistent within systematic uncertainties with the measurements for the 2010 data-set, although it is important to note that they can not be directly applied due to changes in the vertex reconstruction software and the vastly different pileup environment.

6.4 Measuring the event tagging efficiency

6.4.1 Jet tagging efficiency in $t\bar{t}$ simulation

To apply the scale factors some parameterisation of the jet tagging efficiencies and mistag rates must be measured in $t\bar{t}$ simulation. The identification of b-jets is fundamentally tied to the reconstruction of tracks and their association to jets and vertices. Anything that affects the tracking efficiency and track position resolution would be expected in turn to affect the efficiency and mistag rates. This includes a number of specific jet characteristics, such as the transverse momentum, track multiplicity in the jet and the pseudorapidity. In addition to the characteristics of jets, the reconstruction efficiency may be impaired by a large amount of activity in the event. The quantity H_T, which is defined as the scalar sum of the jet transverse momenta and electron transverse momentum in equation 6.4, can be thought of as a measure of the overall event activity.

$$H_T = \sum jet_{p_T} + \sum electron_{p_T} \tag{6.4}$$

The distributions of the jet tagging efficiencies in simulation are shown in Figure 6.6 against the transverse momentum of the jet, the absolute value of pseudorapidity of the jet, the number of tracks in the jet (N_{trk}) and the H_T of the event. The efficiency is calculated from the ratio of tagged jets of a particular flavour to the total number of jets of a particular flavour in each of the bins. The efficiencies are measured after the standard selection defined in Section 5.2 has been applied, excluding the jet multiplicity cuts.

The parameterisation chosen for the analysis is the jet efficiency vs transverse jet momentum. This distribution shows a strong dependence for all of the jet flavours and in particular, the mean transverse momentum of jets in the $t\bar{t}$ simulation sample is approximately 60 GeV/c^2, a region in which the efficiency is rising quite rapidly with increasing jet p_{T}.

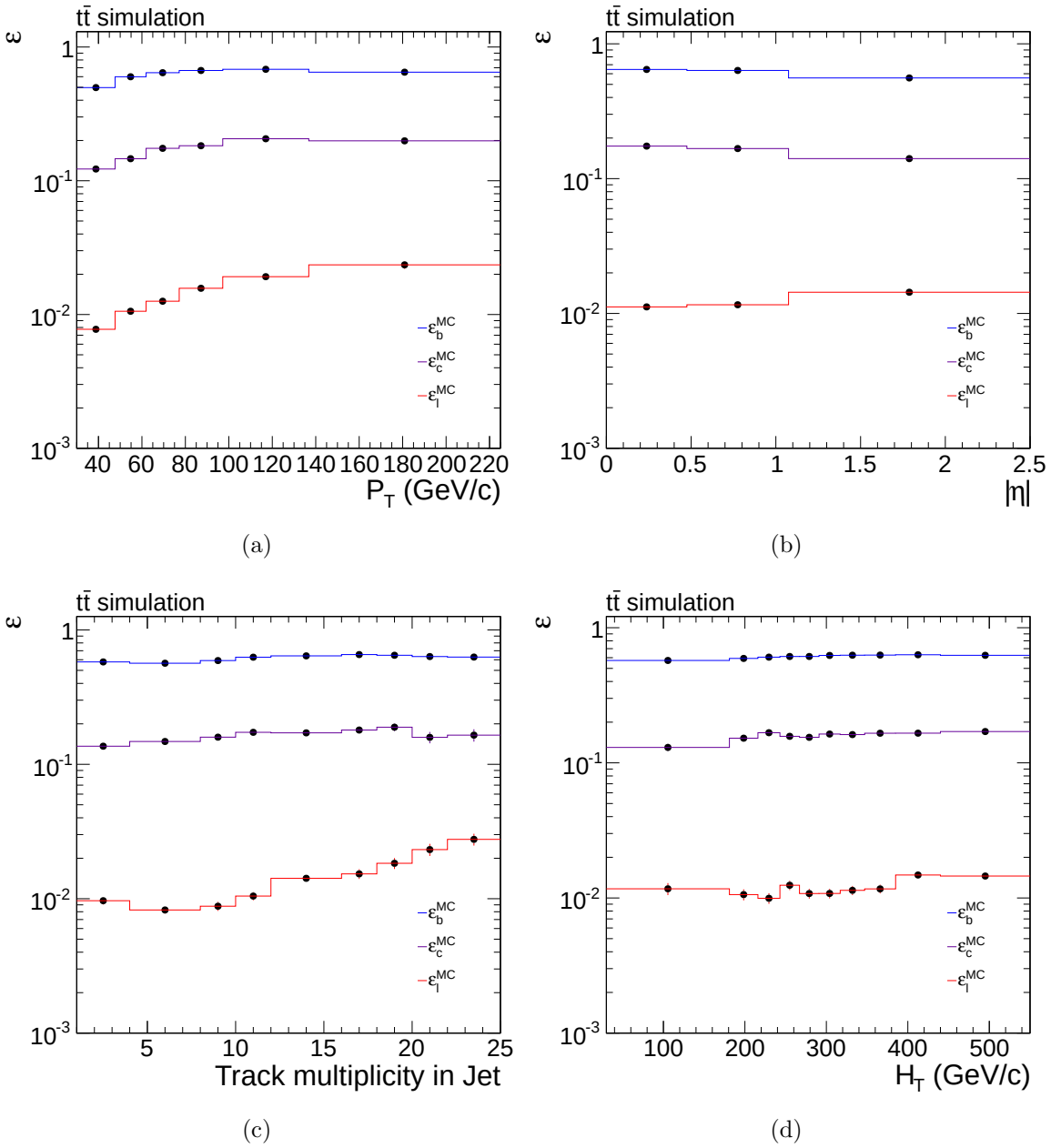


Figure 6.6: Jet tagging efficiency for b, c and l-jets vs p_T (a), $|\eta|$ (b), N_{trk} (c) and H_T (d), for the Simple Secondary Vertex algorithm in $t\bar{t}$ simulation. Error bars indicate the uncertainty due to simulation statistics.

6.4.2 Calculating the probability to tag an event

The selection efficiency for the one or more b-tagged jets requirement (ϵ^{tag}) is equivalent to the average of the probability that at least one of the jets in a given event is tagged as a b-jet. The average probability for a given jet in an event to be tagged is provided by the jet-level efficiencies corresponding to the characteristics of that jet. Therefore, the probability for an individual event to pass this selection cut which contains $nJets$ with b-tagging efficiencies of ε_i is given by equation 6.5.

$$\begin{aligned} P(\geq 1 \text{ b-tag}) &= 1 - P(0 \text{ jets tagged}) \\ &= 1 - \prod_{i}^{nJets} (1 - \varepsilon_i) \end{aligned} \tag{6.5}$$

As an additional exercise, this logic can also be extended to a selection cut which requires two or more b-tagged jets. For this case, the additional probability to tag only a single jet exclusively is required as given in equation 6.6, again for $nJets$ with efficiencies of ε_i. This does include the assumption that the probability of tagging each individual jet is uncorrelated with each other.

$$P(1 \text{ jet tagged}) = \sum_{i}^{nJets} \varepsilon_i \prod_{j;j\neq i}^{nJets} (1 - \varepsilon_j) \tag{6.6}$$

The probability to tag two or more jets in a given event is thus given by equation 6.7.

$$\begin{aligned} P(\geq 2 \text{ jets tagged}) &= 1 - P(1 \text{ jet tagged}) - P(0 \text{ jets tagged}) \\ &= 1 - \{ \sum_{i}^{nJets} \varepsilon_i \prod_{j;j\neq i}^{nJets} (1 - \varepsilon_j) \} - \prod_{k}^{nJets} (1 - \varepsilon_k) \end{aligned} \tag{6.7}$$

These equations can be used to correct the jet multiplicity distributions after the b-tagging selection has been applied from the events remaining at the previous selection cut. For each event remaining the probability that it would pass the selection is calculated and used to weight the event in the jet multiplicity distribution. To test the method, a "closure test" is performed on the $\text{t}\bar{\text{t}}$ simulation sample using the measured

jet efficiencies in Figure 6.6a without applying any scale factors from data. The resulting jet multiplicity distribution is shown in Figure 6.7. The discrepancy between the number of reconstructed and the number of counted events without scale factors (-0.3% for three or more jets and -0.2% for four or more jets) is taken into account as the systematic uncertainty due to the model in Section 8.1.

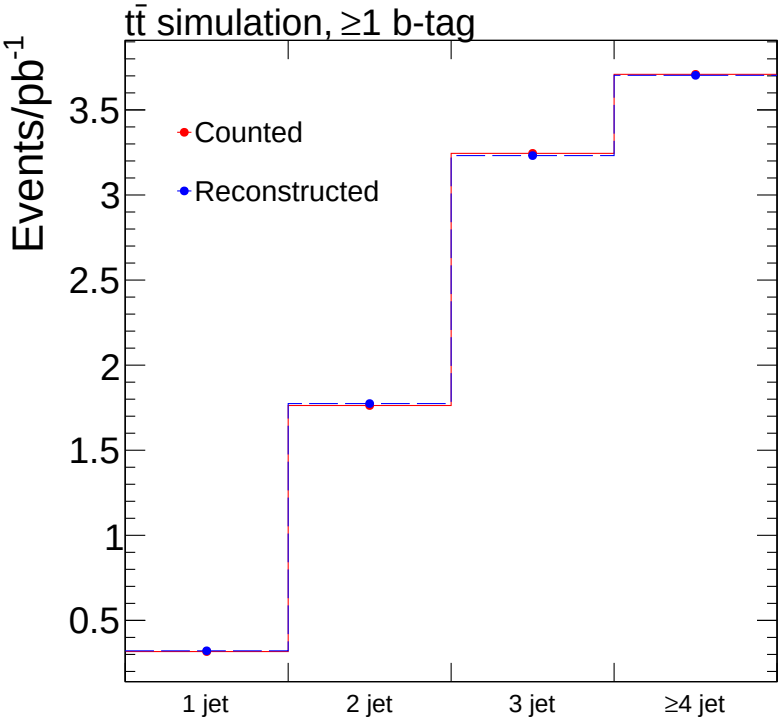


Figure 6.7: The reconstructed number of events per inverse picobarn of integrated luminosity in comparison to the counted number of events in $t\bar{t}$ simulation after full event selection is applied without any data-driven scale factors. Statistical uncertainties are included but not visible on this scale.

6.4.3 $t\bar{t}$ event tagging efficiency

The overall event efficiency is defined as the number of reconstructed events that pass the b-tagging requirement with the method above (N^{tag}), divided by the number of events before b-tagging ($N^{\text{pre-tag}}$) as in equation 6.8:

$$\epsilon^{\text{tag}} = \frac{N^{\text{tag}}}{N^{\text{pre-tag}}} \tag{6.8}$$

To measure the selection efficiency for $t\bar{t}$ events in data, the measured scale factors from Section 6.3.4 are applied to the jet-level efficiencies in the reconstruction to measure N^{tag}. The final selection efficiencies for one or more b-tagged jets and two or more b-tagged jets are shown in Table 6.1.

Table 6.1: Efficiency to tag one or more jets and two or more jets in the event (ϵ^{tag}) measured with data-driven scale factors applied to the $t\bar{t}$ jet-level efficiencies after full selection on a $t\bar{t}$ simulation sample. Quoted uncertainties are due to simulation statistics only.

	≥ 3 Jets	≥ 4 Jets
≥ 1 tagged jet	0.727 ± 0.004	0.755 ± 0.005
≥ 2 tagged jets	0.242 ± 0.001	0.276 ± 0.002

Chapter 7

Background Estimations

7.1 Introduction

The standard model backgrounds to this process are primarily composed of events with leptonically-decaying $W^{\pm}$ bosons with additional jets (W+jets), leptonically-decaying Z^0 bosons with additional jets (Z+jets), QCD multi-jet production and single-top production. There are additional small contributions from photon production with jets and electroweak di-boson production processes. The photon+jets contribution is typically included with the QCD estimation and will not be addressed separately. The estimated number of events from simulation for these processes after the event selection are given in Table 5.4. The following sections will discuss how each of the backgrounds are estimated in this analysis.

7.2 Background from W+jets events

The W+jets background estimation proceeds in two steps. Firstly, an estimation of the number of W+jets events before the application of the b-tagger is made ($N_{\mathrm{W+jets}}^{\mathrm{pre-tag}}$). Secondly, the simulation is used to calculate a data-corrected b-tagging selection effi-

ciency as in Section 6.4.2, to estimate the number of b-tagged W+jets events remaining after the complete selection.

7.2.1 W+jets pre-tag estimate

To determine the number of pre-tagged W+jets events remaining in the data sample a process known as Berends Scaling is exploited [70]. The principle is that to produce an extra jet there must be an extra QCD vertex which would introduce a factor of $1/\alpha_s^2$ in the cross-section. Thus for an ideal system where each additional final state parton produced a jet, the ratio of the inclusive W+ $\geq n$Jets production to the W+ $\geq (n+1)$Jets production should be constant (defined here as C). This theoretically enables the W+jets yield to be measured in the lower jet multiplicities where the signal fraction is greatly reduced and extrapolated to the higher jet multiplicities used in the event selection.

The application of this in a real analysis is more complicated. Each additional parton will not necessarily produce additional jets due to selection and reconstruction efficiencies, and due to effects such as jet merging if the parton is produced sufficiently collinear with an existing jet. In practice the constant is treated as a linear function of the jet multiplicity which should better approximate the diminishing returns from producing extra partons. It is also difficult to extract a pure sample of W+jets in the lower jet multiplicities. To suppress the QCD background, which is also typically very large in the lower jet multiplicities, an additional requirement of a missing transverse energy greater than 20 GeV is imposed. The remaining QCD contribution is measured using the technique discussed in Section 7.3. The contribution from other processes are also non-negligible, including $\mathrm{t\bar{t}}$ production. These contributions are estimated from simulation. The remaining events in data as well as the other processes before b-tagging and with the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut are summarised in Table 7.1.

The contributions from other samples are subtracted and a linear fit is performed to extract $C(n)$. The results of the fit in data are shown in Figure 7.1 along with the

Table 7.1: The number of pre-tagged events remaining after the application of 20 GeV $E_{\mathrm{T}}^{\mathrm{miss}}$ selection. QCD estimation is measured in data as in Section 7.3, the other contributions are estimated from simulation and include only statistical uncertainties.

