Brunel University

Doctoral Thesis

Study of Single Top Quarks in Association with Vector Bosons

Author:
Duncan Leggat

Supervisors:
Dr. Joanne Cole
Dr. Dawn Leslie

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Department of Electronic and Computer Engineering
College of Engineering, Design and Physical Sciences

August 2015

Declaration of Authorship

The work described within this thesis was conducted solely by the author, except where collaboration with others occurred as stated within the text, during their time as a candidate for a research degree at this University. No part of this thesis has been previously submitted to this or any other university as part of the requirement for a higher degree. When the published work of others has been consulted, it has been clearly attributed within the text.

The work described in Chapter 4 is the author's contribution to the continued running of the CMS detector. Along with the work described there, they were also the official convener of the tracker DQM group from July 2014 until January 2015. The responsibilities of this role included expanding the group and supervising activities that ranged from the day-to-day maintenance of the code to creating new code for future detector upgrades.

Duncan Leggat
May 2015

"We learn as we age, we've learned nothing"

Jesse Lacey

Abstract

The search for single top production in association with a massive electroweak vector boson using data collected by the CMS detector at the Large Hadron Collider is presented. Two analyses are discussed: the search for a single top produced in association with a W boson (tW production) and the search for t-channel single top production with a radiated Z boson (tZq production). Both analyses make use of proton-proton collision data at a centre-of-mass energy of 8 TeV, and focus on the leptonic decays of the vector bosons. A cut and count based approach is employed for the tW search, searching for a final state containing two leptons, two neutrinos and a jet originating from the decay of a b-quark. The main backgrounds to be understood and controlled in this analysis are those arising from the production of top quark pairs and Z bosons with one or more jets. Using a set of data corresponding to an integrated luminosity of 12.2 fb^{-1}, a production cross section of 33.9 ± 8.6 pb was measured for the process. This corresponds to an observed (expected) signal excess of 3.6σ ($2.8^{+0.9}_{-0.8}\sigma$) over the background-only hypothesis.

The tZq search was performed on events containing three leptons, one neutrino and two or three jets, at least one of which originated from the decay of a b-quark. A multivariate analysis based on the kinematic properties of the selected events is used to separate the signal from the main backgrounds: WZ diboson production and the production of a top quark pair in association with a Z boson. Additionally, backgrounds arising from top quark pair production and the production of a Z boson with additional jets are estimated from data. Using a set of data corresponding to an integrated luminosity of 19.7 fb^{-1}, a production cross section of 783^{+1000}_{-543} fb was measured for the process. This corresponds to an observed (expected) signal excess of 3.432σ ($1.389 \pm 0.005\sigma$) over the background-only hypothesis. This constitutes the first evidence for this rare Standard Model process at the LHC.

Acknowledgements

There are many people that I need to thank for their help and continuous support in the production of this thesis. First and foremost, I would like to thank my supervisor, *Joanne Cole*, for always having time to offer me advice when I've needed it, and for her impeccable work correcting this and many other documents in my time as her student. I am consistently impressed with her patience through my disorganisation, and her mastery of email shouting when this gets out of hand. I can only hope that it's been as much fun being my supervisor as it has been being your first student. We made it!

I must also thank *Kate Mackay* for her work on the statistical analysis of the tZq analysis; without her contribution the time-scale of the analysis would have been daunting at best. I (and indeed all of the Brunel group) am also hugely indebted to *Ivan Reid*, whose constant and timely technical support made the work presented in this thesis possible. I can only apologise for the times that my poorly thought-out code have caused problems on pion.

For my final acknowledgement from Brunel University, I would be callous not to thank the other students of the HEP group. Your presence has made even the most stress-filled times fun and relaxing. I would particularly like to thank *Nik Berry*, without whom my time in Switzerland would have been filled with far less adventure.

My two years working at CERN owe their productivity and enjoyment to a great number of people. Firstly, I would like to thank *Rebeca Gonzalez Suarez* with whom I started working on the tW-channel search, and without whose advice and support I would have struggled to complete the work. During my time at CERN, she also became the convener of the single top subgroup, and in this capacity has been a fantastic ambassador for both analyses that I have worked on. I must also thank *Jeremy Andrea*, whose uncompromising thoroughness has made the tZq search the robust analysis that it is today.

My abilities outside of physics analyses owe themselves in no small way to *Dan Duggan*, who took me under his wing as a fledgling student at CERN and gave me a detailed understanding of the (sometimes very complicated) pixel DQM code. In my future position as a post-doc, I will certainly attempt to emulate his tireless willingness to help others and impeccable knowledge of his subject area.

The large community of PhD students from around the globe at CERN made my time there some of the best in my life. They are far too numerous to name, but there was always somebody who would go for a beer, play a computer game, jump on a train (or plane) to some random destination, or just sit by the lake and watch the world go by. I doubt I will ever be able to repay them for the impact they've made on my life.

I must also thank my many house-mates over the past three and a half years, who have not only put up with my general disorganisation, but have been great friends. In particular I would like to thank *Mark Smith* and *Thomas Bird*, who made our apartment in St Genis feel like a home throughout our time there, and *Kate Auty* for not only offering to view flats for me in a remote part of London, but for also giving me somewhere to live when those attempts proved fruitless. I am even more indebted to her for being a sounding board for the past few months of thesis writing, as well as nagging me almost constantly to complete my work. I doubt I would have finished so efficiently without her.

I would also like to acknowledge the constant support and love of my family, and the many times that they've sat and listened eagerly to the detailed and technical descriptions of my work. The sarcastic quips and days out based around food have often brightened my mood.

Finally, I would like to thank *Peter Hobson* and the Science and Technologies Facilities Council for believing in me and allowing this work to happen.

Contents

List of Figures

List of Tables

For my family

Chapter 1

Introduction

Science is driven by the human desire to understand the universe around us and how it works. To this end, man has devised and rigorously tested countless theories over the centuries and millennia of our existence. One of our most successful theories is that underpinning our understanding of the fundamental building blocks of the universe and their interactions, the Standard Model (SM) of particle physics. For some 50 years it has provided startlingly accurate predictions about the existence and properties of particles, often long before the means with which to test them have been devised. With the observation of a SM-like Higgs boson in 2012 [1, 2], the pantheon of particles in the SM is now complete, and it remains a remarkably self-consistent and powerful theory.

Despite the mounting evidence supporting SM predictions, there are indications that it cannot be a complete theory of matter at a fundamental level. The existence of *dark matter*, a mysterious form of non-interacting matter that, according to astronomical measurements, is over five times more plentiful in the universe than regular matter, is not explained by the SM [3]. Neutrinos are massless in the SM, but measurements have shown that they oscillate between flavour states over time, implying they must have distinct mass eigenstates [4]. There is also a conceptual 'hierarchy' problem in the SM, in that the model requires many free parameters to be arbitrarily fine-tuned, an idea abhorred by many scientists [5]. All of these problems require new physics, beyond the SM, to be introduced.

As the heaviest known fundamental particle, the top quark offers a unique insight into electroweak physics and therefore potential models beyond the SM. Not only does its high

mass give it the largest interaction coupling with the Higgs boson, but it also means that the top quark, unlike the other quarks, decays before it hadronises. This affords a unique opportunity to study the physics of a bare quark through its decay products.

The high mass also means that the top quark can only be produced at the most powerful particle colliders in the world. It has been estimated that the Large Hadron Collider (LHC) at CERN will, at its design luminosity and frontier pushing centre-of-mass energy of 14 TeV, produce a top quark pair once a second [6]. Whilst previous installations have been able to observe the top quark [7, 8], it is only with the high fluence of top quarks available at the LHC that its properties can be fully understood.

This thesis will concentrate on rarer events in which a single top quark is produced through electroweak interactions, along with one of the weak interaction mediating W and Z bosons. The first analysis presented is the search for, and cross section measurement of, the associated production of a single top quark with a W boson. This 'tW' production is a previously unobserved SM process that directly probes the Wtb coupling.

The main challenge when searching for tW production is the large background presented by top quark pair production. Not only is the cross section of pair production some ten times greater than that for tW production, but the signals are very similar and interfere at next-to-leading order (NLO). The ability to distinguish the two processes experimentally has wider-reaching implications than this analysis alone: separating SM processes that mix at NLO means that models of new physics containing similar interferences should be experimentally accessible.

The second analysis presented in this thesis is the search for a t-channel single top process with a radiated Z boson. This process is primarily of interest because of its sensitivity to the Z boson's couplings to the top quark and W boson. If new physics exists in the electroweak sector, the increase in these couplings should cause an increase in cross section that would be obvious in data already collected at the LHC.

Both analyses were carried out using proton-proton collision data with a centre-of-mass energy $\sqrt{s} = 8$ TeV collected by the Compact Muon Solenoid (CMS) experiment at the LHC. Only leptonic decays of the W and Z bosons will be considered, and only those decays involving electrons and muons. Tau leptons are included in event simulation but are not directly studied.

Chapter 2 of this thesis will briefly introduce the Standard Model of particle physics, paying particular attention to the top quark, its production and decays, and its unique properties as a particle. Chapter 3 contains a description of the experimental setup: the LHC accelerator and CMS detector. Chapter 4 outlines the process by which data from the central silicon pixel detector is certified for use in physics analyses. Having been recorded in the detector and certified as good, the data must be reconstructed and corrected for various detector effects. The reconstruction algorithms employed by CMS, along with the methods of event simulation used, are described in Chapter 5. The kinematic and topological requirements of the reconstructed objects used to select signal events are discussed in Chapter 6. Also discussed here are methods of background estimation and/or rejection employed by the two analyses, including data-driven background estimates and multivariate techniques. Chapter 7 describes the sources of systematic uncertainty present in both analyses, and the way in which they are accounted for in the result. The statistical model used to calculate the cross section and associated limits, as well as the significance of the observed result is described in Chapter 8, along with the calculated results. Also included in this chapter is a comparison of similar results from the LHC. Finally, Chapter 9 provides a summary of the work presented in this thesis, along with an outlook for the future of the measurements with the continued running of the LHC.

Chapter 2

The Standard Model and the Top Quark

The Standard Model (SM) of particle physics is a theory that describes all the known fundamental particles and their interactions via the electromagnetic, weak and strong nuclear forces within the framework of a quantum field theory. The fourth fundamental force of nature, gravity, is not included in the SM because there is, as yet, no complete quantum field theory understanding of it. In practice this has very little impact on particle physics, as the relative weakness of gravity means it does not impact interactions at the energy scales of modern colliders.

The extensive and immensely successful theoretical framework of the SM will be briefly described in the first section of this chapter. The second half of this chapter will reflect on the particle that will be studied extensively in this thesis: the top quark.

2.1 The Standard Model

The SM comprises three generations of fermions, four force-mediating gauge bosons and the Higgs boson. The Higgs boson, originating in electroweak symmetry breaking [9], is the smallest possible excitation of the Higgs field [10], interactions with which give the elementary particles their mass.

In order to distinguish between fermions and bosons, an important property of elementary particles must be introduced: spin. Spin may be considered analogous to, but separate from, classical angular momentum; it is an intrinsic property of a particle and each type has a specific value. In composite particles the spin may be considered the angular momentum around a given point, but this analogy breaks down when considering elementary particles that, despite being point-like with no internal structure, still carry a value of spin. Spin is described by the quantum number s, which may take any half-integer value, $s = 0, \frac{1}{2}, 1, \frac{3}{2}$, etc. The sign of the spin indicates whether it is in a so-called 'up' state, where the spin is in the same direction as the z-axis, or a 'down' state, where the spin points in the opposite direction to the z-axis.