	Data	$\mathrm{t\bar{t}}$	Single top
1 jet	11682.00 ± 108.08	25.87 ± 0.34	39.78 ± 0.24
2 jets	2871.00 ± 53.58	98.46 ± 0.66	44.05 ± 0.25
3 jets	739.00 ± 27.18	153.88 ± 0.83	20.55 ± 0.16
4 jets	242.00 ± 15.56	108.79 ± 0.69	6.46 ± 0.09
	Z+jets	Di-Boson	QCD
1 jet	170.88 ± 2.73	50.99 ± 0.19	$1444.19^{+57.73}_{-57.25}(\mathrm{stat.}) \pm 30.38(\mathrm{syst.})$
2 jets	62.98 ± 1.66	31.09 ± 0.14	$413.76^{+28.88}_{-28.94}(\mathrm{stat.}) \pm 29.94(\mathrm{syst.})$
3 jets	17.72 ± 0.88	7.80 ± 0.07	$128.45^{+16.68}_{-16.59}(\mathrm{stat.}) \pm 16.14(\mathrm{syst.})$
4 jets	4.50 ± 0.44	1.46 ± 0.03	$32.36^{+5.27}_{-5.20}(\mathrm{stat.}) \pm 13.61(\mathrm{syst.})$

pre-tag results from three different simulation samples. The q^2 up and down samples refer to systematic samples for W+jets production in which the factorisation and renormalisation scales controlled by this parameter q^2 in the generator configuration are increased and decreased by a factor of 2.0 and 0.5 respectively. The original scale for the process is set to $q^2 = (m_W)^2 + (\sum p_{\mathrm{T}}^{jet})^2$, with $m_W = 80.398$ GeV. This parameter in particular is sensitive to the jet multiplicity in the simulated sample.

The reason for the sensitivity to the q^2 parameter is as follows. The renormalisation scale is what determines the energy available to the hard process (the part of the event considered to be "high energy"), which if it is increased could lead to additional jet production or higher energy jets more likely to pass the selection threshold. The factorisation scale is what determines the separation point between what is a part of the hard process and what is part of the underlying event. In practical terms, this determines if a parton has sufficient energy to be treated as a separate component of the event which interacts directly with the rest of the hard process and could potentially

lead to a jet. The q^2 systematic samples are not used as part of the determination of $C(n)$ as this is measured in the data-sample, and are included for comparison with the data and the primary W+jets simulated sample. The value for $C(n)$ is shown in Table 7.2.

Table 7.2: The measured value of $C(n)$ for the one, two and three jet multiplicities.

central value $\pm$(fit.const) $\pm$ (fit.slope)			
	n=1 jet	n=2 jets	n= 3jets
$C(n)$	$4.912 \pm 0.125 \pm 0.028$	$5.353 \pm 0.125 \pm 0.100$	$5.794 \pm 0.125^{+0.173}_{-0.172}$

The final W+jets pre-tag estimate is calculated using the number of pre-tagged W+jets events in the one or more jets bin. For this, the pre-tag numbers for the other samples are once again subtracted using the same procedure as above without the missing transverse energy selection. The estimate for the number of pre-tagged W+jets in the three or more and four or more jet bin are calculated with equation 7.1 and equation 7.2 respectively.

$$N^{\text{pre-tag}}_{\text{W+jets},\geq 3\text{jets}} = \frac{N^{pre-tag,\text{data}}_{\text{W+jets},\geq 1\text{jets}}}{C(1)C(2)} \tag{7.1}$$

$$N^{\text{pre-tag}}_{\text{W+jets},\geq 4\text{jets}} = \frac{N^{pre-tag,\text{data}}_{\text{W+jets},\geq 1\text{jets}}}{C(1)C(2)C(3)} \tag{7.2}$$

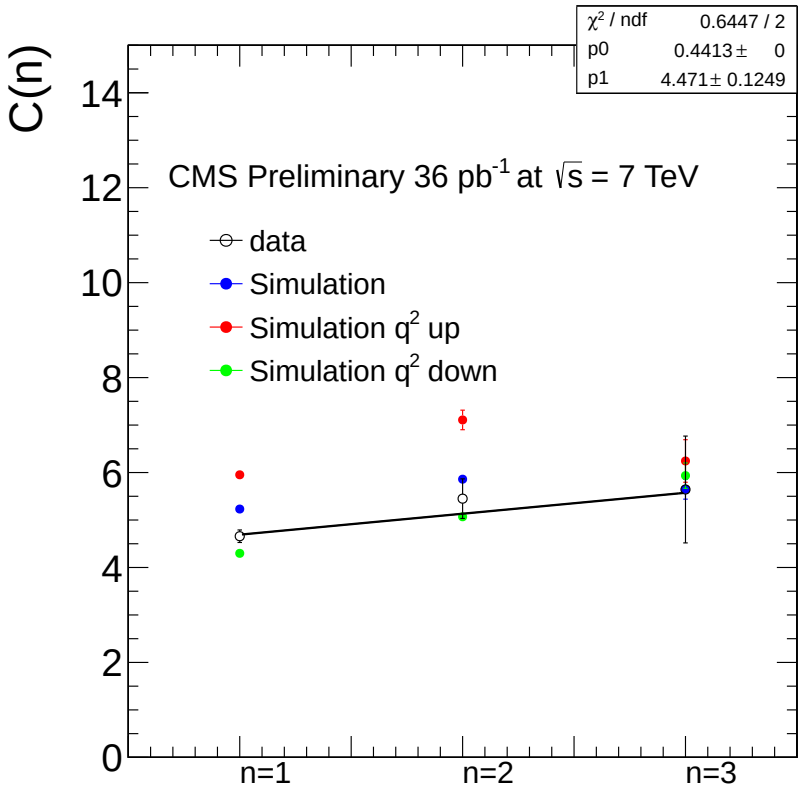


Figure 7.1: The ratio of pre-tagged W+$\geq n$jets to W+$\geq (n+1)$jets on data and in simulation. Also shown are the two q^2 systematic simulated samples. The error bars indicate statistical uncertainties. The resulting fit to the data is indicated by the black line.

7.2.2 W+jets b-tagged estimate

To proceed from the pre-tagged W+jets estimation to determine the number remaining after the b-tagging selection, the sample is separated into light flavour and heavy flavour sub-samples. The separation of these two sub-samples in simulation is a non-trivial process with the generator information that is available and a dedicated tool has to be employed to attempt to identify the flavour production processes in the event. The *Flavour History Tool* and its usage with W+jets is discussed in Appendix B.1.

The motivations for separating the sample by flavour are two fold. The light and heavy flavour sub-samples will obviously have very different average efficiencies with the former being dominated by mistags and the latter real heavy flavour. As well as this, the b-quark components of the heavy flavour sub-sample have to be re-weighted to reflect recent measurements of the relative W+b($\mathrm{b\bar{b}}$) contribution in W+jets events which are poorly reproduced in simulation at the higher jet multiplicities [71]. To accomplish this, the Top Physics Analysis Group decided to apply a scale factor, $K_{\mathrm{b}} = 2.0 \pm 1.0$, which provides a conservative estimation based on the available results. For the W+c($\mathrm{c\bar{c}}$) component, no re-weighting is necessary, although a scale factor $K_{\mathrm{c}} = 1.0^{+1.0}_{-0.5}$ is used for systematic studies.

To calculate the b-tagging event selection efficiency for the separate W+jets contributions, the same method as in Section 6.4.2 is applied. The jet-level b-tagging efficiency measurements are repeated for the inclusive W+jets sample and are shown in Figure 7.2. The data scale factors are then used to calculate the selection efficiency for each of the different flavour categories in the W+jets simulation sample. The results for the different sub-samples and their overall fraction, F_i, are given in Table 7.3 for three or more jets and in Table 7.4 for four or more jets. The fraction shown is after the re-weighting of the W+b($\mathrm{b\bar{b}}$) contribution. The abbreviations refer to the matching done by the Flavour History Tool with GenJets to the matrix element (ME) or parton shower (PS). The ΔR component of each flavour path refers to the separation between matched GenJets as detailed in Appendix B.1.

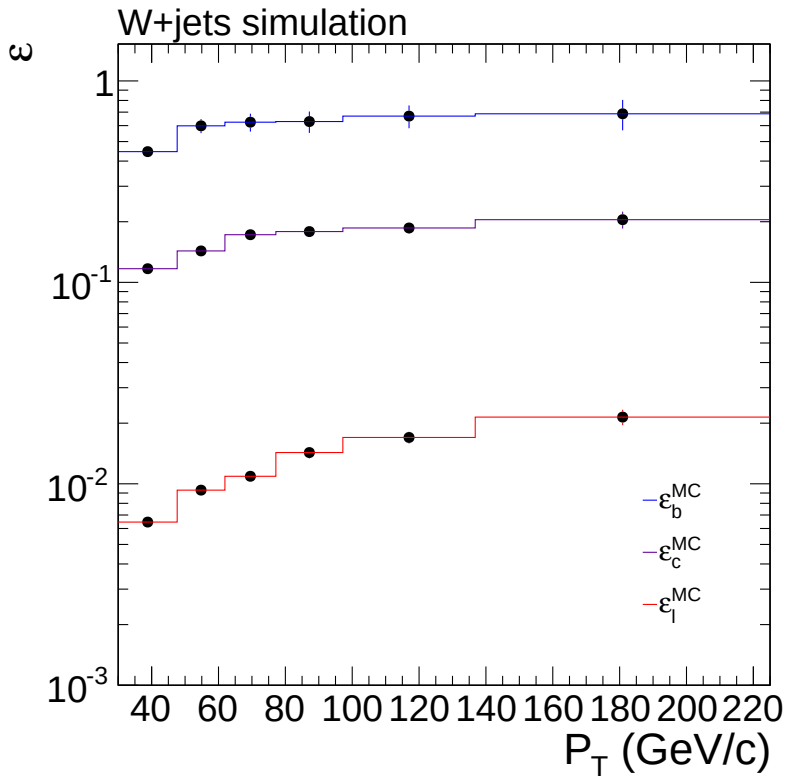


Figure 7.2: Jet tagging efficiency for b, c and l-jets vs the transverse momentum of the jet, for the Simple Secondary Vertex algorithm measured in W+jets simulation. The error bars indicate statistical uncertainties.

Table 7.3: The event tagging efficiency, ϵ_i^{tag}, and re-weighted fractional contribution, F_i, for W+jets events separated by flavour production paths, for selected events with three or more jets. The abbreviations refer to matching to the matrix element (ME) and the parton shower (PS). The ΔR component of each flavour path refers to the separation between matched GenJets.

Sub-sample	ϵ_i^{tag}	F_i
W+b$\bar{\text{b}}$, 2 jets from ME (ΔR>0.5)	0.603 ± 0.074	0.04
W+b or W+b$\bar{\text{b}}$, 1 jet from ME	0.344 ± 0.097	0.0087
W+c$\bar{\text{c}}$, 2 jets, from ME (ΔR>0.5)	0.207 ± 0.019	0.034
W+c or W+c$\bar{\text{c}}$, 1 jet from ME	0.131 ± 0.007	0.12
W+$g \to b\bar{b}$ 1 jet from PS	0.272 ± 0.068	0.012
W+$g \to c\bar{c}$ 1 jet from PS	0.063 ± 0.009	0.021
W+b$\bar{\text{b}}$, 1 jet from ME(ΔR=0.0)	0.540 ± 0.279	0.0023
W+c$\bar{\text{c}}$, 1 jet from ME(ΔR=0.0)	0.180 ± 0.098	0.001
W+$g \to b\bar{b}$, 2 jets from PS(ΔR>0.5)	0.403 ± 0.052	0.04
W+$g \to c\bar{c}$, 2 jets from PS(ΔR>0.5)	0.121 ± 0.010	0.045
W+light	0.050 ± 0.001	0.68

Table 7.4: The event tagging efficiency, ϵ_i^{tag}, and re-weighted fractional contribution, F_i, for W+jets events separated by flavour production paths, for selected events with four or more jets. The abbreviations refer to matching to the matrix element (ME) and the parton shower (PS). The ΔR component of each flavour path refers to the separation between matched GenJets.

Sub-sample	ϵ_i^{tag}	F_i
W+b$\bar{\text{b}}$, 2 jets from ME(ΔR>0.5)	0.655 ± 0.142	0.060
W+b or W+b$\bar{\text{b}}$, 1 jet from ME	0.337 ± 0.211	0.0071
W+c$\bar{\text{c}}$, 2 jets from ME (ΔR>0.5)	0.231 ± 0.050	0.031
W+c or W+c$\bar{\text{c}}$, 1 jet from ME	0.143 ± 0.017	0.100
W+$g\rightarrow b\bar{b}$ 1 jet from PS	0.358 ± 0.211	0.0084
W+$g\rightarrow c\bar{c}$ 1 jet from PS	0.083 ± 0.022	0.024
W+b$\bar{\text{b}}$, 1 jet from ME(ΔR=0.0)	0.606 ± 0.495	0.0043
W+c$\bar{\text{c}}$, 1 jet from ME(ΔR=0.0)	0.000 ± 0.000	0
W+$g\rightarrow b\bar{b}$, 2 jets from PS(ΔR>0.5)	0.500 ± 0.116	0.056
W+$g\rightarrow c\bar{c}$, 2 jets from PS(ΔR>0.5)	0.154 ± 0.023	0.065
W+light	0.069 ± 0.004	0.64

The final contribution for the number of b-tagged W+jets separated into heavy and light flavour sub-samples is calculated using equation 7.3, where "x" refers to either light or heavy.

$$N^{\text{tag}}_{\text{W+x}} = (\sum_{i}^{path} F_i \varepsilon_i^{tag}) \cdot N^{\text{pre-tag}}_{\text{W+jets}} \tag{7.3}$$

The results for the number of b-tagged W+light flavour events are as follows:

- $N^{\text{tag}}_{\text{W+light}} = 23.1 \pm 0.2(\text{stat}) \pm 0.2(\text{stat.MC}) \pm 0.8(\text{fit.const}) \pm 0.5(\text{fit.slope})$ for three or more jets.
- $N^{\text{tag}}_{\text{W+light}} = 5.22 \pm 0.05(\text{stat}) \pm 0.05(\text{stat.MC}) \pm 0.21(\text{fit.const}) \pm 0.19(\text{fit.slope})$ for four or more jets.

The results for the number of b-tagged W+heavy flavour events are as follows:

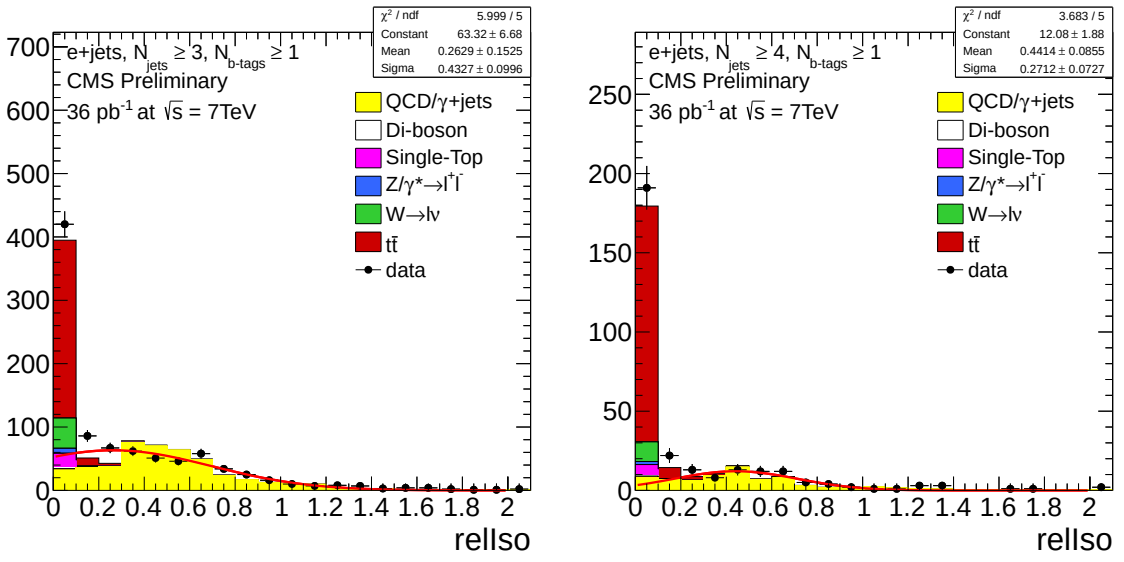
- $N^{\text{tag}}_{\text{W+heavy}} = 52.5 \pm 0.5(\text{stat}) \pm 0.5(\text{stat.MC}) \pm 1.8(\text{fit.const}) \pm 1.0(\text{fit.slope})$ for three or more jets.
- $N^{\text{tag}}_{\text{W+heavy}} = 12.78 \pm 0.13(\text{stat}) \pm 0.13(\text{stat.MC}) \pm 0.52(\text{fit.const})^{+0.46}_{-0.45}(\text{fit.slope})$ for four or more jets.

7.3 Background from QCD events

The QCD background is reduced greatly by the selection criteria discussed previously. In particular, there is no source of single real isolated electrons in these processes and as such it is characterised by fake isolated electrons. Of course electrons can be produced in the final state as part of the decay chain but these are not isolated electrons in the vast majority of cases. As such, the variable RelIso of the electron, as defined in equation 5.2, is used to characterise the QCD distribution in regions of higher RelIso where QCD dominates. The fits in the poorly isolated regions can then be extrapolated into the signal region.

To perform this background estimation, a function is fitted to the QCD distribution in the RelIso region from 0.2 to 1.0 after the full selection has been applied, but without the RelIso cut in the electron selection. A variety of functional forms for this purpose were studied in simulation to attempt to find the one that provided the best agreement with the QCD simulation in the signal region. The function that performed the best in simulation and the one that is used for the central value is the Gaussian function. It is important to note that there is no physical motivation behind the choice of the functional form. A number of other functional forms are used as part of the systematic studies for this estimate (see Section 8.4).

The fit results are shown in Figure 7.3 and summarised in Table 7.5. The estimation for the number of b-tagged QCD events in the signal region is found by integrating the fitted function across the signal region (0.0 to 0.1).



(a) (b)

Figure 7.3: The fit to the RelIso distribution after selection excluding the RelIso cut in the three or more jet bin (a), and four or more jet bin (b).

Table 7.5: Number of b-tagged QCD events estimated in data with the RelIso method for three or more and four or more jets. The given uncertainty is due to the uncertainty on the parameters in the fit.

Multiplicity	$N_{\mathrm{QCD}}^{\mathrm{tag}}$	χ^2/d.o.f
$\geq$3	56.0 ± 17.5(fit)	6.00/5
$\geq$4	4.29 ± 3.87(fit)	3.68/5

7.4 Background from Z+jets events

The majority of Z+jets events are removed during event selection with the veto applied to the di-electron invariant mass around the Z^0 mass window. The remaining contribution from Z+jet events is measured in simulation. The cross-section for Drell-Yan production for di-lepton invariant masses greater than 50 GeV is 3048 ± 187 pb, calculated with FEWZ at NNLO [72]. This includes only decays to di-leptons. An uncertainty of 30% on the cross-section is assumed as part of a conservative estimate as mandated by the Top Physics Analysis Group. The number of tagged Z+jet events for three or more and four or more jets is given in Table 7.6, calculated by scaling the remaining events with the cross-section above and the integrated luminosity.

Table 7.6: The number of Z+jet events after selection estimated from simulation with statistical uncertainties for three or more and four or more jets.

	$N_{\mathrm{Z+jets}}^{\mathrm{tag}}$
≥ 3 Jets	7.5 ± 0.6(stat.)
≥ 4 Jets	1.8 ± 0.3(stat.)

7.5 Background from single-top events

The single-top contribution even at the higher jet multiplicities is non-negligible as it has a high efficiency to pass the selection criteria for $t\bar{t}$ events. The dominant production modes for single-top are the t-channel and tW modes, with the s-channel production being suppressed at the LHC. The leading order production diagrams for single-top are shown in Figure 7.4. The theoretical cross-sections for single-top production at 7 TeV are as follows:

- $64.6^{+3.4}_{-3.2}$ pb for t-channel, NLO calculated with MCFM [11].
- 10.6 ± 0.8 pb for tW, NLO calculated with MCFM [11].
- $4.21^{+0.13}_{-0.10}$ pb for s-channel, calculated at NNLL [73].

The number of single-top events remaining after event selection is estimated using simulation and summarised in Table 7.7. An uncertainty of 30% on the cross-section is assumed as part of a conservative estimate as mandated by the Top Physics Analysis Group.

Table 7.7: The number of single-top events after selection estimated from simulation with statistical uncertainties for three or more and four or more jets.

	$N^{\text{tag}}_{\text{single-top}}$
≥ 3 Jets	23.1 ± 0.2(stat.)
≥ 4 Jets	7.08 ± 0.09(stat.)

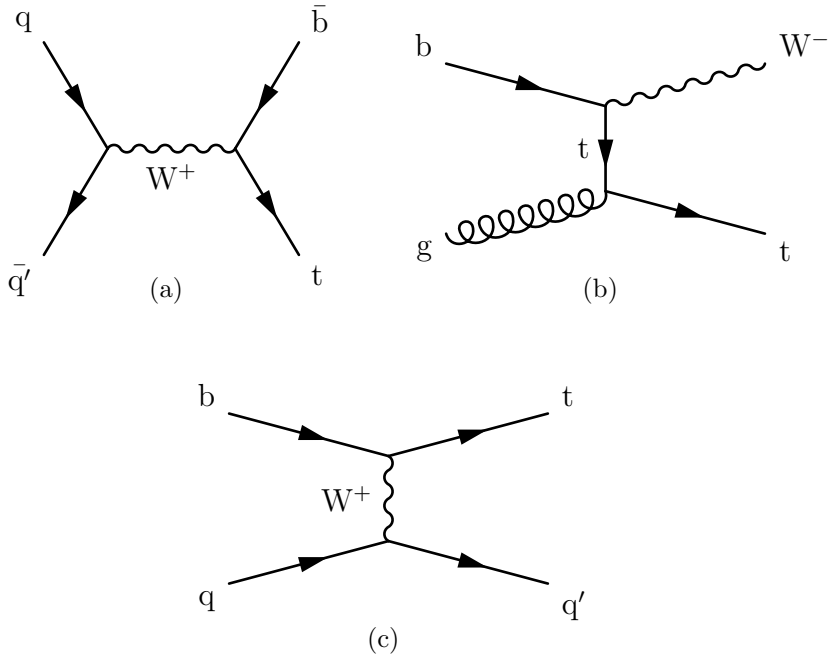


Figure 7.4: The leading order production diagrams for single-top at the LHC, in the *s-channel* (a), W-associated production, tW (b), and the *t-channel* (c).

7.6 Background from di-boson production

The production of $W^{\pm}W^{\pm}$, $W^{\pm}Z^0$ and Z^0Z^0 bosons presents a number of ways in which they could emulate signal events. Each of the bosons can decay leptonically to provide isolated leptons or hadronically to provide jets in configurations that would pass the selection criteria. The contribution from these processes, however, is still expected to be small as their production cross-sections are small in comparison to $t\bar{t}$. The inclusive production cross-sections calculated at NLO calculated with MCFM [11] are as follows:

- 43.0 ± 1.5 pb for $W^{\pm}W^{\pm}$.
- 18.2 ± 0.7 pb for $W^{\pm}Z^0$.
- 5.9 ± 0.2 pb for Z^0Z^0.

The di-boson background is estimated using simulated samples with their estimated contributions given in Table 7.8. An uncertainty of 30% on the cross-section is assumed as part of a conservative estimate as mandated by the Top Physics Analysis Group.

Table 7.8: The number of di-boson events after selection estimated from simulation with statistical uncertainties for three or more and four or more jets.

	$N^{\text{tag}}_{\text{di-boson}}$
≥ 3 Jets	1.73 ± 0.03(stat.)
≥ 4 Jets	0.40 ± 0.02(stat.)

Chapter 8

Systematic Uncertainties

8.1 Systematic uncertainties due to b-tagging

There are a number of effects that contribute to the systematic uncertainty on the cross-section due to b-tagging:

- The uncertainty on the b-jet efficiency data scale factor (SF_{b}), which is applied to both the b-jet and c-jet efficiency. This uncertainty is 15% for b-jets and 30% for c-jets.
- The statistical uncertainty on b-jet efficiency measurements in simulation ($\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$).
- The statistical uncertainty on c-jet efficiency measurements in simulation ($\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$).
- The uncertainty on the l-jet mistag rate data scale factor (SF_{l}), which varies by up to approximately 20%.
- The statistical uncertainty on l-jet mistag rates in simulation ($\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$).
- The systematic uncertainty of the method (see Section 6.4.2), which is evaluated by using $\varepsilon_{\mathrm{b,c,l}}^{\mathrm{MC}}$ without any data-driven correction factors to reconstruct the number of b-tagged events and then compared with the measured result from

counting.

The effect of each systematic uncertainty on the b-tagging selection efficiency, ϵ^{tag}, is measured by the change in ϵ^{tag}, when each parameter is varied separately by ± 1 standard deviation. The measurement technique and jet-level uncertainties are discussed in Chapter 6. The total uncertainty is quoted by combining these results in quadrature. The results for the systematic uncertainties measured in $t\bar{t}$ simulation are given in Table 8.1.

Table 8.1: Systematic uncertainties on ϵ^{tag} due to the b-tagging parameters for one or more tagged jets in the event summarised for both three or more and four or more jets measured on $t\bar{t}$ simulation.

Effect	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
	-1σ	$+1\sigma$	-1σ	$+1\sigma$
Systematic uncertainty SF_{b}	-11.0%	$+9.5\%$	-10.5%	$+8.9\%$
Statistical uncertainty $\varepsilon_{\text{b}}^{\text{MC}}$	-0.6%	$+0.6\%$	-0.6%	$+0.6\%$
Statistical uncertainty $\varepsilon_{\text{c}}^{\text{MC}}$	-0.1%	$+0.1\%$	-0.1%	$+0.1\%$
Systematic uncertainty SF_{l}	-0.1%	$+0.1\%$	-0.1%	$+0.1\%$
Statistical uncertainty $\varepsilon_{\text{l}}^{\text{MC}}$	-0.1%	$+0.1\%$	-0.1%	$+0.1\%$
Systematic uncertainty of model	-0.3%		-0.2%	
Total	-11.0%	$+9.5\%$	-10.5%	$+8.9\%$

The same procedure is also applied to the W+jets simulation to estimate the systematic uncertainty for the light and heavy flavour sub-samples separately. The re-weighting factor of $K_{\text{b}} = 2.0 \pm 1.0$ is applied to the W+b($\text{b}\bar{\text{b}}$) component of the heavy flavour sub-sample (see Section 7.2). The systematic uncertainties due to the re-weighting factors are not considered here and are discussed later in Section 8.3. This systematic treatment has also been applied to each individual flavour sub-sample of the W+jets simulation and those studies can be found in Appendix B.2. The results for the W+jets light flavour and heavy flavour sub-samples are summarised in

Table 8.2 and Table 8.3, respectively.

Table 8.2: Systematic uncertainties due to b-tagging parameters for the efficiency to pass one or more tagged jets in the event summarised for both three or more and four or more jets measured on W+jets simulation in the light flavour sub-sample.

Effect	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
	-1σ	$+1\sigma$	-1σ	$+1\sigma$
Systematic uncertainty SF_{b}	-6.8%	$+6.8\%$	-6.9%	$+6.8\%$
Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	-2.0%	$+2.0\%$	-2.4%	$+2.4\%$
Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-0.7%	$+0.7\%$	-0.8%	$+0.8\%$
Systematic uncertainty SF_{l}	-7.2%	$+7.2\%$	-7.0%	$+7.0\%$
Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-4.2%	$+4.2\%$	-4.4%	$+4.4\%$
Systematic uncertainty of model	-5.6%		-5.3%	
Total	-12.3%	$+10.9\%$	-12.3%	$+11.0\%$

Table 8.3: Systematic uncertainties due to b-tagging parameters for the efficiency to pass one or more tagged jets in the event summarised for both three or more and four or more jets measured on W+jets simulation in the heavy flavour sub-sample.

Effect	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
	-1σ	$+1\sigma$	-1σ	$+1\sigma$
Systematic uncertainty SF_{b}	-17.2%	$+16.7\%$	-15.4%	$+14.7\%$
Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	-4.9%	$+4.8\%$	-5.5%	$+5.2\%$
Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-1.6%	$+1.6\%$	-1.3%	$+1.3\%$
Systematic uncertainty SF_{l}	-0.9%	$+0.9\%$	-0.9%	$+0.9\%$
Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-0.5%	$+0.5\%$	-0.6%	$+0.6\%$
Systematic uncertainty of model	-0.1%		$+5.3\%$	
Total	-18.0%	$+17.4\%$	-16.5%	$+16.6\%$

8.2 Systematic uncertainties due to the jet energy scale and resolution

The jet energy scale is an integral part of jet reconstruction which was previously discussed in Section 4.5. The typical uncertainty on the jet energy scale varies between 8 and 2%, dependent on the transverse momentum and pseudorapidity of the jet. This uncertainty alters the acceptance, A, as a change in the jet energy scale changes the number of jets that will pass or fail the 30 GeV/c transverse momentum threshold and can consequently promote or demote events between jet multiplicity bins. The uncertainty is also largest at lower transverse momenta, so it is anticipated to be a large effect.

As well as the uncertainty parameterisation in jet p_{T} and η, corrections are included for pileup effects and also an additional uncertainty is added for b-jets to take into account the difference in detector response. The pileup correction is dependent on the jet p_{T} and corresponds to an uncertainty of 4.4% on a jet with a transverse momentum of 30 GeV/c. The b-jet correction is a flat value of 3% applicable only to b-jets (identified by generator information as discussed previously).

To estimate the uncertainty due to the jet energy scale, the energy of the jets in $\mathrm{t\bar{t}}$ simulation are varied by one standard deviation up and down and the resulting change in the acceptance from the central measured value is the uncertainty. The b-tagging selection efficiency, ϵ^{tag}, is not anticipated to be sensitive to this result as it is track based. The results are summarised in Table 8.4.