The concept of spin introduces two further important properties: helicity and chirality. If the spin is in the same direction as the momentum of the particle, it is said to have right-handed helicity. Conversely, a left-handed particle has a spin in the opposite direction from its motion. For massless particles, this property is equivalent to the chirality, an intrinsic property of the particle. It differs from the helicity when the particle has mass, as the chirality of a particle is always either left- or right-handed, whereas the helicity depends on the reference frame of observation. The chirality of a particle influences how it interacts with the weak interaction.

The fermions are half-integer spin particles (i.e. spin-$1/2$) subdivided into leptons and quarks depending on their interactions with the four fundamental forces of nature. Quarks, which experience all of the fundamental forces, are the fundamental building blocks of hadronic matter: the baryons that form nuclear matter and mesons. The first, lightest generation contains the up and down quarks that form the protons and neutrons of the atomic nucleus, and therefore the observable matter in the universe. The second generation, more massive than the first, contains the strange and the charm quark, whilst the third and, to our current understanding final, generation contains the bottom and top quarks.

Leptons, which do not experience the strong nuclear force, each have an associated lepton neutrino which, because of their zero electric charge, only interact through the weak force in the SM. The three known leptons are, in order of ascending mass, the electron, muon and tau. Table 2.1 summarises the charge and masses of the three generations of fermions in the SM. Every particles in the SM has an antimatter partner, which has the same

TABLE 2.1: The mass and charge of the Standard Model fermions. All mass measurements have been taken from the Particle Data Group [5], except the top mass, which was taken from the world top mass combination [11]. The listed charge is in units of the magnitude of the electron charge.

Particle	Quarks				Leptons			
Charge	2/3			-1/3	-1		0	
Mass	u	$0.0023^{+0.0007}_{-0.0005}$	d	$0.0048^{+0.0007}_{-0.0003}$	e	0.000511	ν_e	< 2 eV
(GeV/c^2)	c	1.275 ± 0.025	s	0.095 ± 0.005	μ	0.106	ν_μ	< 0.19 MeV
	t	173.34 ± 0.76	b	4.18 ± 0.03	τ	1.777	ν_τ	< 18.2 MeV

mass but opposite charge. When a particle comes into contact with its antiparticle, they annihilate into a pair of photons.

The four gauge bosons are integer spin particles (spin-1) that mediate the three forces included in the SM. The massless photon mediates the electromagnetic force, the $W^{\pm}$ and Z^0 bosons the weak nuclear force and the gluon, of which there are eight variations, the strong force.

2.1.1 Gauge Theory of Interactions

The SM is mathematically formulated using Quantum Field Theory (QFT) [12], in which particles represent the excitations of fields permeating the universe and the dynamics of a system are described by the so-called Lagrangian formalism [13]. The Lagrangian of a system, L, is the difference between the kinetic (T) and potential (T) energy: $L = T - V$. Within QFT it is more convenient to consider the Lagrangian Field Density, $\mathcal{L}$, than L itself, although the Lagrangian may be recovered by integrating $\mathcal{L}$ over the spatial component d^3x. The general form of $\mathcal{L}$ is;

$$\mathcal{L} = \mathcal{L}(\psi, \psi_\mu) \tag{2.1}$$

where $\psi_\mu \equiv \frac{\partial \psi}{\partial x^\mu} \equiv \partial_\mu \psi$ is the four-gradient of ψ.

The quantum state of an isolated system of one or more particles may be described by a wave function, ψ. ψ can be interpreted as a probability amplitude from which physical observables may be derived. It is important to note that the wave function describes all

particles in the system, rather than each particle individually, i.e. $\psi \equiv \psi(x_1, x_2, ...)$ where x_1 and x_2 are the individual particles. By considering the swap of two particles in the wave function the differences in the behaviour of the different classes of particles may begin to be understood. Bosons are said to have 'symmetric' wave functions, such that the wave function remains the same under the swap of two bosons, whilst fermions are 'anti-symmetric', such that the swapping of two fermions results in the negation of the wave-function. This gives rise to the famous Pauli exclusion principle, as seen in equation 2.2, whist states that two fermions may not exist in the same quantum state.

$$\psi_f(x_a, x_b) = -\psi_f(x_b, x_a), \quad \therefore \quad \psi(x_a, x_a) = 0 \tag{2.2}$$

This introduces us to an important concept: observable phenomena arise from the imposition of transformations upon a system. Specifically, the symmetries that are observed in nature exist because the Lagrangian density of a system is invariant under a certain set of transformations on the underlying wave functions. For example, if ψ represents a spinor field (a vector field representing the direction of spin of a Euclidean space), an arbitrary phase α may be added;

$$\psi \rightarrow e^{-i\alpha}\psi \tag{2.3}$$

If α has no dependence on a space-time coordinates then the Lagrangian of the system remains invariant everywhere, and the transformation is said to have a global symmetry. In order to describe the nature of fundamental interactions it is necessary to consider the special case where these transformations are local, i.e. they contain a dependence on the space-time coordinate. When the Lagrangian density remains invariant under these conditions it is known as *gauge invariance*. The SM is constructed by requiring gauge invariance on $\mathcal{L}$ to reflect natural symmetries. If α is allowed to depend on space-time coordinates, equation 2.3 becomes;

$$\psi(x) \rightarrow e^{-i\alpha(x)}\psi(x) \tag{2.4}$$

Clearly $\mathcal{L}$ is not invariant under such transformations as its dependence on ∂_μ will introduce extra terms into the equation. To impose gauge invariance, a vector field, A_μ, is introduced that transforms in such a way that $\mathcal{L}$ remains constant:

$$A_\mu \rightarrow A_\mu + \frac{1}{c}\partial_\mu \alpha(x) \tag{2.5}$$

where c is a constant. By introducing the covariant derivative $D_\mu = \partial_\mu - icA_\mu$ the Lagrangian becomes invariant under the local gauge transformation. This interaction can be interpreted as the particles represented by ψ interacting with excitations in the vector field A_μ, which are the force carrying bosons. The constant, c, depends on the Lagrangian governing the force and indicates the strength of the interaction; this is called the 'coupling constant'.

Noether's Theorem [14] states that a consequence of symmetry in a dynamic system is an associated conserved physical quantity. This can be seen in several instances in classical mechanics: space-time translational symmetry leads to conservation of energy and momentum, whilst rotational symmetry leads to conversation of angular momentum. In electromagnetism, the symmetries imposed require the electric charge be conserved, as observed in nature. The strong and weak nuclear forces have analogous charges that are conserved, corresponding to the quantum numbers of their Lagrangian densities.

2.1.2 Electroweak Interaction

The *electromagnetic interaction*, described within the SM by Quantum Electrodynamics (QED), is the best understood of the four forces, and governs interactions between all charged particles. It is mediated by the photon and, because the photon has zero mass, the force has an infinite range. As the photon also carries zero charge, no self-interaction is allowed. As previously mentioned, the electric charge, Q, is the conserved quantity within QED.

The strength of the electromagnetic force is governed by the charge of an electron, e. The coupling constant, commonly referred to as the *fine structure constant*, is given by $\alpha = \frac{e^2}{4\pi}$. This value has been measured experimentally to a high degree of accuracy with a value $\alpha \sim \frac{1}{137}$ at zero energy [15].

The *weak nuclear interaction*, responsible for radioactive decay, couples to weak isospin, T, and is mediated by the exchange of the massive charged W and neutral Z bosons. Weak isospin is analogous to the electric charge in QED but, unlike the electric charge, is a property of all fermions. As a result, the weak interaction is the only fundamental force experienced by all fermions. The conserved value of the weak interaction is the projection of T along the z axis, T_3. The chirality of the fermion governs its value of T_3; a left-handed fermion has $T_3 = \pm 1/2$, whilst right-handed fermions have $T_3 = 0$. As a consequence, left-handed fermions form isospin doublets in which each up-type quark pairs with the down-type quark from the same generation, and each lepton pairs with its lepton neutrino. Conversely, each right-handed fermion forms a singlet state, although it has been observed that only left-handed neutrinos exist and there are therefore no right-handed singlets for these.

The relatively large mass of the weak gauge bosons, 80.385 ± 0.015 GeV for the W and 91.1876 ± 0.0021 GeV for the Z [5], means that they are short-lived, with a half-life of $O(10^{-25})$s. This causes the strength of the weak force to be much less than that of the electromagnetic or the strong force, and to act over a shorter range. As a consequence of the two gauge bosons, there are two types of current observed in weak interactions: the charged current carried by the W boson and the neutral current associated with the Z boson. The W boson couples only to left-handed fermions or right-handed anti-fermions via the particle doublets. This interaction is unique in the SM for two reasons: it is the only interaction that can change the flavour of quarks and the only one to violate parity, i.e. exhibit a preference for one chirality. The Z boson couples to fermion-anti-fermion pairs and, in the SM, is flavour conserving.

The weak flavour- and mass-eigenstates of the three down-type fermions do not coincide; that is to say that each of the possible flavour states, d', s' and b' is made up of a certain fraction of the mass eigenstates of the down-type quarks, d, s and b. The two vectors are related by the Cabibbo-Kobayashi-Maskawa (CKM) matrix in the relationship shown in equation 2.6.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.6}$$

The CKM matrix describes the fraction of the mass eigenstate of the down-type quarks, on the far right of equation 2.6, that is present in the flavour eigenstates of the down-type quarks (shown on the left hand side). This means that each element of the matrix is the relative coupling of, or decay fraction between, two quark flavours via the weak interaction.

The parameters of the CKM matrix have been measured at various different experiments, an overview of which can be found in [5]. A global fit of all these results produces the best estimate of each element in the CKM matrix, seen in equation 2.7. It should be noted that although the CKM matrix is expected to be unitary in the SM, this is not necessarily a requirement made when determining the elements. For example, the value of V_{tb} is consistent with 1 and therefore is still acceptable within the SM.

$$V_{CKM} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2253 \pm 0.0008 & 0.00413 \pm 0.00049 \\ 0.225 \pm 0.008 & 0.986 \pm 0.016 & 0.0411 \pm 0.0013 \\ 0.0084 \pm 0.0006 & 0.040 \pm 0.0027 & 1.021 \pm 0.032 \end{pmatrix} \tag{2.7}$$

The relative proportion of decays of the form $t \to WX$, upon which the analyses presented in this thesis will rely, are given by the square of the amplitude of the relevant CKM matrix elements, $|V_{tX}|$.

The coupling constants that govern the fundamental interactions are not actually, as their name implies, constant; they vary with the energy scale of the interaction, a process known at the 'running' of the coupling constants. Above a threshold of the order of the Z mass the electromagnetic and weak interactions, which manifest separately at low energies, can be described as one uniform *electroweak interaction*. According to electroweak theory, massless W and Z bosons should mediate the weak force, along with the massless photon, and both forces should be equally strong and far reaching. The observed weak interaction is much weaker than this proposes and the bosons do, in fact, have mass, so there must be an additional phenomenon at work in the universe.