In addition to the jet energy scale, the jet energy resolution has been measured to be about 15% worse in data in comparison to jets reconstructed in simulation. To account for this, the value of $A \cdot \epsilon^{\mathrm{tag}}$ is remeasured whilst smearing the jet energy from matched GenJets by an additional random Gaussian factor which results in a 15% wider resolution in simulation. This results in a small loss in selection efficiency:

- $A \cdot \epsilon^{\text{tag}}$ changes by -0.2% for three or more jets.
- $A \cdot \epsilon^{\text{tag}}$ changes by -0.6% for four or more jets.

Table 8.4: Systematic uncertainty on the acceptance due to jet energy scale uncertainty for three or more and four or more jets.

	ΔA	
	-1σ	$+1\sigma$
≥ 3 jets	-4.5 %	+4.7 %
≥ 4 jets	-10.1 %	+11.4 %

8.3 Systematic uncertainties on the W+jets background estimation

8.3.1 Pre-tagged W+jets estimation

The method for calculating the W+jets background contribution is discussed in Section 7.2. The systematic uncertainty due to the fitting parameters on the Berends scale are given in Section 7.2.2.

The calculation of the Berends scale relies on the subtraction of the other simulation samples to estimate the number of pre-tagged W+jets events. In particular, this also includes the $t\bar{t}$ contribution which is non-negligible even in the lower jet multiplicities. Subtracting the $t\bar{t}$ contribution relies on an assumption of the cross-section value which introduces a bias in the measurement. To correct for this effect the number of pre-tagged W+jets events is remeasured whilst adjusting the $t\bar{t}$ cross-section by $\pm 30\%$, which is the same as the uncertainty on other theoretical cross-sections used in the measurement. The difference in the number of pre-tagged W+jets events is shown in Table 8.5.

Table 8.5: The resultant change in the number of pre-tagged W+jets events from adjusting the $t\bar{t}$ cross-section in simulation by ±30% for three or more and four or more jets.

	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
	-1σ	$+1\sigma$	-1σ	$+1\sigma$
$\sigma_{t\bar{t}}$	+9.0%	−7.5%	+12.7%	-10.3%

Under the assumption that the dependence on the $t\bar{t}$ cross-section is approximately linear, the gradient calculated from the results in Table 8.5 can be used to iteratively correct the number of pre-tagged W+jets events. The iterative correction is performed using equation 8.1, where $\sigma_{t\bar{t}}^{\text{meas}}$ is the measured cross-section, $\sigma_{t\bar{t}}^{\text{sim,low}}$ is the simulation cross-section reduced by 30% and $N_{\text{W+jets}}^{\text{pre-tag,low}}$ is the number of pre-tagged W+jets measured with the reduced cross-section. The parameter $D_{\text{W+jets}}^{t\bar{t}}$ is the dependence of the number of pre-tagged W+jets events on the $t\bar{t}$ cross-section (the gradient above). The cross-section is then measured repeatedly using $N_{\text{W+jets}}^{\text{pre-tag}\prime}$ and fed back into equation 8.1. The three or more jets cross-section measurement converged at three decimal places after five iterations with a resulting shift in the cross-section of −0.6 pb. The four or more jets cross-section measurement also converged at three decimal places after five iterations with a resulting shift in the cross-section of −1.8 pb.

$$N_{\text{W+jets}}^{\text{pre-tag}\prime} = N_{\text{W+jets}}^{\text{pre-tag,low}} + D_{\text{W+jets}}^{t\bar{t}} \cdot (\sigma_{t\bar{t}}^{\text{meas}} - \sigma_{t\bar{t}}^{\text{sim,low}}) \tag{8.1}$$

Similarly, the QCD subtraction also has an associated systematic uncertainty which is calculated in the same manner as in Section 8.4. The effect on the W+jets pre-tagged contribution is summarised in Table 8.6.

Table 8.6: Systematic uncertainties due to the QCD content on the estimated number of pre-tagged W+jets events for three or more and four or more jets.

	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
	-1σ	$+1\sigma$	-1σ	$+1\sigma$
QCD	+7.91%	−7.47%	+9.47%	−8.79%

8.3.2 Tagged W+jets estimation

The tagged W+jets simulation sample has two scale factors applied to re-weight the W+b($b\bar{b}$) and W+c($c\bar{c}$) contributions of $K_b = 2.0 \pm 1.0$ and $K_c = 1.0^{+1.0}_{-0.5}$, respectively. The systematic uncertainties on these scale factors are evaluated by recalculating the number of tagged events for each light and heavy flavour sub-sample using equation 7.3 with one parameter varied by ± 1 standard deviation. The results are summarised for K_b in Table 8.7 and for K_c in Table 8.8.

Table 8.7: Systematic uncertainties due to the flavour content re-weighting factor K_{b} on W+light and W+heavy flavour sub-samples.

	$k_{\mathrm{b}} - 1\sigma$	$k_{\mathrm{b}} + 1\sigma$
W+light $\geq$ 3 Jets	+5.4%	-4.9 %
W+light $\geq$ 4 Jets	+7.2%	-6.3%
W+heavy $\geq$ 3 Jets	-27.4%	+24.7%
W+heavy $\geq$ 4 Jets	-29.7%	+25.9%

Table 8.8: Systematic uncertainties due to the flavour content re-weighting factor K_{c} on W+light and W+heavy flavour sub-samples.

	$k_{\mathrm{c}} - 1\sigma$	$k_{\mathrm{c}} + 1\sigma$
W+light $\geq$ 3 Jets	+12.0%	-17.7%
W+light $\geq$ 4 Jets	+12.5%	-18.2%
W+heavy $\geq$ 3 Jets	-9.1%	+13.4%
W+heavy $\geq$ 4 Jets	-5.0%	+7.2%

8.4 Systematic uncertainties on the QCD background estimation

The systematic uncertainties on the QCD background are estimated by testing the robustness of the fit to the RelIso distribution. This is done by repeating the measurement whilst varying the following parameters in turn:

- The fit range in the RelIso distribution is varied to include the ranges listed in Table 8.10.
- The Fermi-Dirac function and a second order polynomial are used to fit the RelIso function instead of a Gaussian function.
- The number of $t\bar{t}$ events remaining in simulation in the fit region of 0.2 to 1.0 are subtracted from the data before fitting.

The fits for the two different functions and the fits for the varied fitting range are shown in Appendix C. For the purposes of this measurement, the Fermi-Dirac function is defined in equation 8.2, where parameters A, B and C are varied in the fit.

$$F(x) = \frac{A}{e^{x-B} + 1} + C \tag{8.2}$$

The largest difference in each set of results from the central value is then taken to be the systematic uncertainty for that study. The results of each study are summarised in Table 8.9 for the different functional forms, Table 8.10 for the fit range and Table 8.11 for the $t\bar{t}$ subtraction.

Table 8.9: The results of systematic studies on the dependence of the number of tagged QCD events on the fitting function in the RelIso distribution for three or more and four or more jets.

	Function	$N_{\mathrm{QCD}}^{\mathrm{tag}}$±fit	χ^2/d.o.f	$\Delta N_{\mathrm{QCD}}^{\mathrm{tag}}$
≥ 3 jets	Fermi-Dirac	63.1±7.90	5.08/5	+7.1
≥ 3 jets	Quadratic	68.0±15.5	4.83/5	+12.0
≥ 4 jets	Fermi-Dirac	11.5±1.85	3.34/5	+7.21
≥ 4 jets	Quadratic	9.36±6.45	3.97/5	+5.07

Table 8.10: The results of systematic studies on the dependence of the number of tagged QCD events on the range of the fit in the RelIso distribution for three or more and four or more jets.

	Range	$N_{\mathrm{QCD}}^{\mathrm{tag}}$±fit	χ^2/d.o.f	$\Delta N_{\mathrm{QCD}}^{\mathrm{tag}}$
≥ 3 jets	0.2-1.1	52.7±15.2	6.22/6	−3.3
≥ 3 jets	0.2-1.2	53.0±14.3	6.23/7	−3.0
≥ 3 jets	0.2-1.3	58.3±15.0	7.85/8	+2.3
≥ 3 jets	0.3-1.0	35.7±21.3	5.17/4	−20.3
≥ 4 jets	0.2-1.1	4.32±3.54	3.68/6	+0.03
≥ 4 jets	0.2-1.2	4.88±3.81	3.99/7	+0.59
≥ 4 jets	0.2-1.3	5.60±4.44	6.58/8	+1.31
≥ 4 jets	0.3-1.0	0.94±1.36	1.29/4	−3.35

Table 8.11: The results of systematic studies on the dependence of the number of tagged QCD events on the $t\bar{t}$ content in the RelIso distribution for three or more and four or more jets.

	$N_{\mathrm{QCD}}^{\mathrm{tag}}$±fit	χ^2/d.o.f	$\Delta N_{\mathrm{QCD}}^{\mathrm{tag}}$
≥ 3 jets	50.9±16.3	5.93/5	−5.1
≥ 4 jets	2.76±2.76	3.83/5	−1.53

The systematic studies give a final result for the number of tagged QCD events of:

- $N_{QCD}^{\text{tag}} = 56.0 \pm 17.5(\text{fit})^{+0}_{-5.1}(\text{t}\bar{\text{t}})^{+2.3}_{-20.3}(\text{range})^{+12.0}_{-0}(\text{function})$ for three or more jets.
- $N_{QCD}^{\text{tag}} = 4.29 \pm 3.87(\text{fit})^{+0}_{-1.53}(\text{t}\bar{\text{t}})^{+1.31}_{-3.35}(\text{range})^{+7.21}_{-0}(\text{function})$ for four or more jets.

A minimum systematic uncertainty of 50% is applied to the three or more jets case as mandated by the Top Physics Analysis Group. For the four or more jets case, the estimate above is larger than 50% and this is used instead.

8.5 Systematic uncertainties due to the simulation generator

There are a number of dedicated simulation samples which are used to evaluate the systematic uncertainties due to the parameters and configuration of the generator used in simulation. These systematic uncertainties directly affect the acceptance A and b-tagging event selection efficiency, ϵ^{tag}. The following dedicated systematic simulation samples are used:

- A different PYTHIA tune for the underlying event, known as Z2 [74].
- Two samples in which the initial and final state radiation (ISR/FSR) are varied up and down by 50%.
- Two samples in which the parton level matching threshold is varied up and down by a factor of 2.0 and 0.5.
- Two samples in which the factorisation and renormalisation scale (q^2) are varied up and down by a factor of 2.0 and 0.5.
- A sample which includes pile-up vertices consistent with the pileup distribution in the 2010 data-set.

The acceptance and b-tagging selection efficiency are remeasured with each of these simulation samples and the difference with respect to the reference $t\bar{t}$ simulation is taken to be the systematic uncertainty due to this parameter. The systematic uncertainty due to these parameters are summarised for the acceptance in Table 8.12 and for the b-tagging selection efficiency in Table 8.13.

Table 8.12: Summary of the systematic uncertainties on the acceptance evaluated with systematic samples for three or more and four or more jets.

Sample	$N_{jets} \geq 3$	$N_{jets} \geq 4$
Z2 PYTHIA Tune	+ 1.0%	+ 2.7%
Increased ISR/FSR	− 1.5%	− 1.9%
Smaller ISR/FSR	+ 0.1%	− 0.3%
Matching up	− 0.3%	− 1.8%
Matching down	− 0.8%	+ 0.8%
Scale up	− 1.0%	− 5.4%
Scale down	+ 2.7%	+ 8.4%
Pileup	+ 1.6%	+ 4.9%

Table 8.13: Summary of the systematic uncertainties on the b-tagging selection efficiency evaluated with systematic samples for three or more and four or more jets.

Sample	$N_{jets} \geq 3$	$N_{jets} \geq 4$
Z2 PYTHIA Tune	+ 0.2%	+ 0.5%
Increased ISR/FSR	− 1.1%	− 1.1%
Smaller ISR/FSR	− 0.2%	+ 0.0%
Matching up	− 0.2%	+ 0.0%
Matching down	− 0.2%	− 0.3%
Scale up	+ 0.3%	+ 0.9%
Scale down	+ 0.0%	+ 0.2%
Pileup	− 1.9%	− 2.0%

8.6 Systematic uncertainties due to the parton distribution functions

The uncertainty due to the choice of parton distribution functions (PDF) affects the acceptance, A, and b-tagging selection efficiency, ϵ^{tag}. To evaluate the effect of this systematic a dedicated tool named LHAPDF is used [75]. This tool provides the interface to the PDFs from which the probability to find partons with certain characteristics can be retrieved. The error PDFs provided by the CTEQ6.6 [46] PDF set are used to calculate the systematic uncertainty. The CTEQ6.6 PDF set is what is used to generate the simulated samples. For comparison, a second set from the MSTW2008 [76] PDF set is included here but not propagated through to the final systematic uncertainty calculation as decided by the Top Physics Analysis Group. The CTEQ6.6 PDF set contains 44 error PDFs for 22 different parameters. Similarly, the MSTW2008 PDF set contains 40 error PDFs for 20 different parameters. Each of the error PDFs in each set correspond to the resulting PDF with one parameter varied up or down by one standard deviation.

Using these error PDFs an event weight is derived from the probability to find the two underlying partons of flavour f, with momentum fraction x and at the factorisation scale Q, for the central value and the error PDF. The event weight (w_j) is calculated by equation 8.3, where PDF^0 is the central value and PDF^j is the error PDF for parameter j.

$$w_j = \frac{\text{PDF}^j(x_1, f_1, Q) \cdot \text{PDF}^j(x_2, f_2, Q)}{\text{PDF}^0(x_1, f_1, Q) \cdot \text{PDF}^0(x_2, f_2, Q)} \tag{8.3}$$

The acceptance and the b-tagging selection efficiency are remeasured for each error PDF, weighting each event by w_j. The total uncertainty is calculated according to equation 8.4, where X is the measured parameter (acceptance or b-tagging selection efficiency), X_0 is the central value and X_i^+ (X_i^-) corresponds to the result from varying

the PDF parameter up (down) one standard deviation.

$$\Delta X^{+} = \sqrt{\sum_{i=1}[\max(X_i^{+} - X_0, X_i^{-} - X_0, 0)]^2} \tag{8.4}$$

$$\Delta X^{-} = \sqrt{\sum_{i=1}[\max(X_0 - X_i^{+}, X_0 - X_i^{-}, 0)]^2} \tag{8.5}$$

The results for both CTEQ6.6 and MSTW2008 for three or more and four or more jets are summarised in Table 8.14.

Table 8.14: Systematic uncertainties due to PDFs evaluated with CTEQ6.6 and MSTW2008 for the acceptance and event-tagging efficiency.

	A		ϵ^{tag}	
	-1σ	$+1\sigma$	-1σ	$+1\sigma$
≥ 3 jets (CTEQ6.6)	-0.1%	$+0.1\%$	-0.1%	$+0.1\%$
≥ 4 jets (CTEQ6.6)	-0.1%	$+0.1\%$	-0.2%	$+0.2\%$
≥ 3 jets (MSTW2008)	-0.1%	$+0.1\%$	-0.1%	$+0.1\%$
≥ 4 jets (MSTW2008)	-0.3%	$+0.3\%$	-0.4%	$+0.5\%$

8.7 Systematic uncertainties due to the electron energy scale

The electron energy scale is applied as part of the electron reconstruction to correct the detector response to the real energy of the electron candidates. This energy scale has an associated uncertainty of 0.5% in the barrel region and 2.5% in the endcap region. To evaluate this systematic uncertainty, the electron energy is varied up and down by one standard deviation with the uncertainty for the corresponding region. This results in a small change to the acceptance and event-tagging efficiency of $A \cdot \epsilon \pm 0.6\%$, independent of the jet multiplicity.

8.8 Systematic uncertainties due to the lepton branching ratio in MadGraph

The lepton branching fraction in the MadGraph simulation samples for the process $W \to l\nu$ is fixed to the leading order value of $\frac{1}{9}$. The world average of this measurement is 0.1080 ± 0.0009 [3]. To correct for this, the event fractions have to be re-weighted. The effect of the re-weighting on the $t\bar{t}$ simulation is of the order of 0.2%. To account for this effect in all of the MadGraph samples, a 1% systematic uncertainty on the total cross-section is assumed.

8.9 Systematic uncertainties on simulation-based background estimates

For a conservative estimate of the uncertainty on the cross-section for background estimations from simulation, $\pm 30\%$ is used as mandated by the Top Physics Analysis Group. This is to account for the theoretical uncertainty, as well as the uncertainty on the acceptance and b-tagging event selection efficiency. This applies separately to the following background estimations:

- Z+jets.
- Single-top.
- Di-boson production.

8.10 Systematic uncertainty on the luminosity

The systematic uncertainty on the integrated luminosity is 4% (see Section 5.4). This applies directly to the calculation of the cross-section as well as to the normalisation

of the background estimates from simulation which is a small but non-negligible effect. The luminosity uncertainty is quoted separately.

8.11 Summary of systematic effects

The resulting change in the cross-section for three or more and four or more jets is given in Table 8.15 and in Table 8.16, respectively for the dominant systematic uncertainties. The total systematic uncertainty is determined from adding all of the systematic uncertainties in quadrature. The uncertainty due to the luminosity is summarised in Table 8.17. A complete list of all systematic effects can be found in Appendix D.

Table 8.15: Summary of the shift in the measured $t\bar{t}$ cross-section due to the dominant systematic effects for three or more jets.

Effect	≥ 3 jets	
	-1σ (pb)	$+1\sigma$ (pb)
SF_{b}	-20.8	+28.4
QCD	-18.2	+18.2
W+jets K_{b}	-7.7	+8.5
Jet energy scale	-7.6	+8.0
Total systematic uncertainty	-31.6	+37.4

Table 8.16: Summary of the shift in the measured $t\bar{t}$ cross-section due to the dominant systematic effects for four or more jets.

Effect	≥ 4 jets	
	-1σ (pb)	$+1\sigma$ (pb)
SF_{b}	-18.6	+26.3
Jet energy scale	-20.2	+22.1
q^2 $t\bar{t}$	-15.6	+9.4
Total Systematic Uncertainty	-34.9	+38.1

Table 8.17: The shift in the measured $t\bar{t}$ cross-section due to the uncertainty on the integrated luminosity for three or more and four or more jets.

	-1σ (pb)	$+1\sigma$ (pb)
Luminosity, ≥ 3 jets	-7.3	+7.9
Luminosity, ≥ 4 jets	-8.0	+8.6

Chapter 9

Measurement of the $t\bar{t}$ Production Cross-Section

9.1 Summary of results

The measurements described in the preceding Chapters are combined to calculate the cross-section using equation 5.1. Summaries of the inputs for the cross-section calculation are given in Table 9.1 for three or more jets and in Table 9.2 for four or more jets.

Table 9.1: A summary of the cross-section inputs for three or more jets.

Input	Value	Reference
N_{obs}	424	Chapter 5, Table 5.5
Backgrounds		
$N^{\text{tag}}_{\text{W+light}}$	23.1	Chapter 7, Section 7.2
$N^{\text{tag}}_{\text{W+heavy}}$	52.5	Chapter 7, Section 7.2
$N^{\text{tag}}_{\text{QCD}}$	56.0	Chapter 7, Section 7.3
$N^{\text{tag}}_{\text{Z+jets}}$	7.5	Chapter 7, Section 7.4
$N^{\text{tag}}_{\text{Single-top}}$	23.1	Chapter 7, Section 7.5
$N^{\text{tag}}_{\text{Di-Boson}}$	1.73	Chapter 7, Section 7.6
Acceptance, A		
Simulation Acceptance	0.06329	Chapter 5, Table 5.3
Trigger correction	0.982	Chapter 5, Section 5.2.1
Electron ID correction	0.950	Chapter 5, Section 5.2.3
ϵ^{tag}	0.727	Chapter 6, Table 6.1
Luminosity, $\mathcal{L}$ (pb^{-1})	35.8	Chapter 5, Section 5.4

Table 9.2: A summary of the cross-section inputs for four or more jets.

Input	Value	Reference
N_{obs}	194	Chapter 5, Table 5.5
Backgrounds		
$N^{\text{tag}}_{\text{W+light}}$	5.22	Chapter 7, Section 7.2
$N^{\text{tag}}_{\text{W+heavy}}$	12.78	Chapter 7, Section 7.2
$N^{\text{tag}}_{\text{QCD}}$	4.29	Chapter 7, Section 7.3
$N^{\text{tag}}_{\text{Z+jets}}$	1.8	Chapter 7, Section 7.4
$N^{\text{tag}}_{\text{Single-top}}$	7.08	Chapter 7, Section 7.5
$N^{\text{tag}}_{\text{Di-Boson}}$	0.40	Chapter 7, Section 7.6
Acceptance, A		
Simulation Acceptance	0.03269	Chapter 5, Table 5.3
Trigger correction	0.982	Chapter 5, Section 5.2.1
Electron ID correction	0.950	Chapter 5, Section 5.2.3
ϵ^{tag}	0.755	Chapter 6, Table 6.1
Luminosity, $\mathcal{L}$ (pb^{-1})	35.8	Chapter 5, Section 5.4

Using the inputs above the final cross-section measurements with the systematic uncertainties from Chapter 8 are as follows:

- $169 \pm 13(\text{stat.})^{+37}_{-32}(\text{sys.})^{+8}_{-7}(\text{lumi.})$ pb, for three or more jets.
- $197 \pm 17(\text{stat.})^{+38}_{-35}(\text{sys.})^{+9}_{-8}(\text{lumi.})$ pb, for four or more jets.

9.2 Comparison of cross-section measurements

A naive analysis of the two results may conclude that they are consistent within systematic uncertainties. However, the two samples are not statistically independent and the systematic uncertainties will also be correlated and have differing contributions to each cross-section measurement. A method to resolve these issues would be to establish a statistically independent measurement using the exclusive three jet bin. Unfortunately, although estimates for the background events from the current results and the b-tagging efficiencies can be established with minimal extra work, a complete set of systematic studies needed for a comparison is not available for the majority of the results that contribute to the cross-section measurement.

9.3 Comparison with other measurements

9.3.1 Other cross-section measurements at the LHC

A comparison of the three or more and four or more jet cross-section measurements with separate measurements at ATLAS and CMS at $35\,\text{pb}^{-1}$ and $36\,\text{pb}^{-1}$ respectively, is shown in Figure 9.1. For both the ATLAS and CMS measurements, a lepton+jets cross-section measurement (where "lepton" excludes τ decay channels) with b-tagging are shown [77, 71]. The di-lepton results also exclude the τ decay channels and are performed without b-tagging, these are detailed in Reference [78] for CMS and in

Reference [79] for ATLAS. The combination of these results are detailed for ATLAS in Reference [80] and for CMS in Reference [10]. The measurements in this thesis were not part of the CMS combined result, but were published as a cross-check to the CMS lepton+jets measurement with b-tagging in Reference [71]. The theory prediction is calculated with the HATHOR programme (see Section 1.2).

The most recent measurements of the cross-section from ATLAS and CMS are shown in Figure 9.2. This includes updates to the di-lepton measurements [81, 82], a measurement in the fully hadronic channel [83] and measurements in the di-lepton channel with tau and muon decays [84, 85]. A more recent measurement in the lepton+jets channel from ATLAS is also shown [86][1].

[1]These are the latest public results from CMS and ATLAS at the time of writing and cover a range of different integrated luminosities.

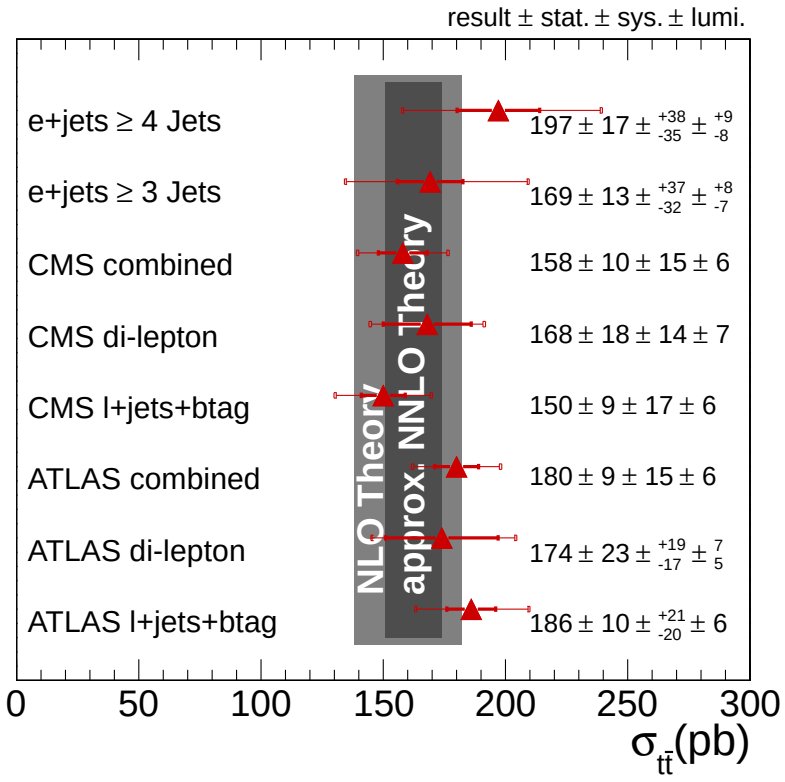


Figure 9.1: A comparison of the results from this analysis (labelled e+jets) with three separate results from CMS at $36\,\mathrm{pb}^{-1}$ and three separate results from ATLAS at $35\,\mathrm{pb}^{-1}$. The statistical uncertainty is represented by the thick error bars, the thin error bars indicate the statistical and systematic uncertainty combined in quadrature and the square brackets indicate the total uncertainty on the measurement. The NLO and approximate NNLO theory bands in light and dark grey respectively were calculated with HATHOR (see Section 1.2).

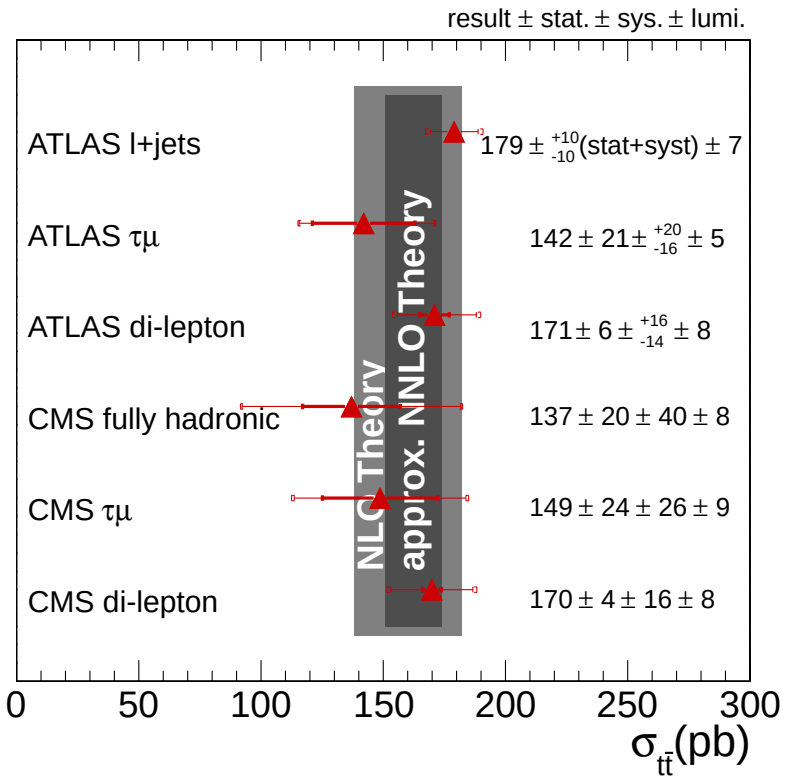


Figure 9.2: A comparison of the latest $t\bar{t}$ cross-section measurements from ATLAS and CMS at luminosities ranging between 0.70 and 1.19 fb^{-1}. The NLO and approximate NNLO theory bands in light and dark grey respectively were calculated with HATHOR (see Section 1.2).

9.3.2 Top production at the LHC in comparison to the Tevatron

A comparison of the two combined measurements from ATLAS and CMS to the results from the two Tevatron experiments, CDF and DØ has been performed in Reference [10]. The results are illustrated in Figure 9.3.

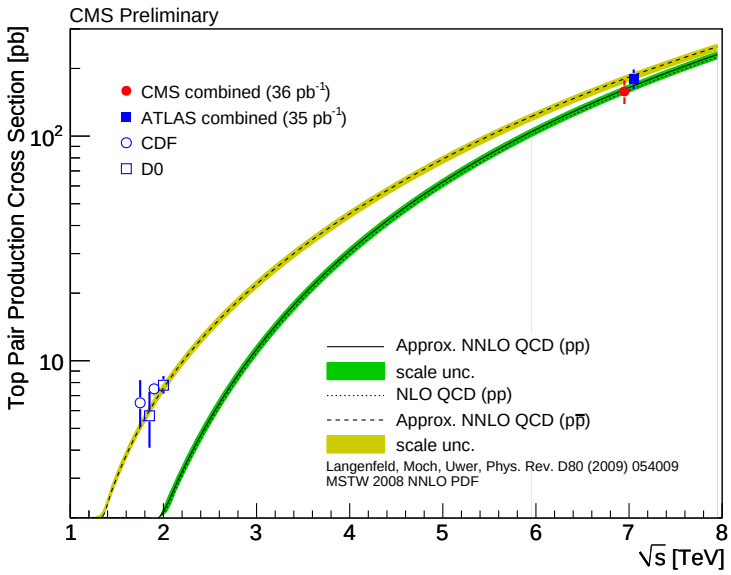


Figure 9.3: A comparison of the results from the combined cross-section measurements at ATLAS and CMS to those of CDF and DØ at the centre of mass energies $\sqrt{s} = 1.8\,\text{TeV}$ and $\sqrt{s} = 1.96\,\text{TeV}$. Data-points are shown slightly displaced horizontally for clarity.