The Higgs mechanism of electroweak symmetry breaking explains this by proposing an additional scalar field for the universe: the Higgs field. Non-zero mass means that the vacuum expectation value corresponding to the minimum of the Higgs potential cannot be located at 0, which means that it takes on the form of the colloquially named 'mexican hat', with an infinite degeneracy in the ground state (see Figure 2.1). In the high temperatures of the early universe the symmetry in the ground state remained, and the electroweak

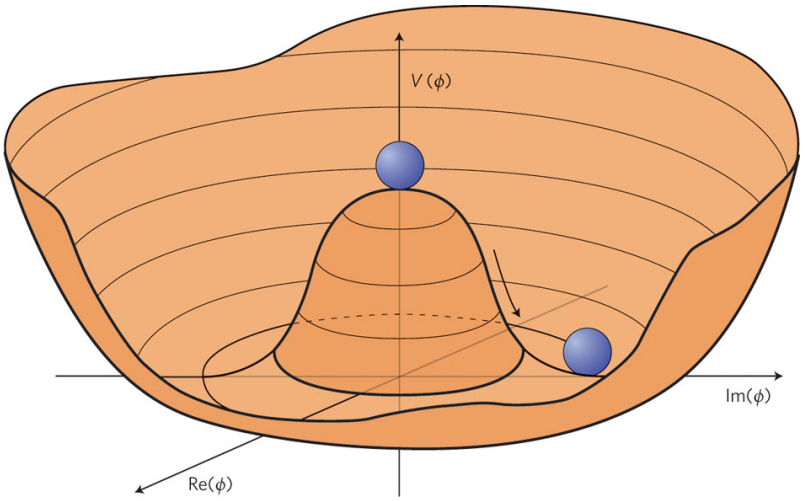


FIGURE 2.1: The effective Higgs potential chosen such that the minimum is not at the zero value of the vacuum expectation [16].

gauge bosons were massless. As the universe cooled this symmetry was spontaneously broken when a single ground state was chosen. Interactions between this field and the W and Z bosons result in their masses. The scalar Higgs boson - the final particle in the SM - exists as a by-product of the Higgs field, and was first observed at CERN in 2012 [1, 2]. The relative coupling of any given particle with the Higgs field is what gives that particle its mass.

2.1.3 The Strong Nuclear Force

The strong force is described within the SM by Quantum Chromodynamics (QCD). It is mediated by 8 massless gauge bosons called gluons and acts on the conserved charge of the strong force, known as colour. Unlike the singular electric charge, there are three types of colour charge: red, green and blue. The name 'colour' is meant for illustrative purposes only, and does not bear a direct relation to visual colour. Unlike the photon in the electromagnetic force, the gluons themselves carry colour charge. This means that self-interaction between the gluons is possible; this is the defining characteristic of the strong force. Each quark carries an associated colour charge, and each anti-quark carries an anti-colour. Gluons carry a colour and an anti-colour charge, whilst the other particles do not carry any colour charge. The strong force therefore only acts on quarks and gluons.

Unlike the electroweak force, the strong coupling constant, α_s, increases with particle separation. The strong force is therefore characterised by a potential that grows with distance. This effect, known as *asymptotic freedom*, is a direct consequence of the gluon's

self-interaction. In QED, any physical separation between two electric charges may be considered to be filled with virtual electron-positron pairs. These extra charges create an effective screening between the two 'bare' charges, reducing the effective charge seen by either. As the two charges approach each other the cloud of virtual pairs becomes smaller, and therefore the effective charge increases. Unlike the photon, the gluon can self-interact, meaning that the analogous virtual quark-anti-quark cloud between separated colour charges also contains virtual gluons. These gluons have the opposite affect to the screening quarks, and increase α_s, greatly increasing the strong potential between two quarks at large distances. Conversely, at very small distances the strong potential between the quarks becomes negligible, and they effectively act as free particles.

A consequence of this asymptotic freedom is known as *colour confinement*. The increase in potential with separation means that no free quarks can exist; they must instead remain in bound 'colourless' states. A colourless state comes about when there are either an equal number of colour charged quarks as anti-colour charged anti-quarks, or through the combination of a red, green and blue charge (analogously to red, green and blue light combining to form white, or 'colourless' light). Two such states are currently known to exist in nature: the three-quark hadrons, such as the proton and neutron, and the quark-anti-quark mesons, such as pions. If enough energy is applied to a quark to bring it away from its confined state, the increasing strong potential means that at some point it becomes more energetically viable to create a new quark-anti-quark pair out of the vacuum and so create a new confined state than it is to displace the quark any further. These hadronising quarks lead to the particle jets that are measured by particle detectors.

The idea that an exact number of quarks exists within a confined state is, however, overly simplified. In reality gluon splitting and quark annihilation processes are constantly occurring, creating a 'sea' of quarks and gluons within the hadron [17]. When studied with a low momentum probe, the hadron acts as if the three quarks that define it, known as 'valence' quarks, are free particles, each carrying a fraction of the hadron's total momentum, referred to as Bjorken x. As the probe's momentum increases, it is able to resolve increasingly small momentum fractions, revealing the presence of both gluons and the sea quarks. The dependence of the number density of quarks and gluons (known collectively as 'partons') on x and how they evolve with increasing hard scale is encapsulated within the Parton Distribution Functions (PDFs). These PDFs are vital in creating accurate simulations of hadron collisions, as they provide a detailed description of the internal

structure of each hadron in the initial state. The x dependence of the PDFs is impossible to calculate within perturbative QCD, although the dependence on the hard scale can be predicted using the so-called DGLAP evolution equation [18–20]. The x-dependence is parametrised, with the parameters determined by fitting data from a wide variety of experiments. A number of the fits exist; they are generally performed by collaborations such as the Coordinated Theoretical-Experimental Project on QCD (CTEQ) [21].

A hadron collider can impart enough energy to a parton to bring it away from its confined state. As the separation increases, the intermediate gluon field forms a narrow colour tube to hold the quarks together. At some point it becomes more energetically viable for the separating force to pull a new quark-anti-quark pair from the vacuum than to stretch the colour tube any further. This creates a new confined state in a process known as *hadronisation*.

A consequence of α_s changing so rapidly with distance is that it makes creating a consistent and satisfactory simulation of QCD processes very difficult. At very high energies, such as at the point of hard scatter in the LHC, the particles involved in a QCD interaction can be considered free particles, allowing calculations to be performed using a perturbation theory. However, once the energy of scale drops below a certain threshold, α_s becomes too large and perturbative QCD no longer applies. Therefore, in order to simulate QCD interactions, a two stage approach must be used. The first step uses perturbative QCD to develop the particles produced in the hard scattering, along with the proton remnants, into a more complete description of the event. This stage is known as *parton showering* and continues for as long as the partons produced have sufficient energy for perturbative QCD to remain valid. When the energy of the partons drops below threshold and perturbation theory no longer applies. This is known as *hadronisation* and is a non-perturbative process for which different types of models have been developed, the most common being the Lund String Model [22] and the Cluster Model [23]. To ensure the two procedures work cohesively, different 'matching' algorithms have been developed, examples of which are extensively discussed in [24].

2.2 The Top Quark

The top quark was originally proposed along with the bottom quark in 1973 as the pair that would constitute the third generation of quarks within the SM [25]. The discovery of the τ lepton in 1975 [26] and the bottom quark in 1977 [27], along with the implied lepton-quark symmetry of the existing models, strongly implied that the top quark must exist, although no experiment had the ability to observe it at the time. The search for the top quark continued in the following years but, despite the discovery of the W and Z bosons [28, 29] at the ever-increasingly energetic Super Proton Synchrotron (SPS) at CERN, it remained elusive until a larger, more powerful particle collider was built. It was first observed in 1995 by the CDF [7] and D0 [8] experiments at the Tevatron accelerator at the Fermilab facility in the USA.

The exact value of the top quark mass, m_{top}, is an important fundamental parameter of the SM. Its large mass gives the top quark the largest coupling to the Higgs boson of all particles in the SM, and as such it appears to have a special place in electroweak symmetry breaking [30, 31]. Along with the mass of the W boson, the top mass is one of the most important input parameters to global electroweak fits that constrain the properties of the Higgs, including the stability of the Higgs potential [32]. The value of the top quark production cross section is also heavily influenced by additional radiative contributions that a large mass brings in the form of virtual fermionic loops.

Although it is clearly vital to accurately measure m_{top}, there is a conceptual problem surrounding the definition of quark mass. For free particles such as the leptons the mass is well defined in a classical way, but when considering a confined state there is no direct way with which to measure the mass. The mass then becomes a property of the quark that must be inferred, and may have different values depending on the scheme being used to define it. The two main interpretations of the top mass [33] are the pole-mass scheme and the $\overline{MS}$ scheme. The pole mass treats m_{top} (and indeed the other quark masses) as a physical mass term in the quark propagator similar to that of an electron. Whilst this scheme works well in perturbation theory, such as the mathematical foundation of the electroweak interaction, the non-perturbative infrared effects of QCD are not accounted for. These effects, caused by the additional loops in the production diagrams from the self-interacting gluon, lead to the propagator, and, therefore, quark mass from QCD, increasing dramatically. In order to maintain a finite and realistic value for the mass, renormalisation

schemes are required. The most common of these is the $\overline{MS}$ scheme. The mass determined from the $\overline{MS}$ scheme can be calculated from the pole mass [34], but it may be extracted directly from data and as such is the preferential definition.

The current best estimate of the top mass is $m_{top} = 173.34 \pm 0.76$ GeV/c^2 [11], making the top quark the most massive fundamental particle known to date. Its large mass gives the top quark one of its most interesting features: its lifetime is much shorter than the timescale over which hadronisation occurs. Other quarks can hadronise with different quarks, forming a wide range of mesons and hadrons, each with their own decay topologies. Top quarks, on the other hand, can only decay via the weak interaction to a bottom quark and a W boson. This affords a unique opportunity to study the properties of a bare quark through its subsequent decay products, such as its polarisation [35] and the helicity of the W bosons produced in the decay [36].

2.2.1 Top Quark Topology

Its large mass means that the top quark can only be produced at the most energetic particle colliders; the Tevatron was able to observe the top quark, but could only begin to probe its properties. The LHC's higher centre-of-mass energy and integrated luminosity result in the production of a large number of top quark events, bestowing upon it the moniker of 'top factory'.

The top quark is predominantly produced via the strong interaction alongside an anti-top quark, the leading order (LO) Feynman diagrams for which can be seen in Figure 2.2. At the Tevatron the dominant production channel for $t\bar{t}$ pairs was quark-anti-quark annihilation, with $\sim 90\%$ of $t\bar{t}$ pairs produced in this way [37]. Conversely, at the LHC $t\bar{t}$ production is dominated by gluon fusion; about 80% of $t\bar{t}$ pairs are produced this way at $\sqrt{s} = 8$ TeV. Once the accelerator restarts at $\sqrt{s} = 14$ TeV this will increase to $\sim 90\%$. The reasons for this difference in production mechanism are twofold. Firstly, the PDFs change substantially between $\sqrt{s} = 1.96$ TeV and $\sqrt{s} = 8$ TeV, with a much larger fraction of the hadron's energy carried by the gluons at higher energies. Secondly, the Tevatron was a proton-anti-proton collider as opposed to the LHC, which is a proton-proton machine. This means that the annihilating quarks at the Tevatron may both be valence quarks, whereas at the LHC at least one of the quarks in an annihilation process must originate from the sea quarks.