9.4 Conclusion and future outlook

The top-quark pair production cross-section has been measured with the requirement of three or more and four or more jets in the electron+jets channel. The results are consistent with the theoretical predictions at NLO and approximate NNLO.

The original concept of this analysis was to take a conservative approach, as at the time the conditions present during data taking and the accuracy of the simulation were unknown. However, in comparison to other measurements performed at the LHC, the counting experiment method in which the backgrounds are characterised separately, offers relatively poor performance in terms of the systematic uncertainties. In particular, the large uncertainties on the QCD and W+jets contributions for the three or more jets measurement, contribute significantly to a systematic uncertainty double that of comparable lepton+jets measurements at CMS in the same jet bin using different methods. The other measurements are also able to take advantage of techniques to reduce their sensitivity to the larger sources of uncertainty such as the b-tagging data scale factors and the jet energy scale. Despite this, the measurement does still represent a very useful cross-check of the more sophisticated measurement techniques, although it is unlikely to be repeated with larger data-sets for the reasons given above.

Bibliography

[1] CDF Collaboration, "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631. doi:10.1103/PhysRevLett.74.2626.

[2] DØ Collaboration, "Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} =$ 1.8 TeV", *Phys. Rev. Lett.* **74** (Mar, 1995) 2422–2426. doi:10.1103/PhysRevLett.74.2422.

[3] K. Nakamura and others (Particle Data Group), "The Review of Particle Physics", *J. Phys. G 37, 075021* (2010).

[4] CMS Collaboration, "Search for Resonances in Semi-leptonic Top-pair Decays Close to Production Threshold", *CMS PAS TOP-10-007* (2010).

[5] CMS Collaboration, "Measurement of the Charge Asymmetry in Top Quark Pair Production", *CMS PAS TOP-11-014* (2011).

[6] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. doi:10.1103/PhysRevLett.13.508.

[7] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration et al., "Search for the Standard Model Higgs boson at LEP", *Phys. Lett. B* **565** (July, 2003) 61–75, arXiv:arXiv:hep-ex/0306033. doi:10.1016/S0370-2693(03)00614-2.

[8] TEVNPH (Tevatron New Phenomina and Higgs Working Group), CDF and DØ Collaboration, "Combined CDF and DØ Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data", arXiv:1107.5518. Sumitted to the EPS 2011 Conference.

[9] N. Kidonakis and B. D. Pecjak, "Top-quark production and QCD", arXiv:1108.6063.

[10] CMS Collaboration, "Combination of top pair production cross sections in pp collisions at 7 TeV and comparisons with theory", *CMS PAS TOP-11-001* (2011).

[11] http://mcfm.fnal.gov/.

[12] R. Kleiss and W. J. Stirling, "Top quark production at hadron colliders: Some useful formulae", *Z. Phys.* **C40** (1988) 419–423. doi:10.1007/BF01548856.

[13] http://wwwhep.ucl.ac.uk/pdf4lhc/.

[14] N. Kidonakis, "Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution", *Phys.Rev.* **D82** (2010) 114030, arXiv:1009.4935. doi:10.1103/PhysRevD.82.114030.

[15] M. Aliev, H. Lacker, U. Langenfeld et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput.Phys.Commun.* **182** (2011) 1034–1046, arXiv:1007.1327. doi:10.1016/j.cpc.2010.12.040.

[16] U. Langenfeld, S. Moch, and P. Uwer, "Measuring the running top-quark mass", *Phys. Rev.* **D80** (2009) 054009, arXiv:0906.5273. doi:10.1103/PhysRevD.80.054009.

[17] CDF and DØ Collaboration, "Combination of CDF and DØ Measurements of the Single Top Production Cross Section", arXiv:0908.2171.

[18] DØ Collaboration, "Observation of Single Top-Quark Production", *Phys. Rev. Lett.* **103** (2009) 092001, arXiv:0903.0850. doi:10.1103/PhysRevLett.103.092001.

[19] CDF Collaboration, "First Observation of Electroweak Single Top Quark Production", *Phys. Rev. Lett.* **103** (2009) 092002, arXiv:0903.0885. doi:10.1103/PhysRevLett.103.092002.

[20] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008), no. 08, S08001.

[21] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[22] CMS Collaboration, "The CMS tracker system project : Technical Design Report". CERN, Geneva, 1997.

[23] ROSE Collaboration, "2nd RD48 Status Report : R & D on silicon for future experiments", *CERN LHCC-98-039* (1998).

[24] CMS Collaboration, "CMS technical design report, volume I: Detector Performance and Software". CERN/LHCC, Geneva, 2006.

[25] CMS Collaboration, "Electromagnetic calorimeter calibration with 7 TeV data", *CMS PAS EGM-10-003* (2010).

[26] P. Adzic, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *JINST* **2** (2007), no. 04, P04004.

[27] CMS Collaboration, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *JINST* **5** (2010), no. 03, T03012.

[28] CMS HCAL Collaboration, "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", *Eur. Phys. J. C* **60** (2009) 359–373. 10.1140/epjc/s10052-009-0959-5.

[29] CMS HCAL Collaboration, "Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges", *CMS NOTE-2006-138* (2007).

[30] CMS Collaboration, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *JINST* **5** (March, 2010) 3012, arXiv:0911.4991. doi:10.1088/1748-0221/5/03/T03012.

[31] CMS Collaboration, "Fine synchronization of the CMS muon drift-tube local trigger using cosmic rays", *JINST* **5** (2010), no. 03, T03004.

[32] CMS Collaboration, "Performance of CMS muon reconstruction in cosmic-ray events", *JINST* **5** (2010), no. 03, T03022.

[33] CMS Collaboration, "Aligning the CMS muon chambers with the muon alignment system during an extended cosmic ray run", *JINST* **5** (2010), no. 03, T03019.

[34] CMS Collaboration, "Commissioning and performance of the CMS pixel tracker with cosmic ray muons", *JINST* **5** (2010), no. 03, T03007.

[35] CMS Collaboration, "Commissioning and performance of the CMS silicon strip tracker with cosmic ray muons", *JINST* **5** (2010), no. 03, T03008.

[36] CMS Collaboration, "Alignment of the CMS silicon tracker during commissioning with cosmic rays", *JINST* **5** (2010), no. 03, T03009.

[37] CMS Collaboration, "Measurement of Tracking Efficiency", *CMS PAS TRK-10-002* (2010).

[38] W. Adam, B. Mangano, T. Speer et al., "Track Reconstruction in the CMS tracker", Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec, 2006.

[39] CMS Collaboration, "Measurement of Momentum Scale and Resolution using Low-mass Resonances and Cosmic-Ray Muons", *CMS PAS TRK-10-004* (2010).

[40] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", *CMS PAS TRK-10-005* (2010).

[41] CMS Collaboration, "Studies of Tracker Material", *CMS PAS TRK-10-003* (2010).

[42] K. Prokofiev and T. Speer, "A kinematic and a decay chain reconstruction library",.

[43] J. Alwall, P. Demin, S. de Visscher et al., "MadGraph/MadEvent v4: the new web generation", *JHEP* **2007** (2007), no. 09, 028.

[44] T. Sjstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **2006** (2006), no. 05, 026.

[45] M. L. Mangano, M. Moretti, F. Piccinini et al., "Matching matrix elements and shower evolution for top-pair production in hadronic collisions", *JHEP* **2007** (2007), no. 01, 013.

[46] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev.* **D78:013004** (2008) arXiv:arXiv:0802.0007v3 [hep-ph]. doi:10.1103/PhysRevD.78.013004.

[47] Geant4 Collaboration, S. Agostinelli, J. Allison et al., "Geant4-a simulation toolkit", *Nuclear Instruments and Methods in Physics Research A* **506** (July, 2003) 250–303.

[48] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS PAS EGM-10-004* (2010).

[49] W. Adam, R. Frhwirth, A. Strandlie et al., "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *J. Phys. G* **31** (2005), no. 9, N9.

[50] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV", *CMS PAS MUO-10-002* (2010).

[51] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *J. Instrum.* **5** (Nov, 2009) T03022 . 47 p.

[52] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", *CMS PAS PFT-09-001* (2009).

[53] M. Cacciari, G. P. Salam, and G. Soyez, "The $anti-k_t$ jet clustering algorithm", *JHEP* **4** (April, 2008) 63, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[54] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (October, 1993) 3160–3166, arXiv:arXiv:hep-ph/9305266. doi:10.1103/PhysRevD.48.3160.

[55] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *CERN PH EP-2011-102* (2011).

[56] CMS Collaboration, "Jet Performance in pp Collisions at 7 TeV", *CMS PAS JME-10-003* (2010).

[57] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *CERN PH EP-2011-051* (2011).

[58] CMS Collaboration, "Adaptive Vertex Fitting", *CMS Note 2007/008* (2007).

[59] CMS HCAL Collaboration, "Selection of Top-Like Events in the Dilepton and Lepton-plus-Jets Channels in Early 7 TeV Data",.

[60] CMS Collaboration, "Measurement of CMS Luminosity", *CMS PAS EWK-10-004* (2010).

[61] S. van der Mee, "Calibration of the effective beam height in the ISR", *CERN-ISR-PO-68-31* (1968).

[62] CMS Collaboration, "Algorithms for b Jet identification in CMS", *CMS PAS BTV-09-001* (2009).

[63] CMS Collaboration, "Commissioning of b-jet identification with pp collisions at $\sqrt{s} = 7$ TeV", *CMS PAS BTV-10-001* (2010).

[64] CMS Collaboration, "Effect of misalignment on b tagging", *CMS PAS BTV-07-003* (2007).

[65] UA1 Collaboration, "Study of heavy flavour production in events with a muon accompanied by jet(s) at the CERN proton-antiproton collider", *Z. Phys. C* **37** (1988) 489. doi:10.1007/BF01549709.

[66] B. Clément *PhD Thesis, IPHC, Université de Strasbourg, France, FERMILAB-THESIS-2006-06* (2006).

[67] DØ Collaboration, "b-Jet Identification in the DØ Experiment", *Nucl. Instrum. Meth. A* **620** (2010) 490. doi:10.1016/j.nima.2010.03.118.

[68] CMS Collaboration, "Measuring uds mistag rate of b tag using negative tags", *CMS PAS BTV-07-002* (2007).

[69] CMS Collaboration, "b-tagging performance (efficiency and mistag rate) measurements from CMS", *CMS PAS BTV-11-001* (2011).

[70] F. A. Berends, H. Kuijf, B. Tausk et al., "On the production of a W and jets at hadron colliders", *Nucl. Phys. B* **357** (1991), no. 1, 32 – 64. doi:DOI: 10.1016/0550-3213(91)90458-A.

[71] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in pp collisions at 7 TeV in lepton + jets events using b-quark jet identification", *Phys. Rev. D* **84** (Nov, 2011) 092004. doi:10.1103/PhysRevD.84.092004.

[72] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$", *Phys. Rev.* **D74** (2006) 114017, arXiv:hep-ph/0609070. doi:10.1103/PhysRevD.74.114017.

[73] N. Kidonakis, "NNLL resummation for s-channel single top quark production", *Phys. Rev.* **D81** (2010) 054028, arXiv:1001.5034. doi:10.1103/PhysRevD.81.054028.

[74] R. Field, "Min-Bias and the Underlying Event at the LHC", *ArXiv e-prints* (February, 2012) arXiv:1202.0901.

[75] M. R. Whalley, D. Bourilkov, and R. C. Group, "The Les Houches Accord PDFs (LHAPDF) and Lhaglue", arXiv:hep-ph/0508110.

[76] A. D. Martin, W. J. Stirling, R. S. Thorne et al., "Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections", *Eur. Phys. J.* **C64** (2009) 653–680, arXiv:arXiv:0905.3531v2 [hep-ph]. doi:10.1140/epjc/s10052-009-1164-2.

[77] ATLAS Collaboration, "Measurement of the top quark-pair cross-section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV in the single-lepton channel using b-tagging", *ATLAS CONF-2011-035* (2011).

[78] CMS Collaboration, "Measurement of the top-quark pair-production cross section in the di-lepton channel at 7 TeV", *CMS PAS TOP-10-005* (2010).

[79] ATLAS Collaboration, "Measurement of the top quark pair production cross-section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV in dilepton final states", *ATLAS CONF-2011-034* (2011).

[80] ATLAS Collaboration, "A combined measurement of the top quark pair production cross-section using dilepton and single-lepton final states", *ATLAS CONF-2011-040* (2011).

[81] ATLAS Collaboration, "Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s} = 7$ TeV in dilepton final states with ATLAS", *ATLAS CONF-2011-100* (2011).

[82] CMS Collaboration, "Top pair cross section in dileptons", *CMS PAS TOP-11-005* (2011).

[83] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the fully hadronic decay channel in pp collisions at 7 TeV", *CMS PAS TOP-11-007* (2011).

[84] ATLAS Collaboration, "Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s} = 7$ TeV in μ+τ final states with ATLAS.", *ATLAS CONF-2011-119* (2011).

[85] CMS Collaboration, "First measurement of the top quark pair production cross section in the dilepton channel with tau leptons in the final state in pp collisions at $\sqrt{s} = 7$ TeV", *CMS PAS TOP-11-006* (2011).

[86] ATLAS Collaboration, "Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 7$ TeV using kinematic information of lepton+jets events", *ATLAS CONF-2011-121* (2011).

Appendix A

Trigger Details

The triggering system at CMS is described in Section 2.6 in general and for this specific analysis in Section 5.2.1. The full details of each individual trigger used are described in Table A.1.

Table A.1: The high level triggers used in the analysis and their characteristics. Where separate values for the ECAL endcaps are implemented, they are shown in parentheses.

Name	E_{T} (GeV)	Pixel Match	Track Match	H/E	$\Delta\phi$	$\Delta\eta$	$\sigma_{i\eta i\eta}$
HLT_Ele10_LW_L1R	10	LW	✗	✗	✗	✗	✗
HLT_Ele15_SW_L1R	15	SW	✗	✗	✗	✗	✗
HLT_Ele15_SW_CaloEleId_L1R	15	SW	✗	0.15	✗	✗	0.014 (0.035)
HLT_Ele17_SW_CaloEleId_L1R	17	SW	✗	0.15	✗	✗	0.014 (0.035)
HLT_Ele17_SW_TightEleId_L1R	17	SW	✓	0.1	0.08	0.01	0.012 (0.032)
HLT_Ele22_SW_TighterEleId_L1R	22	SW	✓	0.05	0.1	0.008 (0.007)	0.011 (0.031)

Appendix B

W+jets Flavour Sub-Samples

B.1 Separating W+jets into flavour sub-samples

The need to split the W+jets sample into flavour-separated sub-samples arises from particular concerns related to the application of b-tagging which is obviously more sensitive to the heavy flavour content and the uncertainty on the relative flavour contributions within the sample. Unfortunately, the exact flavour of a jet is not available within the simulated samples and as such a dedicated utility known as the *Flavour History Tool* was developed to handle these samples in a consistent manner.