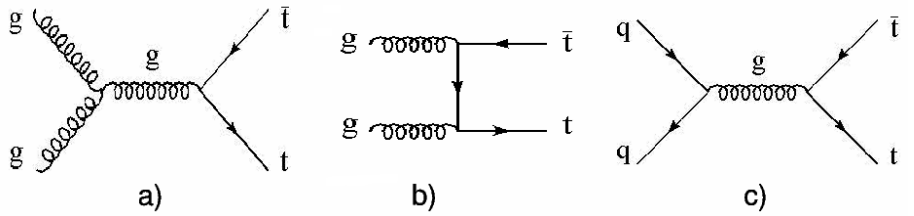


FIGURE 2.2: The three leading order processes for $t\bar{t}$ pair production at a hadron collider. Gluon fusion (a) and scattering (b) processes are the most common at the LHC, whilst quark anti-quark annihilation (c) was the dominant process seen at the Tevatron. [38]

TABLE 2.2: Standard Model cross sections for single top and $t\bar{t}$ pair production at the Tevatron and LHC. All cross sections are calculated at next-to-next-to-leading-order and measured in pb [37].

	$\sqrt{s}$ (TeV)	s-channel	t-channel	tW-channel	$t\bar{t}$
Tevatron	1.96	1.046	2.08	0.266	7.31
LHC	7	4.56	65.9	15.6	163
	8	5.55	87.2	22.2	235.8
	14	11.86	248	83.6	920

Top quarks can also, on rarer occasions, be produced alone via the weak interaction. This can occur through three different channels that are outlined in Figure 2.3. These channels have different initial and final states, and so are treated independently in analyses. Single top channels are excellent probes of the SM: they have direct access to the V_{tb} element of the CKM matrix and can be used to assess the bottom quark contribution to the PDFs. This means that the number of events expected in these channels are very sensitive to many models of physics beyond the SM. The SM cross sections for single top and $t\bar{t}$ pair production are given in table 2.2.

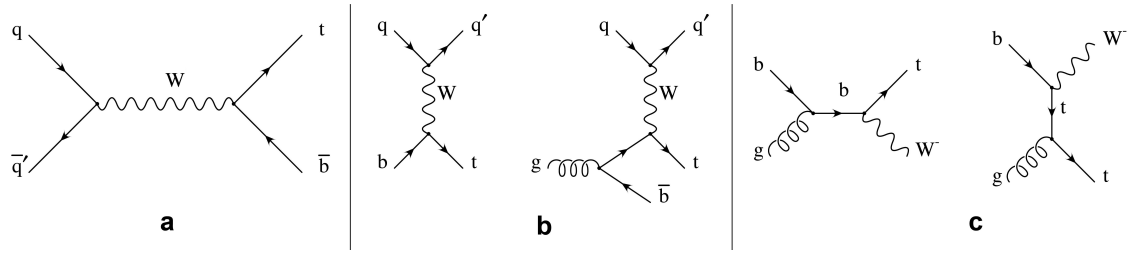


FIGURE 2.3: Leading order Feynman diagrams for the three production channels for a single top quark via the weak interaction. The three channels are (a) the s-channel, (b) the t-channel and (c) tW-channel [39].

The s-channel process, seen in Figure 2.3(a), is a quark-anti-quark annihilation to an off-shell $W^{\pm}$ boson that decays to a top and an anti-bottom quark. Similarly to quark annihilation $t\bar{t}$ production, this process was much more relevant at the Tevatron, which contained valence anti-quarks in the initial anti-proton. The process has been observed

at the Tevatron [40], but due to its slow cross section growth with energy at the LHC, especially in comparison to its main backgrounds, $t\bar{t}$ and single boson production, it is very difficult to make measurements of this channel at the LHC. As such, it remains unobserved there.

The t-channel process, shown in Figure 2.3(b), is quark scattering via the exchange of a virtual W boson. The final state is a single top quark and one other quark of any flavour. The t-channel was first observed at the Tevatron in conjunction with the s-channel [41], and since then, thanks to its relatively large production cross section, has been the most accessible of the single top channels at the LHC.

As seen in Figure 2.3(b), there are two methods of production for the t-channel: the case where the bottom quark is a sea quark and the case where a gluon splits into a $b\bar{b}$ pair before the interaction. Although the two cases essentially describe the same process (as sea quarks are the product of splitting gluons within the proton), the additional quark in the final state requires additional theoretical consideration when analysing this channel. The difference is related to how the b-PDFs are generated theoretically. If the b-quark is assumed to be massless, then it can have its own PDFs, like the other quarks (other than the top). This is known as the five flavour scheme. Conversely, if the b-quark is treated as massive, then it cannot have its own PDF, but must be generated exclusively via gluon splitting into $b\bar{b}$ pairs. This is known as the four flavour number scheme. In both cases the fraction of the proton's momentum carried by the gluons is much higher at the LHC than at the Tevatron, leading to an increased number of events in which gluon splitting occurs and accounting for the large increase in the t-channel cross section (over $30\times$ larger at $\sqrt{s} = 7$ TeV compared to $\sqrt{s} = 1.96$ TeV).

The final channel, on which the first of the two analyses presented in this thesis will concentrate, is the production of a single top quark in association with a real, on-shell W boson, known as the tW-channel. As the initial state contains a gluon and a b-quark, the lower energy of the Tevatron largely limited tW production there, to the point where it was considered negligible. At the LHC tW production is expected to be observable with a cross section that grows favourably with energy: whilst the t-channel grows by a factor of approximately 3 between 8 and 14 TeV, the tW production cross section increases by a factor of almost 4, comparable to the increase of the $t\bar{t}$ production cross section. The

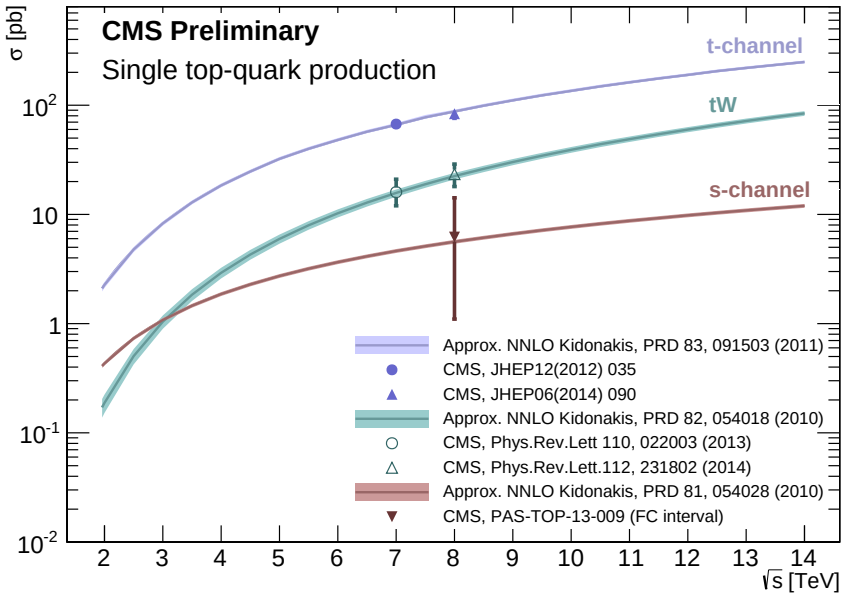


FIGURE 2.4: Theoretical and measured values of the single top production cross sections. The figure shows the evolution of the cross sections from the Tevatron centre-of-mass energy on the far left to the design centre-of-mass energy of the LHC on the far right [42].

evolution of the single top cross sections with increasing centre-of-mass energy can be seen in Figure 2.4.

The tW channel, first evidence for which has previously been seen by the CMS [43] and ATLAS experiments [44], is interesting for several reasons. Firstly, at the time the analysis was begun the channel remained undiscovered. Proving that it existed was an important test of the SM. Secondly, at next-to-leading order (NLO) it interferes with $t\bar{t}$ production, the implications and ramifications of which will be discussed in Section 2.2.2. Thirdly, as only a single top quark and a W boson are present in the final state, the channel is sensitive to physics that directly affects the Wtb interaction. Finally, the channel is a background to many searches, new and old, and a good understanding of its production mechanism aids many other analyses.

As seen from equation 2.7, the top quark decays via the weak interaction to a W boson and a b-quark almost 100% of the time. This means that the decays of a top event are categorised by the decay of the W boson, which can either be leptonic or hadronic. This corresponds to the W boson coupling to weak isospin doublets, i.e. to a lepton and lepton neutrino or quark-anti-quark pair. The leptonic decays therefore take the form $W^+ \to l^+ + \nu_l$ and $W^- \to l^- + \nu_l$, where l represents one flavour of lepton. The leptonic

branching fraction is $BR(W \to l\nu) = (10.80 \pm 0.09)\%$ for each lepton flavour [5]. The hadronic decay occurs when the $W^{\pm}$ decays to a quark-anti-quark pair, for which the branching fraction is $BR(W \to hadrons) = (67.60 \pm 0.27)\%$. The analyses presented in this thesis will consider exclusively the leptonic decay of the W boson. W and Z boson decays to tau leptons are included in event simulation but, due to complications with reconstructing the hadronic decay of the tau, are omitted from event selection in many leptonic analyses.

The first analysis discussed is the search for, and cross section measurement of, the tW channel. The top quark decays to a W boson of opposite sign to the one produced on-shell. Consequently, the two leptons produced by the leptonic W decays are also oppositely signed. The two neutrinos from the two W decays carry a large amount of transverse energy. Finally, there is also a b-quark originating from the top decay. The signature of the tW channel is, therefore, two oppositely signed leptons (in any combination of electrons and muons), a large amount of missing transverse momentum and a single jet originating from a b-quark.

2.2.2 tW production interference with $t\bar{t}$

At LO, tW and $t\bar{t}$ production are well defined and independent processes. A problem arises, however, when considering the NLO corrections in QCD, examples of which are shown in Figure 2.5. Although the corrections themselves are standard, these diagrams can also be considered top quark pair production events with one of the tops decaying to a $W^{\pm}$ boson and a b quark. These diagrams, known as 'doubly resonant' as the top quark propagator can be either on or off shell, represent an interference between top pair production and the tW channel. Conversely, the well-defined diagrams of Figure 2.3(c) are referred to as 'singly resonant'. The interference is worst when the invariant mass of the Wb system is close to m_t. At this point the diagram is most compatible with $t\bar{t}$ production, causing the top propagator to become large and artificially inflating the tW production cross section [45].

In practice, despite the similarity between the two processes it is possible to define kinematic and topological requirements that can separate them. The main difference is the presence of an extra b-quark from the second top quark of $t\bar{t}$ production, resulting in additional jets and b-tags in the event. Although there are additional b quarks present at

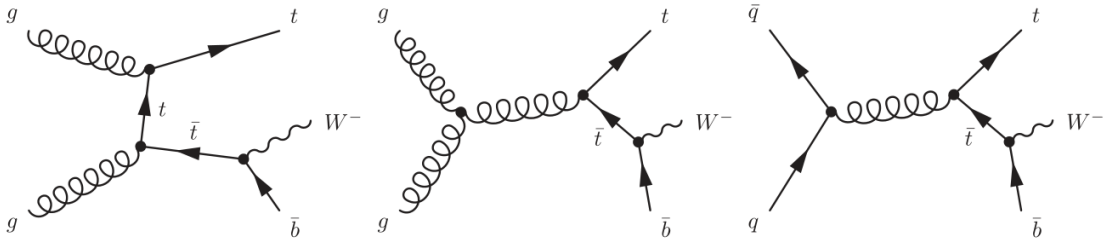


FIGURE 2.5: Feynman diagrams for the NLO tW single top quark production that are removed from the signal definition in the Diagram Removal simulation scheme. Charge conjugate diagrams are implicitly included [43].

NLO in tW, they originate from a gluon splitting event and so tend to be softer than the typical higher energy b quarks originating from $t\bar{t}$ events.

The interference problem, then, is not predominantly one of event selection, but rather of simulation. Accurately simulated events are required to study and optimise analyses in the context of the SM, and NLO corrections are vital in correctly generating these events. One possible solution would be to simulate events ignoring the interference at NLO and then remove by hand the problematic events using a series of kinematic requirements, but this approach is impractical. The production time for such a sample would need to be very large in order to maintain sufficient size and purity for analysis purposes after removing the interference events. In order to be able to generate a large, pure sample, a standardised definition of tW at higher orders must be asserted.

There are two working definitions for the production of simulated tW events which, by design, can be compared to directly estimate the contribution of tW-$t\bar{t}$ interference [45]. These schemes originated as part of the MC@NLO event generator, and have since been included in the POWHEG generator [46]. The two schemes are:

- The Diagram Subtraction (DS) scheme introduces a locally gauge invariant subtraction term into the NLO tW production cross section calculation to cancel the $t\bar{t}$ interference.
- The Diagram Removal (DR) scheme removes all the doubly resonant NLO tW diagrams from the signal definition. Whilst this method produces good results, it is not gauge invariant.

The problematic diagrams, containing the process $\alpha\beta \to tW\bar{b}$ and its charge conjugate process, where $\alpha\beta = gg$ or $q\bar{q}$, are denoted by $\mathcal{D}_{\alpha\beta}$. Within the DS scheme the contribution

from these diagrams is given by the magnitude squared of the amplitude for $t\bar{t}$ production, $|\mathcal{A}^{t\bar{t}}_{\alpha\beta}|^2$. A local, gauge-invariant counteracting term, $\bar{\mathcal{D}}_{\alpha\beta}$, is defined in such a way that it can be subtracted and will cancel the doubly resonant contributions exactly when $M_{bW} = m_{top}$. It therefore has the form:

$$\bar{\mathcal{D}}_{\alpha\beta} = \frac{BW(M_{bW})}{BW(m_{top})}|\hat{\mathcal{A}}^{t\bar{t}}_{\alpha\beta}|^2 \tag{2.8}$$

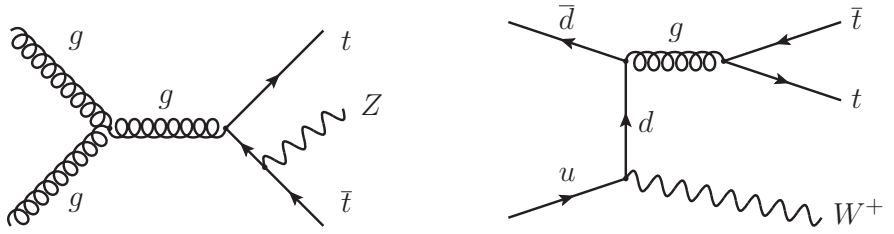
where $BW()$ is a Breit-Wigner function, used to describe the non-interfering cross section of particle resonant states [47], and $|\hat{\mathcal{A}}^{t\bar{t}}_{\alpha\beta}|^2$ is the $|\mathcal{A}^{t\bar{t}}_{\alpha\beta}|^2$ system arranged in such a way that the $\bar{t}$ is on-shell.

The DS scheme has advantages over the DR scheme in that it considers all diagrams, albeit not necessarily equally, and maintains gauge invariance in the calculation. However, the DR scheme is chosen as the default scheme for top physics analyses because DS can lead to unphysical negative weights when simulating events. The choice of DR or DS is taken into consideration in the systematic uncertainties of the tW analysis, as described further in Section 7.3.3.

2.2.3 Single top production in association with a Z boson

The second analysis presented in this thesis is the search for another rare single top interaction: t-channel single top production in association with a radiated Z boson (tZq).

The large mass of the top quark implies that it has a special place within electroweak symmetry breaking. The energy and luminosity frontiers being explored by the LHC allow the electroweak properties of the top quark to be investigated for the first time. This understanding requires measurements of the top quark coupling with the electroweak bosons, the γ, W and Z bosons. The obvious signatures to investigate are the associated production of bosons with a top anti-top quark pair, the $t\bar{t}Z$ and $t\bar{t}\gamma$ processes, because of the high cross section of top pair production at the LHC. The LO Feynman diagram for this process can be seen in Figure 2.6a. Although it might be assumed that the $t\bar{t}W$ process would also give insights into the top electroweak sector, the W boson in this case couples to the initial state quarks and has no interaction with the produced top, as seen in Figure 2.6b. These signatures are sensitive to new physics, including mixing with Z'



(A) $t\bar{t}Z$ production (B) $t\bar{t}W$ production

FIGURE 2.6: The LO Feynman diagrams for $t\bar{t}Z$ (A) and $t\bar{t}W$ (B) production [49]. $t\bar{t}\gamma$ production proceeds analogously to $t\bar{t}Z$, with the Z boson replaced with a photon.

gauge bosons and new heavy fermions [48]. CMS has performed measurements using all three channels [49, 50] and has found them to be consistent with the SM. The SM cross sections for the processes can be seen in table 2.3.

TABLE 2.3: SM cross sections for top quark associated production processes. The $t\bar{t}Z$ cross section was calculated at NLO using MADGRAPH5 and aMC@NLO [49]. The $t\bar{t}\gamma$ cross section was calculated at LO using the WHIZARD event generator and corrected to NLO using a k-factor [50]. A detailed description of the $tZ + \bar{t}Z$ calculation (which is correct to NLO) can be found in [51].

Process	Cross Section (fb)
$t\bar{t}Z$	197^{+22}_{-25}
$t\bar{t}\gamma$	1800 ± 500
$tZ + \bar{t}Z$	236 ± 24

It might be naïvely assumed that the associated production of a single top quark through the t-channel with a Z^0 boson would have a much smaller production cross section than the top quark pair alternative, because of the large difference in cross sections of the two base processes (see table 2.2), such that tZq production would essentially be inaccessible at the LHC. In fact, the additional particles within the $t\bar{t}Z$ final state make it more difficult to produce, meaning that the combined tZ and $\bar{t}Z$ (collectively referred to as tZq) cross section is larger than that for $t\bar{t}Z$ production. As can be seen in Figure 2.7, the tZq cross section scales with centre-of-mass energy similarly to $t\bar{t}Z$ production at the LHC. As evidence for the $t\bar{t}Z$ process has already been found in the data provided by CMS at $\sqrt{s} = 8$ TeV [49], it should also be possible to observe the tZq process.

The tZq channel, the LO Feynman diagrams for which can be seen in Figure 2.8, provides an excellent probe of the SM: as well as being sensitive to the coupling between the top and the Z boson, it also has sensitivity to the coupling between the W and Z boson. It forms an irreducible background to searches for flavour changing neutral currents, a possible area of

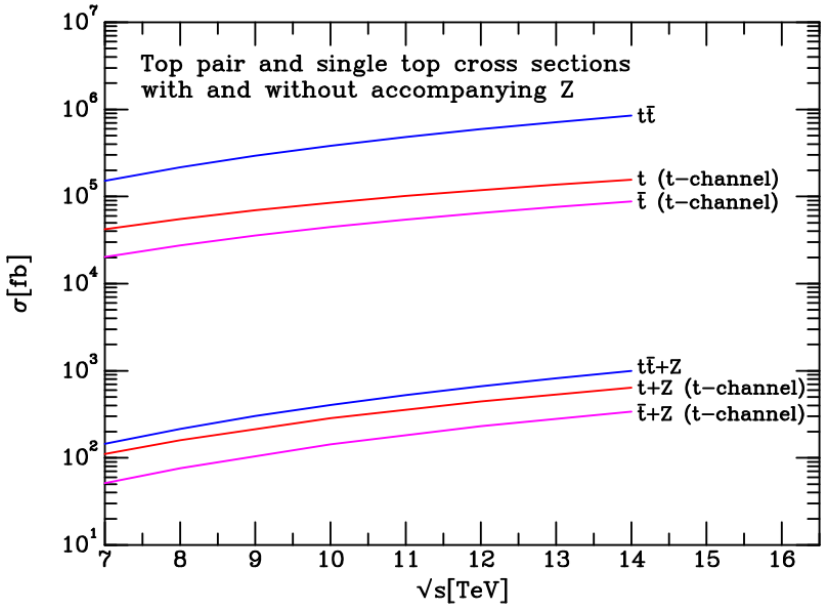


FIGURE 2.7: NLO inclusive cross sections for single top and top quark pair production with and without an associated Z boson [51].

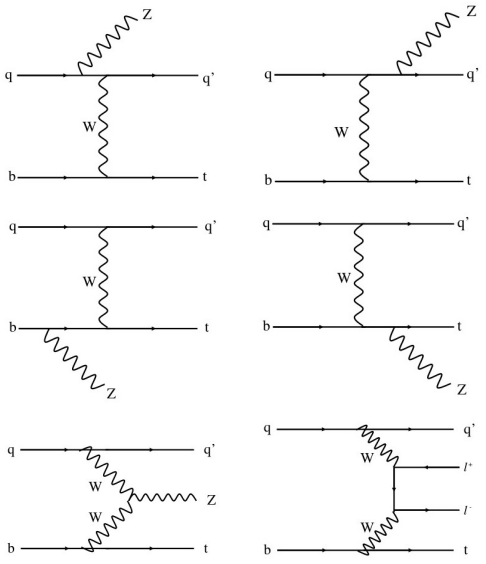


FIGURE 2.8: The LO Feynman diagrams for tZq production [52].

investigation for physics beyond the SM. It can also form a background to the production of a single top in association with a Higgs boson, which at higher centre-of-mass energies will form an important area of study in the Higgs sector.

The easiest way to observe the tZq process will be to look for the fully leptonic decay channel. The leptonic decay of the W boson resulting from the top decay and the on-shell Z boson give a tri-leptonic final state. The two leptons originating from the Z boson decay must be of the same flavour and have opposite sign, and their invariant mass will

be consistent with that of the Z. The leptonic decay of the W boson introduces a large amount of missing transverse energy in the form of a neutrino. The final part of the event topology to consider are the hadronic jets. There will be one jet originating from the b-quark of the top decay, and one from the recoil quark of the t-channel process. This additional jet may also originate from a b-quark; in this case the process is referred as tbZ production. In addition there may be an extra jet originating from the gluon splitting that creates the b-quark in the initial state; this jet can be seen in Figure 2.3(b). The final state is therefore three leptons, two of which must originate from a Z boson, a large amount of missing transverse energy and two or three jets, at least one of which originates from a b-quark. The actual selection requirements used for the analysis will be discussed in detail in Chapter 6.

Chapter 3

LHC and the CMS Detector

3.1 The LHC

The Large Hadron Collider (LHC), at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland, is the highest energy proton-proton collider ever built. Situated on top of the Franco-Swiss border in a tunnel around 100m underground and 27km in circumference, it is also the largest particle collider ever constructed. The LHC's design specifications comprise two 7 TeV proton beams each containing 2808 bunches of up to 1.15×10^{11} protons, colliding at 4 interaction points around the ring once every 25 ns [53]. This corresponds to a design peak instantaneous luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}.

The four interaction points, shown in Figure 3.1, house the four major experiments of the LHC. The two general purpose detectors, A Toroidal LHC Apparatus (ATLAS) [55] and the Compact Muon Solenoid (CMS) [56], are located at point 1 in the CERN Meyrin site and point 5 near Cessy, France, respectively. LHCb, a detector specialising in b physics, is situated in point 8 near Ferney-Volaire [57]. ALICE (A Large Ion Collider Experiment), located at point 2 in Pouilly, France, specialises in heavy ion physics [58].