To separate the sample into the flavour sub-samples the Flavour History Tool performs matching between GenJets and generator level parton information to attempt to ascertain the production process. The matching is performed between the GenJets and partons produced by either the matrix element calculation (ME) in MadGraph or originating from the parton shower (PS) in pythia. A ΔR cut between matched heavy flavour GenJets is used to further classify the events. The paths are defined below.

- Path 1: Two b-flavour GenJets matched to matrix element partons separated by $\Delta R > 0.5$ (W+bb).

- Path 2: One b-flavour GenJet matched to matrix element partons (W+b or W+bb).
- Path 3: Two c-flavour GenJets matched to matrix element partons separated by $\Delta R > 0.5$ (W+cc).
- Path 4: One c-flavour GenJet matched to matrix element partons (W+c or W+cc).
- Path 5: One b-flavour GenJet matched to parton shower partons ($\mathrm{W} + \mathrm{g} \rightarrow \mathrm{b\bar{b}}$).
- Path 6: One c-flavour GenJet matched to parton shower partons ($\mathrm{W} + \mathrm{g} \rightarrow \mathrm{c\bar{c}}$).
- Path 7: One b-flavour GenJet matched to two or more matrix element partons (W+bb),
- Path 8: One c-flavour GenJet matched to two or more matrix element partons (W+cc).
- Path 9: Two b-flavour GenJets matched to parton shower partons separated by $\Delta R > 0.5$ ($\mathrm{W} + \mathrm{g} \rightarrow \mathrm{b\bar{b}}$).
- Path 10: Two c-flavour GenJets matched to parton shower partons separated by $\Delta R > 0.5$ ($\mathrm{W} + \mathrm{g} \rightarrow \mathrm{c\bar{c}}$).
- Path 11: Veto all of the above (W+light).

It is important to note that this matching provides an approximation of the flavour production in the process but it is not perfectly accurate. In particular, the paths 7 - 10 are considered to be unreliable as they contain the apparent collinear production of partons from the matrix element to form a single jet (paths seven and eight) and the production of greatly separated partons from gluon splitting to produce two jets (paths nine and ten). This suggests a problem with the generated event itself or the matching performed by the Flavour History Tool as these situations are physically not possible. The problematic nature of this method is also illustrated in the analysis of the systematic uncertainties on the W+light sub-sample (Section 8.3.2) which show a

significant dependence on the b-jet tagging scale factor SF_{b}.

B.2 Individual flavour path systematic uncertainties

The following systematic studies of the b-tagging uncertainties (see Section 8.1) in individual flavour paths are included for completeness. The different flavour paths are defined above in Appendix B.1. Many of the flavour paths have very few events, resulting in estimates dominated by statistical uncertainties. It is important to note that these studies are not included in the cross-section measurement, but instead the overall systematic uncertainties for the light flavour and heavy flavour sub-samples are used. The results are summarised in Table B.1, Table B.2, Table B.3, and in Table B.4.

Table B.1: Systematic uncertainties due to b-tagging parameters for the efficiency to pass one or more tagged jets in the W+jets simulation sample separated by flavour history path, summarised for both three or more and four or more jets.

Flavour path	Effect	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
		-1σ	$+1\sigma$	-1σ	$+1\sigma$
Path 1	Systematic uncertainty SF_{b}	−12.2%	+11.2%	−11.8%	+10.5%
Path 1	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	−7.2%	+6.8%	−7.8%	+7.2%
Path 1	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	−0.1%	+0.1%	−0.2%	+0.2%
Path 1	Systematic uncertainty SF_{l}	−0.2%	+0.2%	−0.2%	+0.2%
Path 1	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	−0.1%	+0.1%	−0.1%	+0.1%
Path 1	Systematic uncertainty of model	−3.4%		+14.7%	
Path 1	Total	−14.6%	+13.1%	−14.1%	+19.6%
Path 2	Systematic uncertainty SF_{b}	−15.2%	+14.8%	−16.0%	+16.0%
Path 2	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	−10.3%	+10.3%	−6.9%	+6.9%
Path 2	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	−0.6%	+0.6%	−2.0%	+2.0%
Path 2	Systematic uncertainty SF_{l}	−0.6%	+0.6%	−0.7%	+0.7%
Path 2	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	−0.3%	+0.3%	−0.4%	+0.4%
Path 2	Systematic uncertainty of model	+6.2%		−44.3%	
Path 2	Total	−18.4%	+19.1%	−47.7%	+17.6%
Path 3	Systematic uncertainty SF_{b}	−26.3%	+25.4%	−25.5%	+24.5%
Path 3	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	0.0%	0.0%	0.0%	0.0%
Path 3	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	−4.4%	+4.3%	−5.0%	+5.0%
Path 3	Systematic uncertainty SF_{l}	−0.8%	+0.8%	−1.0%	+1.0%
Path 3	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	−0.5%	+0.5%	−0.6%	+0.6%
Path 3	Systematic uncertainty of model	+7.0%		+13.2%	
Path 3	Total	−26.7%	+26.3%	−26.0%	+28.3%

Table B.2: Systematic uncertainties due to b-tagging parameters for the efficiency to pass one or more tagged jets in the W+jets simulation sample separated by flavour history path, summarised for both three or more and four or more jets.

Path 4	Systematic uncertainty SF_{b}	-24.6%	$+24.6\%$	-22.5%	$+22.5\%$
Path 4	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	0.0%	0.0%	0.0%	0.0%
Path 4	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-4.2%	$+4.2\%$	-3.9%	$+3.9\%$
Path 4	Systematic uncertainty SF_{l}	-1.8%	$+1.8\%$	-2.4%	$+2.4\%$
Path 4	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-0.9%	$+0.9\%$	-1.4%	$+1.4\%$
Path 4	Systematic uncertainty of model	$+0.7\%$		$+4.0\%$	
Path 4	Total	-25.0%	$+25.0\%$	-23.0%	$+23.4\%$
Path 5	Systematic uncertainty SF_{b}	-13.8%	$+13.7\%$	-14.3%	$+14.3\%$
Path 5	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	-7.4%	$+7.4\%$	-8.3%	$+8.3\%$
Path 5	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-0.1%	$+0.1\%$	-0.3%	$+0.3\%$
Path 5	Systematic uncertainty SF_{l}	-0.9%	$+0.9\%$	-0.9%	$+0.9\%$
Path 5	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-0.5%	$+0.5\%$	-0.5%	$+0.5\%$
Path 5	Systematic uncertainty of model	-10.6%		-8.4%	
Path 5	Total	-18.9%	$+15.6\%$	-18.5%	$+16.5\%$
Path 6	Systematic uncertainty SF_{b}	-14.0%	$+14.0\%$	-13.0%	$+13.0\%$
Path 6	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	0.0%	0.0%	0.0%	0.0%
Path 6	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-1.8%	$+1.8\%$	-1.9%	$+1.9\%$
Path 6	Systematic uncertainty SF_{l}	-5.6%	$+5.6\%$	-5.8%	$+5.8\%$
Path 6	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-3.3%	$+3.3\%$	-3.6%	$+3.6\%$
Path 6	Systematic uncertainty of model	$+88.0\%$		$+51.4\%$	
Path 6	Total	-15.6%	$+89.5\%$	-14.8%	$+53.3\%$

Table B.3: Systematic uncertainties due to b-tagging parameters for the efficiency to pass one or more tagged jets in the W+jets simulation sample separated by flavour history path, summarised for both three or more and four or more jets.

Flavour path	Effect	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
		-1σ	$+1\sigma$	-1σ	$+1\sigma$
Path 7	Systematic uncertainty SF_{b}	-14.3%	$+14.1\%$	-14.3%	$+14.3\%$
Path 7	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	-12.0%	$+12.0\%$	-13.5%	$+13.5\%$
Path 7	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-0.1%	$+0.1\%$	0.0%	0.0%
Path 7	Systematic uncertainty SF_{l}	-0.2%	$+0.2\%$	-0.2%	$+0.2\%$
Path 7	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-0.1%	$+0.1\%$	-0.1%	$+0.1\%$
Path 7	Systematic uncertainty of model	-4.9%		0.0%	
Path 7	Total	-19.3%	$+18.5\%$	-19.6%	$+19.6\%$
Path 8	Systematic uncertainty SF_{b}	-26.4%	$+26.0\%$	–	–
Path 8	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	0.0%	0.0%	–	–
Path 8	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-3.8%	$+3.8\%$	–	–
Path 8	Systematic uncertainty SF_{l}	-1.0%	$+1.0\%$	–	–
Path 8	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-0.5%	$+0.5\%$	–	–
Path 8	Systematic uncertainty of model	–		–	
Path 8	Total	-26.7%	$+26.3\%$	–	–
Path 9	Systematic uncertainty SF_{b}	-13.3%	$+13.0\%$	-12.3%	$+11.7\%$
Path 9	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	-8.4%	$+8.3\%$	-7.8%	$+7.5\%$
Path 9	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	0.0%	0.0%	0.0%	0.0%
Path 9	Systematic uncertainty SF_{l}	-0.5%	$+0.5\%$	-0.5%	$+0.5\%$
Path 9	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-0.3%	$+0.3\%$	-0.3%	$+0.3\%$
Path 9	Systematic uncertainty of model	-2.4%		$+1.1\%$	
Path 9	Total	-15.9%	$+15.4\%$	-14.5%	$+14.0\%$

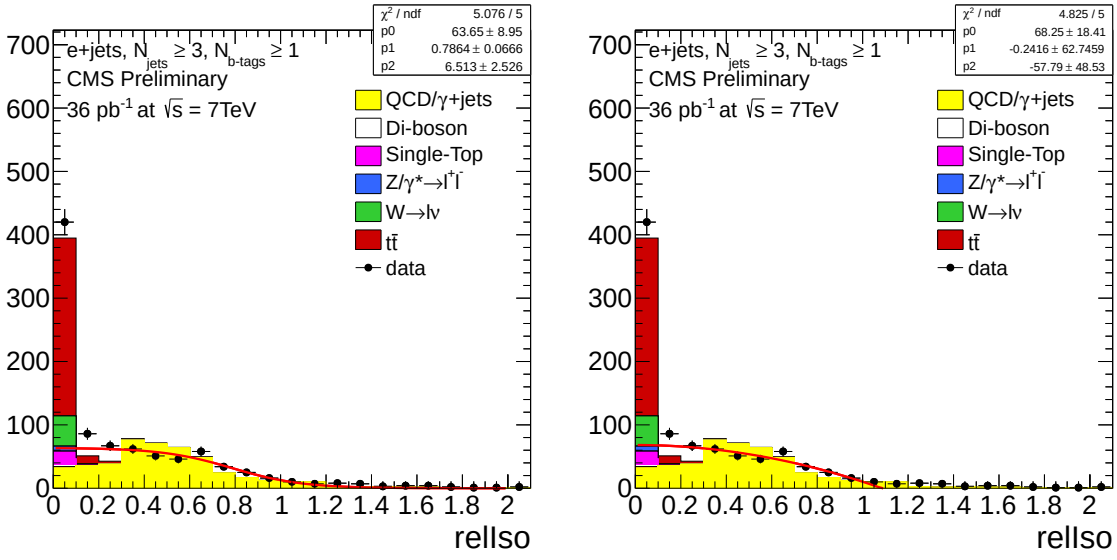
Table B.4: Systematic uncertainties due to b-tagging parameters for the efficiency to pass one or more tagged jets in the W+jets simulation sample separated by flavour history path, summarised for both three or more and four or more jets.

Flavour path	Effect	$N_{jets} \geq 3$		$N_{jets} \geq 4$	
		-1σ	$+1\sigma$	-1σ	$+1\sigma$
Path 10	Systematic uncertainty SF_{b}	-22.2%	$+22.0\%$	-22.1%	$+21.6\%$
Path 10	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	0.0%	0.0%	0.0%	0.0%
Path 10	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-3.5%	$+3.5\%$	-3.6%	$+3.6\%$
Path 10	Systematic uncertainty SF_{l}	-2.5%	$+2.5\%$	-2.4%	$+2.4\%$
Path 10	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-1.6%	$+1.6\%$	-1.6%	$+1.6\%$
Path 10	Systematic uncertainty of model	$+2.9\%$		$+6.0\%$	
Path 10	Total	-22.7%	$+22.6\%$	-22.5%	$+22.9\%$
Path 11	Systematic uncertainty SF_{b}	-6.8%	$+6.8\%$	-6.9%	$+6.8\%$
Path 11	Statistical uncertainty $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$	-2.0%	$+2.0\%$	-2.4%	$+2.4\%$
Path 11	Statistical uncertainty $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$	-0.7%	$+0.7\%$	-0.8%	$+0.8\%$
Path 11	Systematic uncertainty SF_{l}	-7.2%	$+7.2\%$	-7.0%	$+7.0\%$
Path 11	Statistical uncertainty $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$	-4.2%	$+4.2\%$	-4.4%	$+4.4\%$
Path 11	Systematic uncertainty of model	-5.6%		-5.2%	
Path 11	Total	-12.3%	$+10.9\%$	-12.3%	$+11.0\%$

Appendix C

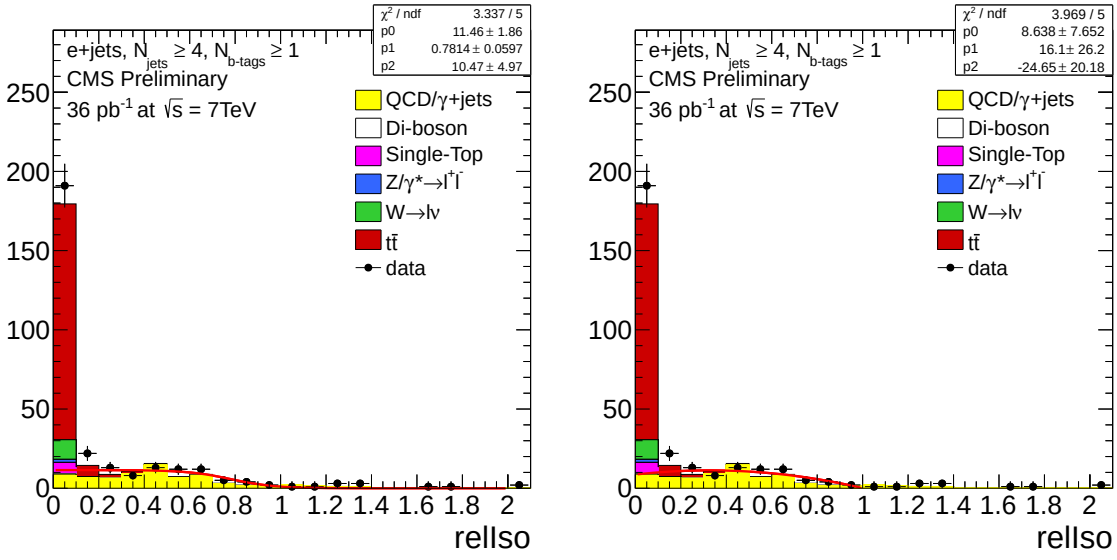
QCD Fit Systematic Uncertainties

The method for evaluating the QCD background and the uncertainty due to the fitting function and fitting range are discussed in Section 7.3 and Section 8.4. The numerical results are presented in Table 8.9 and Table 8.10. The fits to the Fermi-Dirac function and the quadratic function for three or more and four or more jets are shown in Figure C.1 and in Figure C.2 respectively. The fits for the Gaussian function with a varied range are shown in Figure C.3 for three or more jets and in Figure C.4 for four or more jets.