The protons in the LHC originate from a hydrogen gas canister at the beginning of Linear Accelerator 2 (Linac 2) on the main CERN Meyrin site. Here they are accelerated to 50 MeV before injection into the Proton Synchrotron Booster, which in turn accelerates the protons to 1.4 GeV before injection into the Proton Synchrotron (PS). The PS accelerates the protons to 26 GeV before passing them on to the Super Proton Synchrotron (SPS), which further accelerates them to 450 GeV before final injection into the main LHC ring.

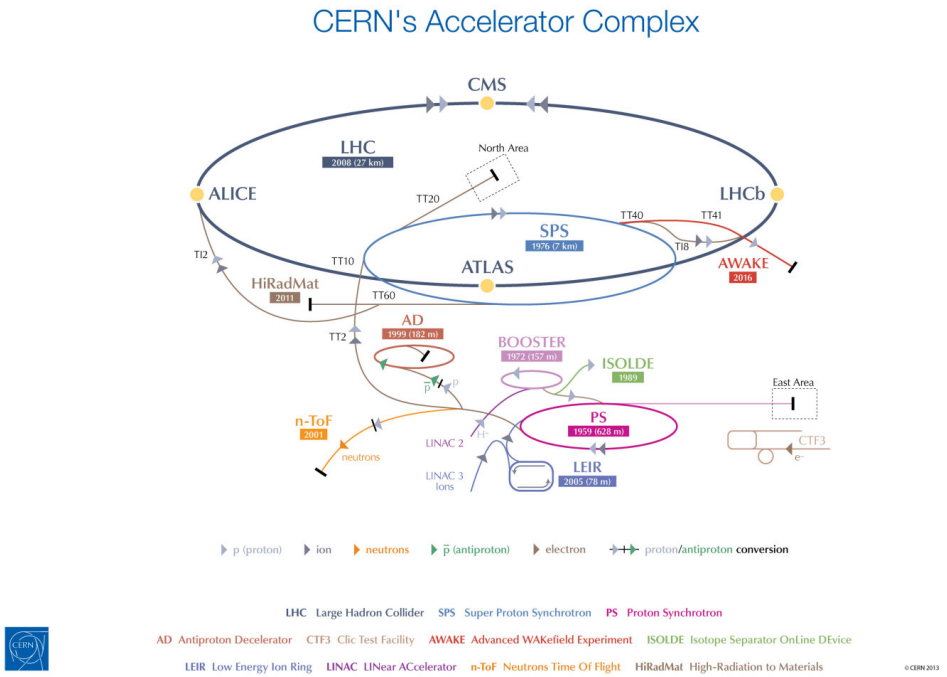


FIGURE 3.1: The LHC accelerator chain and the location of the four interaction points housing the main LHC experiments [54].

Once in the main ring, the protons are collected into bunches and accelerated to collision energy using radio frequency (RF) cavities. The LHC uses eight RF cavities per beam, operating at 4.5 K and delivering 4 MV at a frequency of 400 MHz. The process of fully accelerating each beam to the nominal operating energy of 7 TeV takes around 20 minutes. To keep the proton bunches in the LHC ring, 1232 superconducting magnetic dipoles are used to bend the beam and 858 quadrupoles are used to focus and stabilise it. These magnets operate at a temperature of 1.9 K and produce a field of 8.3 T. A more detailed account of the LHC injector sequence can be found in [59].

Although mostly centred on proton-proton physics, the LHC also operates a heavy-ion physics program. For one month a year lead ions are collided at an energy of 2.76 TeV per nucleon for the purpose of investigating the early universe through study of the quark-gluon plasma created in these collisions.

3.1.1 LHC Performance 2011-2012

In order to operate safely (and avoid incidents such as the magnet quench of September 2008 [60]), the energy and luminosity of the LHC beams have been slowly ramped up over

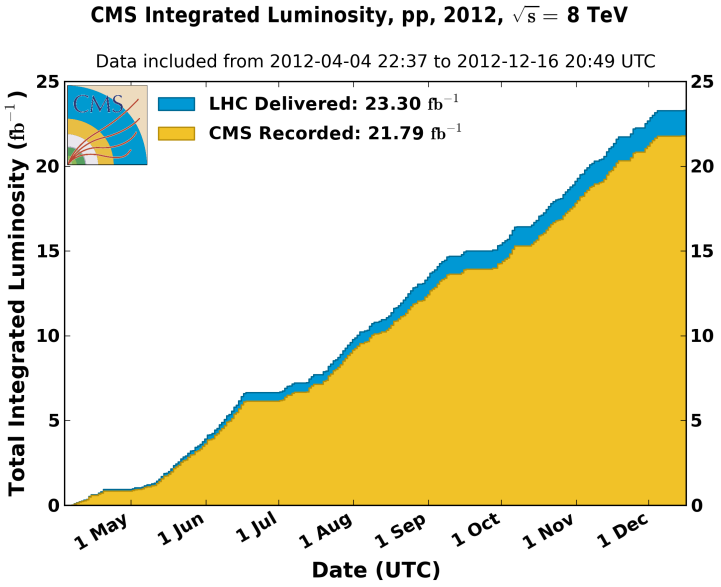


FIGURE 3.2: Delivered versus recorded luminosity for proton-proton collisions in CMS during 8 TeV running in 2012 [62].

the course of several years. During 2011, the LHC operated at a centre-of-mass energy ($\sqrt{s}$) of 7 TeV at an instantaneous luminosity of up to 3.65×10^{33} cm^{-2}s^{-1}, corresponding to 1380 proton bunches at 50ns bunch spacing. In 2012 this was increased to $\sqrt{s} = 8$ TeV, with instantaneous luminosity of up to 7.7×10^{33} cm^{-2}s^{-1} [61]. Figure 3.3 shows clearly the corresponding increase in integrated luminosity during each data-taking period.

As of the end of heavy ion running in early 2013, the LHC has been in a long shut down for the purpose of essential upgrades and repairs, known as LS1. When it restarts in 2015 it is expected to run at $\sqrt{s} = 13$ TeV before eventually reaching its design energy and luminosity in later years.

The analyses presented in this thesis are based on data collected by the CMS detector during proton-proton running in 2012.

3.2 The CMS Detector

The Compact Muon Solenoid (CMS) detector is one of the two general purpose detectors installed at the LHC. CMS is distinguished by its large, high field solenoidal magnet, high precision silicon tracker and homogeneous scintillating crystal electromagnetic calorimeter. The design of CMS was motivated by the expected physics of the LHC, namely the

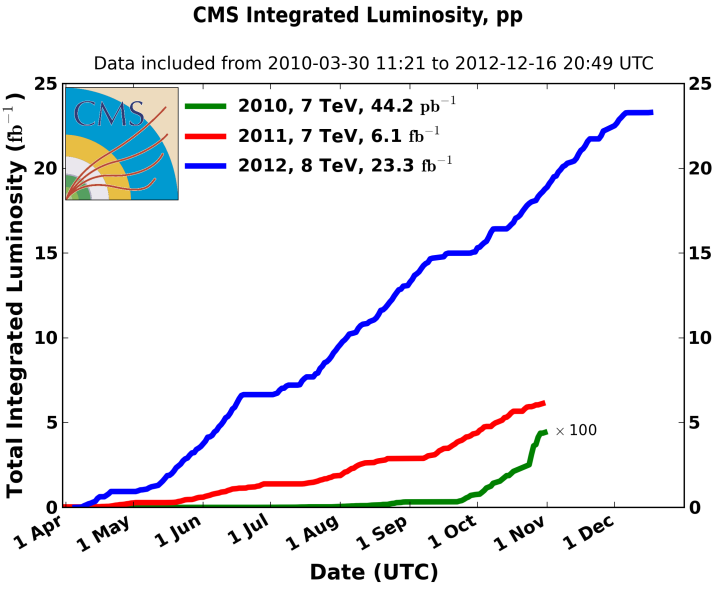


FIGURE 3.3: Integrated luminosity recorded by CMS throughout run 1, 2010 until 2012 [62].

search for the mechanism of electroweak symmetry breaking. In particular this meant a high precision inner tracker for particle identification and b tagging, an electromagnetic calorimeter with good energy and spatial resolution and an accurate muon tracking system. As accurate measurements of missing transverse energy and dijet masses are required, the detector was designed to cover a large solid angle and have good granularity throughout. The detector also had to be designed taking into account the practical implications of the high luminosity environment of the LHC. Fast electronics and triggering systems were a necessity to handle the high bunch crossing frequency, and the sub-detectors close to the beamline, in particular the inner edge of the HCAL and ECAL endcaps, had to be sufficiently radiation-hard to survive the harsh radiation environment.

The co-ordinate system of CMS has its origin at the centre of the detector at the nominal interaction point, and employs a right-handed Cartesian system. The x-axis points towards the centre of the LHC ring, whilst the y-axis points vertically upwards perpendicular to the ground, and the z-axis is parallel to proton beam 2, pointing in the anti-clockwise direction. The polar angle, θ, is measured with respect to the positive z-axis, and the azimuthal angle, ϕ, is defined as the angle in the $x - y$ plane. It is often more convenient to express the polar angle in terms of pseudorapidity (η), which is defined in equation 3.1.

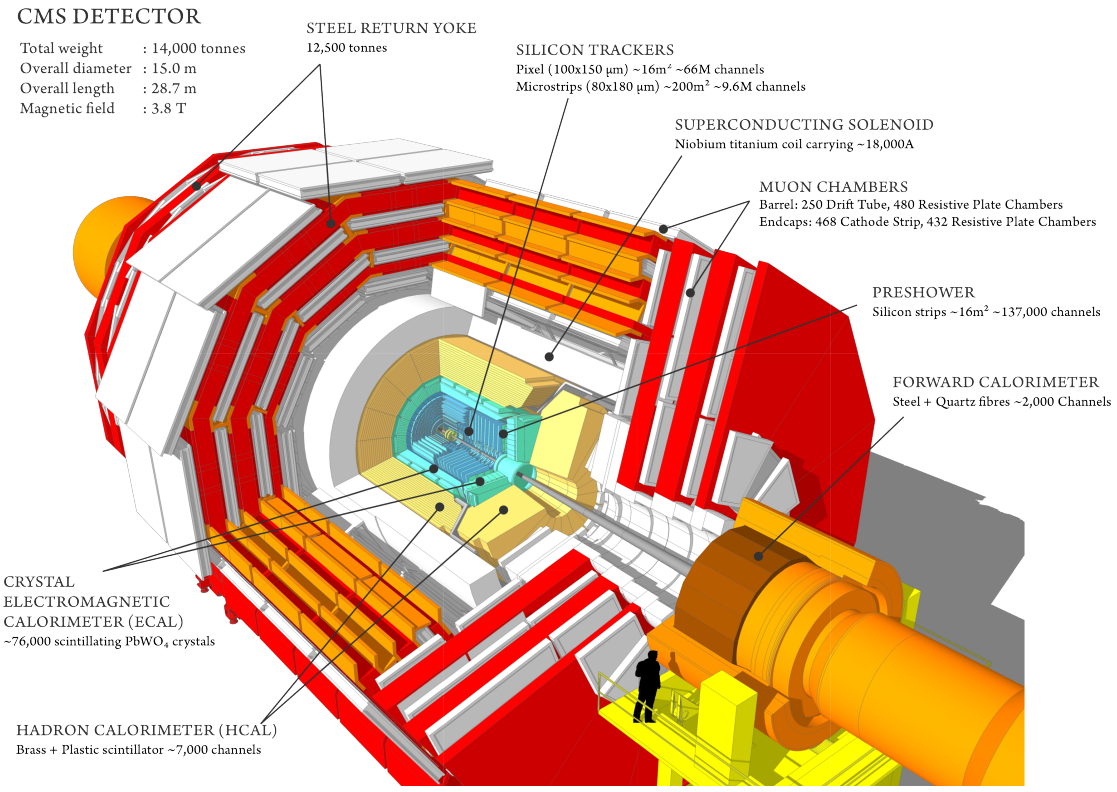


FIGURE 3.4: A cut-away diagram of the CMS detector, with the main sub-detectors labelled. [63]

$$\eta = -\ln[\tan(\frac{\theta}{2})] \tag{3.1}$$

In the massless limit η coincides with the rapidity, y, defined in equation 3.2.

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z} \tag{3.2}$$

This has the property of being invariant under Lorentz transformation along the z-axis, which is useful because it allows us to define observables independent of the momentum component in the z-direction of the initial state. This comes in particularly useful at a hadron collider where parton distribution functions lead to a less well-defined initial state than is the case in a lepton collider.

3.2.1 Overview

The CMS detector is 21.6 metres long and 14.6 metres in diameter and weighs 14,000 tons. The layout of the sub-detectors can be seen in Figure 3.4. The detector is built around the superconducting solenoidal magnet that is 12.5 m long and 6 m in inner diameter. Although designed to produce an axial field of 4 T, the magnet is found to produce a peak stable field of 3.8 T, used during normal running. The inside of the magnet contains the inner tracking system along with the electromagnetic and hadronic calorimeters. The inner tracker is made up of the silicon pixel detector, closest to the interaction point, surrounded by several layers of silicon strip tracker. Together they provide highly accurate measurements of the tracks of charged particles produced at or near the interaction point, allowing primary and displaced vertex location. Surrounding the inner tracker system is a homogeneous electromagnetic calorimeter (ECAL), providing excellent energy resolution for photons and electrons. It is made of lead tungstate ($PbWO_4$) scintillating crystals, attached to silicon photomultipliers to detect the scintillation light. Between this and the inside of the solenoid is a sampling hadronic calorimeter (HCAL), consisting of alternating layers of brass and active plastic scintillator, which is used to measure the energy of hadronic particles. Surrounding the magnet is an iron return yoke that contains and guides the magnetic field, as well as supporting the muon chambers and the forward detectors.

This chapter will give a detailed summary of the sub-detectors important in the reported analyses. A more thorough description of the detector can be found in [56].

3.2.2 The Inner Tracker

The innermost detector of CMS is the silicon tracker [64], designed to resolve the trajectories of charged particles moving within the magnetic field and to provide high resolution vertex reconstruction. The precise tracking information of particles produced during proton-proton collisions allows the measurement of their momentum and charge. Vertex location is important in identifying displaced vertices, such as those from heavy quark decays.

The high flux of particles near the interaction point presents many technical challenges for a tracking detector: it must be granular enough to distinguish between the many hundreds of particles passing through it at design luminosity, and it must be sufficiently radiation

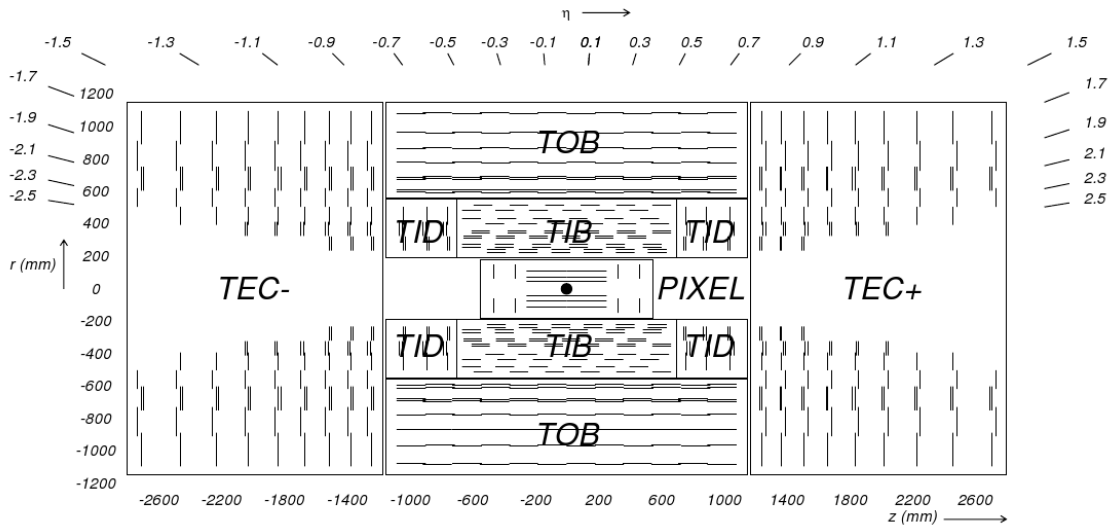


FIGURE 3.5: A schematic view of the CMS inner tracker, including the position of the sub-detectors [65].

hard to survive the harsh radiation environment. Being most suited to these challenges, silicon was chosen as the basis for the tracker. To keep the high granularity at the centre of the detector, but at the same time to keep costs at a manageable level, the tracker is split into two distinct regions: the pixel detector closest to the beamline and the strip tracker surrounding it. In total, the tracker has a length of 5.8 m, a radius of 2.5 m and covers the region up to $|\eta| < 2.5$. Figure 3.5 shows a detailed layout of the full tracking system.

The pixel detector consists of three barrel layers at radii of 4.4, 7.3 and 10.2 cm, respectively, arranged cylindrically with a length of 53 cm each. Two disks at each end, at $|z| = 34.5$ and $|z| = 46.5$ cm, and extending radially from 6 to 15 cm, form the pixel endcaps. The pixel detector contains 66 million pixels, each with an area of $100 \times 150 \mu\text{m}^2$, covering a total of 1.06 m^2.

Surrounding the pixel detector is the silicon strip detector, which extends radially from 20 to 116 cm. The strip tracker is divided into four parts: the Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB), comprising the barrel portion of the detector, and the Tracker Inner Disks (TID) and Tracker EndCaps (TEC) that complete the forward coverage of the detector. The TIB and TID, which make up the inner section, extend radially out to 55 cm and contain 4 layers of strips in the barrel and 3 disk layers at each end. The strips are, in general, parallel to the z-axis, providing measurements in both r and z. The inner two layers of the TIB and the first two disks of the TID additionally contain stereo modules,

modules containing two sets of strips attached back-to-back at an angle of 100 mrad with respect to one another, to provide ϕ measurements. The typical size of the silicon strips in the inner region of the tracker is 10 cm $\times$ 80 μm, with a thickness of 320 μm.

The TOB surrounds the TIB and TID, and extends the coverage to $|z| < 118$ cm and $r < 116$ cm. The TOB contains six layers of silicon strips which, due to their greater distance from the interaction point, have larger dimensions of 25 cm $\times$ 180 μm with a thickness of 500 μm. Like the inner detectors, the first two layers of the TOB contain back-to-back modules, as can be seen in Figure 3.5. The TEC completes the inner tracker structure, each endcap comprising nine layers that extend to $|z| < 280$ cm. The strip tracker contains a total of 9.3 million strips giving an active area of 198 m^2.

Minimum bias events[1] along with simulation have been used to study the spatial and momentum resolution of the tracker detector [66]. The resolution of the track parameters, including that of the transverse momentum of the track, are calculated using the 'track residual' method. Tracks are reconstructed in the tracker volume as described in Section 5.2.1. The track residual is the difference in the $x - y$ co-ordinate of the reconstructed tracker hit and the position that its associated track passes through the layer. The residuals form a Gaussian distribution around 0, the standard deviation of which is the resolution of the measurement.

Figure 3.6 shows the resolution of the transverse momentum and impact parameter of tracks reconstructed in the CMS inner tracker as a function of η, calculated for single, isolated muons of $p_T = 1$, 10 and 100 GeV. The resolution becomes worse at higher values of $|\eta|$ because the extrapolation of the track from its innermost hit to the beamline, where the parameters are calculated, becomes larger. Additionally, there are fewer hits contributing to each track, resulting in a less constrained fit on the track.

The resolution of the impact parameter is dominated by two components: the spatial resolution of the detector and multiple scatterings of the particle. At high momentum a particle undergoes fewer scatterings that significantly influence its course, meaning that the impact parameter resolution is dominated by the spatial resolution of the detector. Lower momentum particles are more heavily influenced by multiple scattering in the tracker volume, causing the visible degradation in resolution at lower p_T.

[1] Data events selected by a minimum bias trigger; a trigger that randomly selects events to record. This creates a dataset in which there is no selection bias introduced by the trigger system itself.

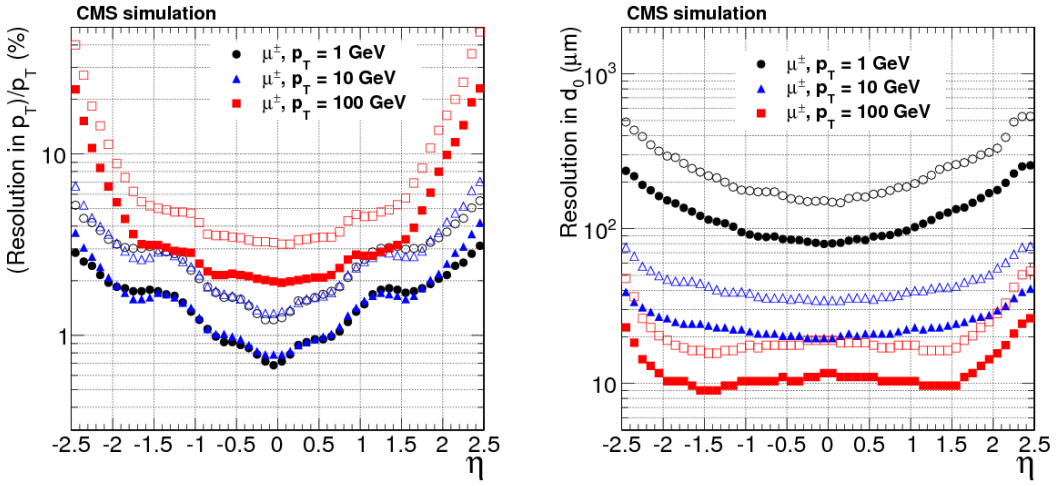


FIGURE 3.6: Resolution, as a function of η, for single, isolated muons of $p_T = 1$, 10 and 100 GeV/c of the transverse momentum (left) and impact parameter (right) of reconstructed tracks in the CMS inner tracker. For each bin in η, solid symbols correspond to one standard deviation of a Gaussian fitted to the residuals distribution, whilst the open symbols correspond to two standard deviations [67].

The momentum resolution of the tracker is limited for particles with high momentum by the strength of the solenoidal magnetic field. At low momentum the resolution is again limited by multiple scattering of the particle within the tracker volume. High momentum particles (100 GeV) have a p_T resolution of $\sim 2-3\%$ up to $|\eta| < 1.6$, where it rapidly deteriorates. The degradation near $|\eta| = 1$ is caused by the gap between the tracker barrel and endcap. The best relative precision in p_T is measured to be for tracks with $p_T \approx 3$ GeV [67].