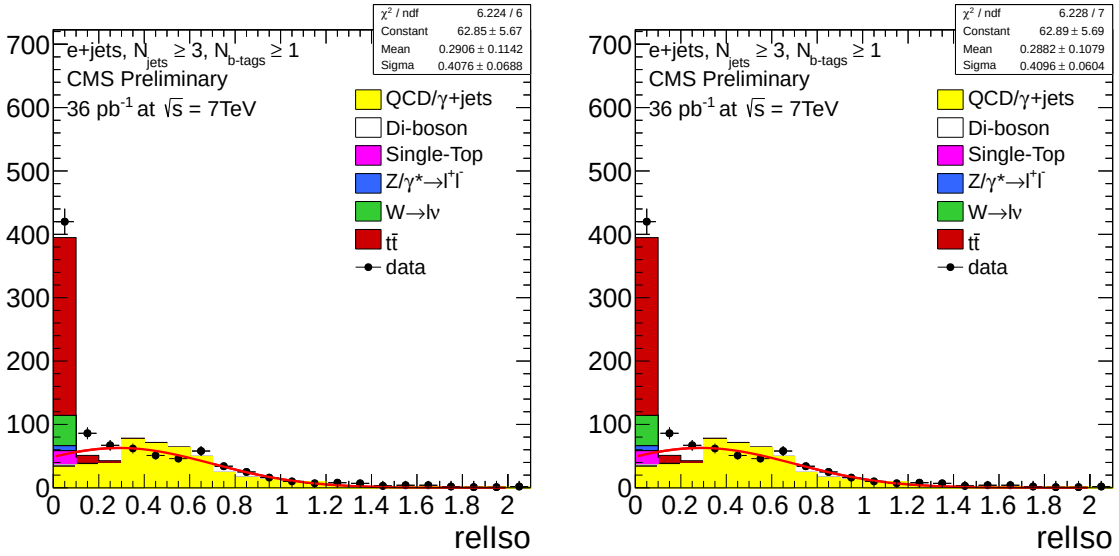
(a) (b)

Figure C.1: The fit to the RelIso distribution after selection excluding the RelIso cut in the three or more jet bin using a Fermi-Dirac function (a) and a quadratic function (b). The range of the fit is set to the default of 0.2 to 1.0.



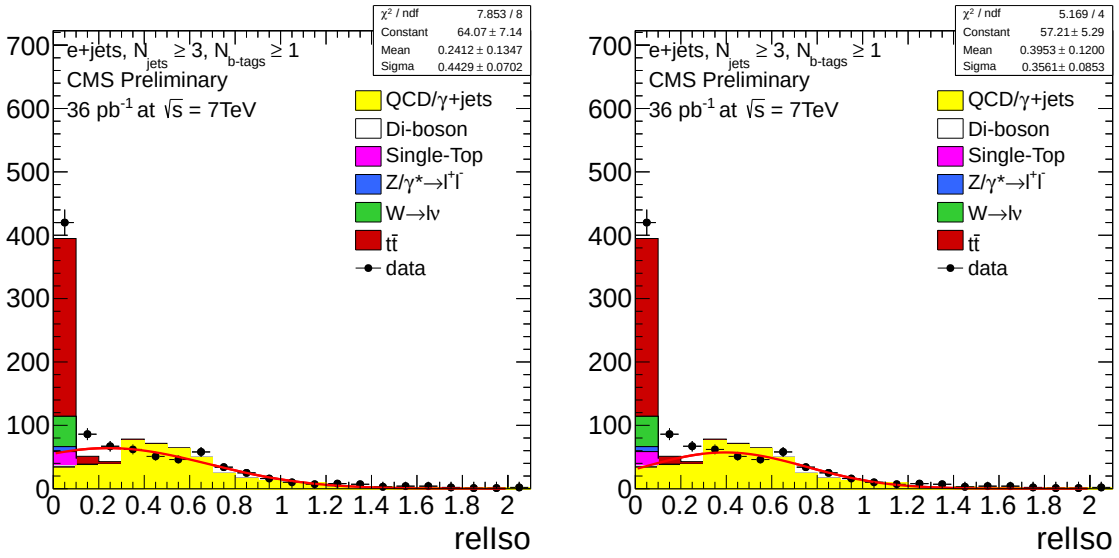
(a) (b)

Figure C.2: The fit to the RelIso distribution after selection excluding the RelIso cut in the four or more jet bin using a Fermi-Dirac function (a) and a quadratic function (b). The range of the fit is set to the default of 0.2 to 1.0.



(a)

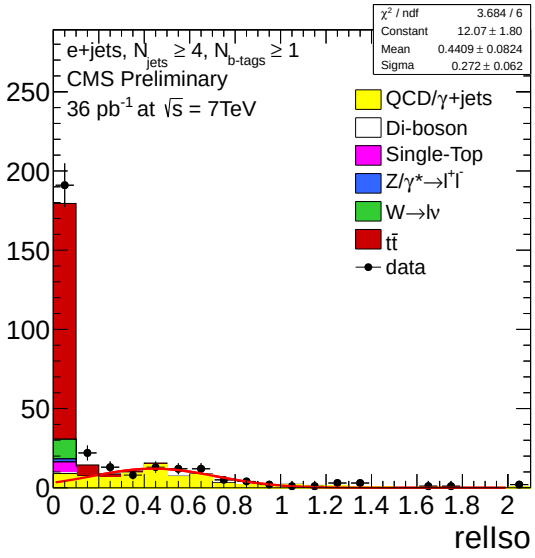
(b)



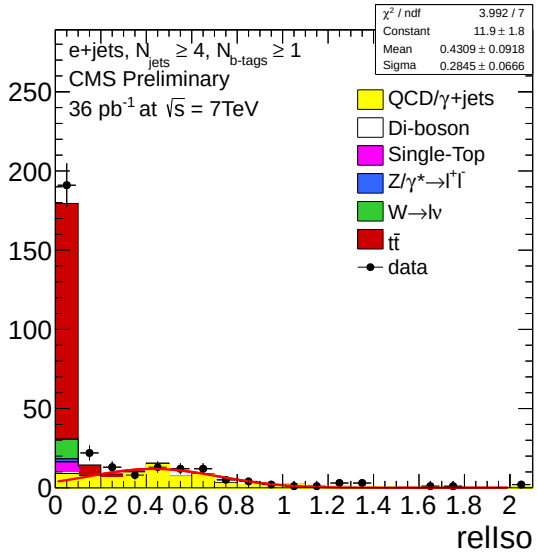
(c)

(d)

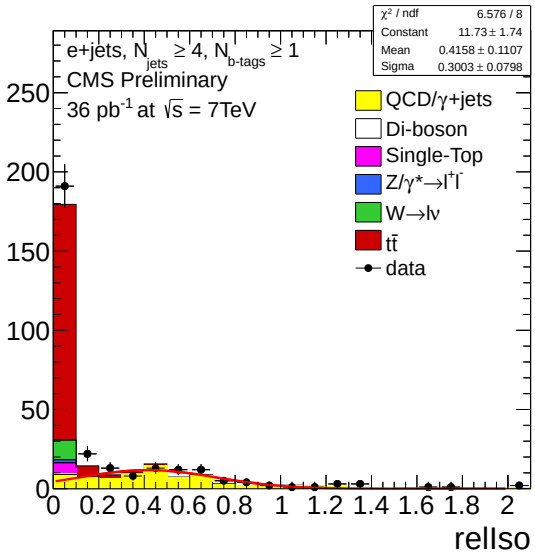
Figure C.3: The fit to the RelIso distribution after selection excluding the RelIso cut in the three or more jet bin using a Gaussian function with the ranges 0.2 to 1.1 (a), 0.2 to 1.2 (b), 0.2 to 1.3 (c), and 0.3 to 1.0 (d).



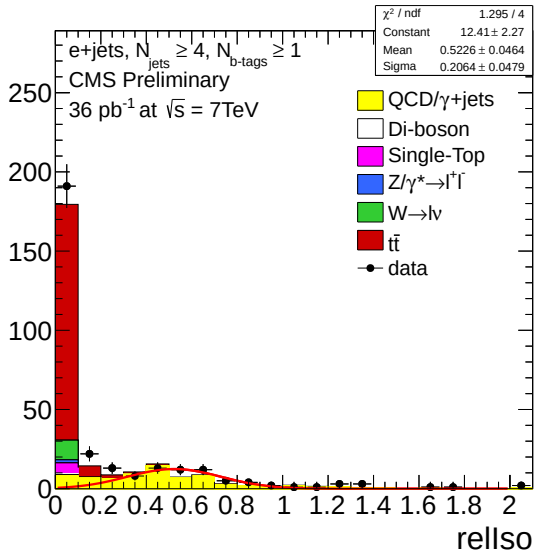
(a)



(b)



(c)



(d)

Figure C.4: The fit to the RelIso distribution after selection excluding the RelIso cut in the four or more jet bin using a Gaussian function with the ranges 0.2 to 1.1 (a), 0.2 to 1.2 (b), 0.2 to 1.3 (c), and 0.3 to 1.0 (d).

Appendix D

Complete Systematic Uncertainties

The complete list of systematic uncertainties for this analysis are given in Table D.1 and Table D.2 for three or more jets and in Table D.3 and Table D.4 for three or more jets.

Table D.1: Summary of the shift in the measured $t\bar{t}$ cross-section due to the dominant systematic effects for three or more jets.

Effect	$N_{jets} \geq 3$	
	-1σ (pb)	$+1\sigma$ (pb)
SF_{b}	-20.8	+28.4
SF_{l}	-1.7	+1.7
$\sigma_{\mathrm{Di-Boson}}$	-0.3	+0.3
$\sigma_{\mathrm{Single-top}}$	-4.5	+4.5
$\sigma_{\mathrm{Z+jets}}$	-1.5	+1.5
q^2 $t\bar{t}$	-4.4	+1.2
Berends scale fit	-1.7	+1.7
Berends scale slope	-0.3	+0.3
CTEQ6.6 Error PDF	-0.3	+0.3
Electron ID Statistical	-2.6	+2.7
Electron ID Systematic	-5.0	+5.3
Electron energy scale	-1.0	+1.0
Event tagging model		+1.4
ISR/FSR	0.0	+4.5
Jet energy resolution		+0.3
Jet energy scale	-7.6	+8.0
Parton matching	0.0	+1.7
QCD	-18.2	+18.2
Simulation with pileup		+0.6
Trigger Efficiency	-0.2	+0.2

Table D.2: Summary of the shift in the measured $t\bar{t}$ cross-section due to the dominant systematic effects for three or more jets continued.

Effect	$N_{jets} \geq 3$	
	-1σ (pb)	$+1\sigma$ (pb)
W+heavy flavour ϵ^{tag} stat.	-1.6	+1.6
W+jets K_{b}	-7.7	+8.5
W+jets K_{c}	-1.9	+1.3
W+jets $\sigma_{t\bar{t}}$ dependence	-0.6	
W+jets QCD subtraction	-3.9	+3.7
W+jets data statistics	-0.2	+0.5
W+jets simulation $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$ stat.	-1.9	+2.0
W+jets simulation $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$ stat.	-0.7	+0.7
W+jets simulation $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$ stat.	-0.8	+0.8
W+jets simulation statistics	-0.5	+0.5
W+light flavour ϵ^{tag} stat.	-0.3	+0.3
MadGraph leptonic W decay correction	-1.7	+1.7
$t\bar{t}$ ϵ^{tag} stat.	-0.9	+0.9
$t\bar{t}$ pythia tune	-2.0	
$t\bar{t}$ simulation $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$ stat.	-1.0	+1.0
$t\bar{t}$ simulation $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$ stat.	-0.2	+0.2
$t\bar{t}$ simulation $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$ stat.	-0.2	+0.2
Total systematic uncertainty	-31.6	+37.4

Table D.3: Summary of the shift in the measured $t\bar{t}$ cross-section due to the dominant systematic effects for four or more jets.

Effect	$N_{jets} \geq 4$	
	-1σ (pb)	$+1\sigma$ (pb)
SF_{b}	-18.6	+26.3
SF_{l}	-1.0	+1.0
$\sigma_{\mathrm{Di-Boson}}$	-0.1	+0.1
$\sigma_{\mathrm{Single-top}}$	-2.6	+2.6
$\sigma_{\mathrm{Z+jets}}$	-0.7	+0.7
q^2 $t\bar{t}$	-15.6	+9.4
Berends scale fit	-0.9	+0.9
Berends scale slope	-0.2	+0.2
CTEQ6.6 Error PDF	-0.6	+0.6
Electron ID Statistical	-3.1	+3.2
Electron ID Systematic	-5.8	+6.2
Electron energy scale	-1.2	+1.2
Event tagging model	-0.1	
ISR/FSR	0.0	+6.1
Jet energy resolution	+1.2	
Jet energy scale	-20.2	+22.1
Parton matching	0.0	+3.6
QCD $t\bar{t} \oplus$ range $\oplus$ function	-8.9	+4.5
QCD fit	-4.7	+4.7
Simulation with pileup	-5.4	
Trigger Efficiency	-0.2	+0.2

Table D.4: Summary of the shift in the measured $t\bar{t}$ cross-section due to the dominant systematic effects for four or more jets continued.

Effect	$N_{jets} \geq 4$	
	-1σ (pb)	$+1\sigma$ (pb)
W+heavy flavour ϵ^{tag} stat.	-0.7	+0.7
W+jets K_{b}	-3.6	+4.1
W+jets K_{c}	-0.0	+0.0
W+jets $\sigma_{t\bar{t}}$ dependence	-1.8	
W+jets QCD subtraction	-2.1	+1.9
W+jets data statistics	-0.2	+0.2
W+jets simulation $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$ stat.	-1.0	+1.0
W+jets simulation $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$ stat.	-0.3	+0.3
W+jets simulation $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$ stat.	-0.4	+0.4
W+jets simulation statistics	-0.2	+0.2
W+light flavour ϵ^{tag} stat.	-0.1	+0.1
MadGraph leptonic W decay correction	-2.0	+2.0
$t\bar{t}$ ϵ^{tag} stat.	-1.3	+1.3
$t\bar{t}$ pythia tune	-2.9	
$t\bar{t}$ simulation $\varepsilon_{\mathrm{b}}^{\mathrm{MC}}$ stat.	-1.2	+1.2
$t\bar{t}$ simulation $\varepsilon_{\mathrm{c}}^{\mathrm{MC}}$ stat.	-0.2	+0.2
$t\bar{t}$ simulation $\varepsilon_{\mathrm{l}}^{\mathrm{MC}}$ stat.	-0.2	+0.2
Total Systematic Uncertainty	-34.9	+38.1