3.2.3 The Electromagnetic Calorimeter

Once it has passed through the tracker system, the next detector that a particle encounters is the electromagnetic calorimeter (ECAL) [68], designed to measure the energy of electrons and photons. The ECAL is made up of the barrel section (EB), covering up to $|\eta| < 1.479$, and two endcaps (EE) in the range $1.479 < |\eta| < 3.0$. In order to discriminate between neutral pions and photons in the endcap region, a silicon preshower detector is placed in front of the endcaps covering the range $1.653 < |\eta| < 2.6$. In order to fit inside the solenoid along with the hadronic calorimeter system, the ECAL is required to be compact, so a homogeneous detector with a short radiation length was selected. The material chosen for construction of the ECAL had to be sufficiently radiation hard to survive long data taking periods, especially in the high flux environment in the forward regions of the

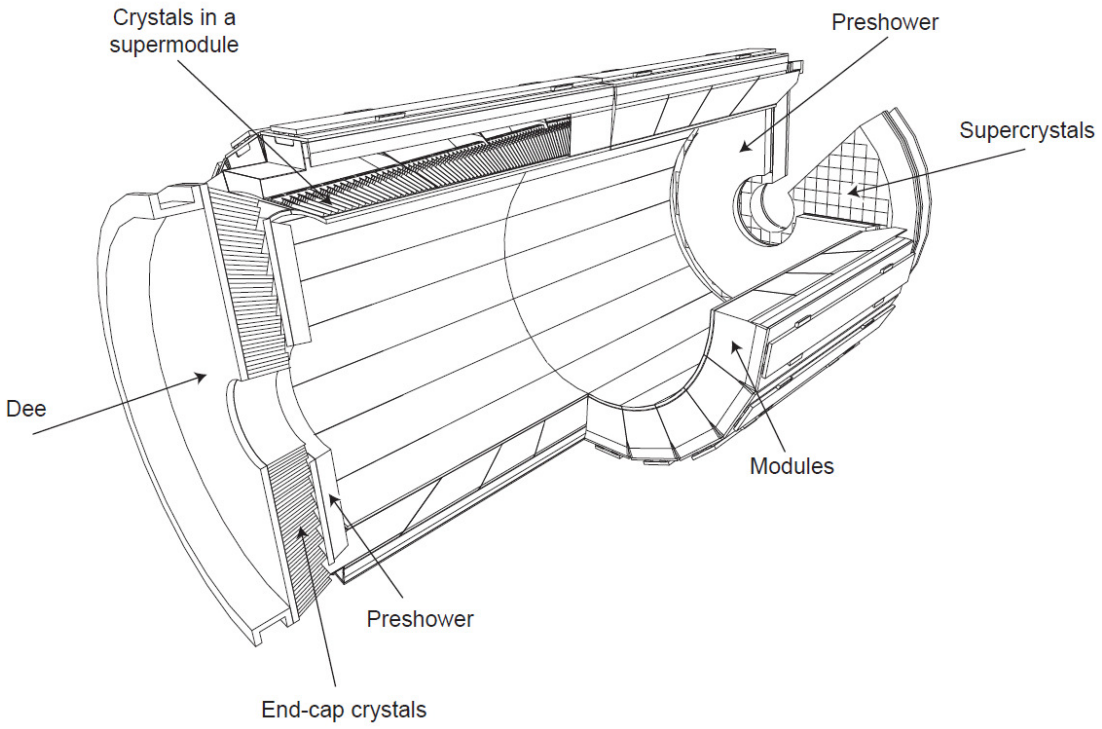


FIGURE 3.7: The layout of the CMS ECAL, including the position of supermodules and the different sub-detectors [56].

endcaps nearest the beamline. In addition, the detector was required to be as hermetic as possible for measurements of missing transvere energy. Fulfilling these criteria with a high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radius (2.2 cm), scintillating lead tungstate ($PbWO_4$) crystals were chosen. The scintillation light is collected by avalanche photo-diodes (APDs) in the barrel region and vacuum photo-triodes (VPTs) in the endcap.

The ECAL barrel contains 61,200 crystals covering the absolute pseudorapidity range up to 1.479. The crystals are grouped into clusters of between 400 and 500, which are, in turn, grouped into 36 supermodules. One supermodule covers 20° in ϕ and makes up one half of the barrel in length, as can be seen in Figure 3.7. Each crystal covers a solid angle of approximately 0.0174 $\times$ 0.0174 in $\eta - \phi$, and is slightly tapered towards the interaction point, although the actual dimensions vary throughout the barrel. In order to avoid particles disappearing through gaps between crystals, they are mounted such that their axes are rotated by an angle of 3° with respect to the nominal interaction point. The length of each crystal is 230 mm, or 25.8 radiation lengths. The total active volume of the barrel is 8.14 m^3 and it weighs 67.4 tons.

The ECAL endcaps cover the region $1.479 < |\eta| < 3.0$, and are placed at $|z| = 3.154$, taking into account the shift caused by the CMS magnet when it is switched on. Each endcap is divided into two halves, called "Dees", which contain 3662 crystals each. The crystals, each of which measures 28.62×28.62 mm^2 and 30×30 mm^2 at the inside and outside face, respectively, are arranged into groups of 5×5 arrays or "supercrystals". The endcap crytals are 220 mm long, corresponding to 24.7 radiation lengths. This gives the endcaps a total volume of 2.90 m^3 and a total mass of 24.0 tons.

The final element of the ECAL is the preshower detector; a two-layered sampling calorimeter located in front of the endcap in the fiducial region $1.653 < |\eta| < 2.6$. The aim of this detector is to discriminate between neutral pions and photons in the endcap region using the energy deposition profile of the particles. To this end, the detector contains two layers of lead radiators to initiate electromagnetic showers, each backed by silicon strip sensors to measure the deposited energy. The two silicon layers are arranged at right angles to each other to provide two co-ordinates for fine position resolution of the hits. The initial lead plate is 2 radiation lengths thick, and the second a single radiation length thick, meaning that 95% of incident single photons will begin showering before the second silicon detector. The silicon strip sensors have a width of 2 mm, and a thickness of 320 μm. The preshower has an overall thickness of 20 cm.

By the end of the first LHC running period in early 2013 the ECAL was running at a very high efficiency, with only a small fraction of non-operational channels: around 1% in EB, 2% in EE and 3% in ES [69]. The observed energy resolution of the ECAL has been measured and found to agree with that expected from simulation and test-beam studies [70]. For electrons showering in the centre of a barrel crystal, the energy resolution is:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{3.3}$$

where E is the electron energy in GeV and $\oplus$ denotes addition in quadrature. The three terms in this equation are the stochastic term, the noise term and the constant term, respectively. The stochastic term arises from event-to-event fluctuations in energy deposition, for example larger or smaller lateral shower containment and varying preshower energy deposition. The noise term originates from the electronic noise associated with

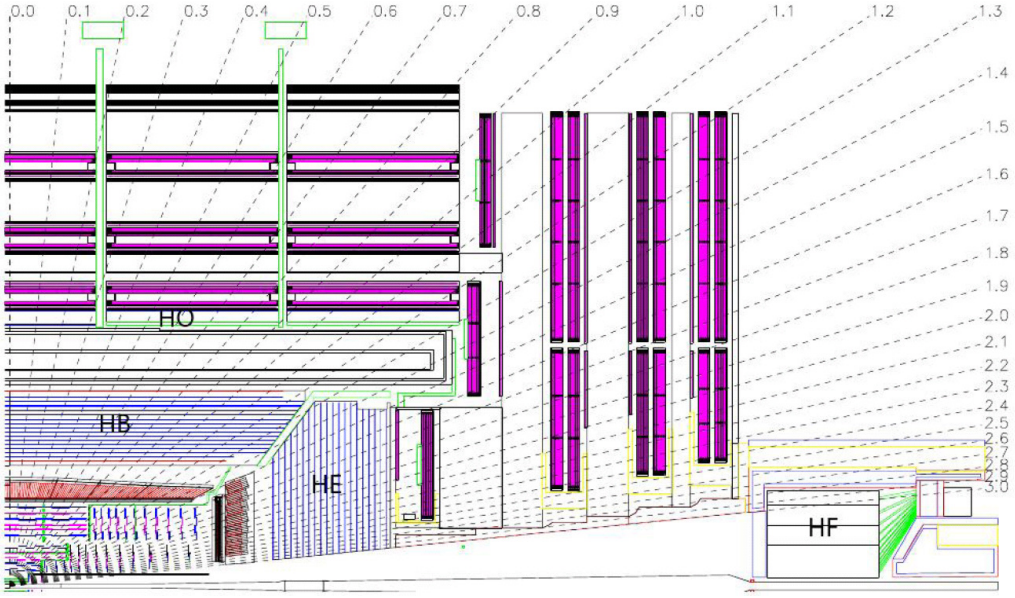


FIGURE 3.8: A schematic view of the HCAL in situe within the CMS detector [56].

digitisation and amplification of the signal. The final constant term is dominated by non-uniformity of longitudinal light collection and energy leakage from the back of the ECAL crystals.

3.2.4 The Hadronic Calorimeter

Particles that have not been stopped by the ECAL will next reach the hadronic calorimeter (HCAL) system [71]. The HCAL is designed to measure the energy of hadronic particles, including protons, neutrons, pions and kaons, as well as to determine the missing transverse energy of events containing neutrinos or other exotic weakly-interacting particles. The HCAL is a sampling calorimeter made up of four sub-detectors: the HCAL Barrel (HB) and Endcaps (HE) that sit between the ECAL and the solenoid, the Outer Hadronic calorimeter (HO) that sits outside the solenoid to catch the end of very high energy hadronic showers that would otherwise escape into the muon system, and the Forward Hadronic calormeter (HF) in a position along the beamline outside the main calorimeter. The locations of the four sub-detectors can be seen in Figure 3.8.

The HB is a sampling calorimeter that covers the range $|\eta| < 1.3$. It contains 36 identical azimuthal wedges made up of eight 50.5 mm-thick and six 56.55 mm-thick flat brass absorber plates interspersed with 3.7 mm thick plastic scintillator tiles. For structural stability the inner- and outermost plates are made of stainless steel (40 mm and 75 mm

thick, respectively). They are assembled using a staggered geometry that ensures particles originating from the nominal interaction point cannot pass through gaps between plates. The plastic scintillators are separated into 16 sectors in η, resulting in a segmentation of 0.087 $\times$0.087 in η and ϕ. Wavelength shifting fibres carry the light from each tile to hybrid photo-diodes (HPDs). At $\eta = 0$ the HB has an effective length of 5.82 interaction lengths, which rises with pseudorapidity to a maximum of 10.6 interaction lengths. The ECAL acts as an additional absorber layer in front of the HCAL, contributing an additional 1.1 interaction lengths in which hadronic showers can develop.

The HE covers the region $1.3 < |\eta| < 3$, a range expected to contain approximately 34% of final state particles. The HE is constructed from the same materials as the HB, with 79 mm-thick brass absorbers and 9 mm thick plastic scintillators. The granularity of the HE matches the HB up to $|\eta| < 1.6$, but decreases to 0.17×0.17 in $\eta \times \phi$ for $|\eta| \geq 1.6$. The total length of the HE, including the ECAL crystals, is approximately 10 interaction lengths.

The restriction in space imposed by the solenoid means that the combined stopping power of the ECAL and the HB is not sufficient for all hadronic showers. For this reason an additional hadronic calorimeter has been placed outside the solenoid to measure the energy of late-developing and/or extended hadronic showers. The HO calorimeter uses the solenoid coil as an additional absorber, equivalent to $1.4/\sin\theta$ interaction lengths, bringing the minimum effective absorber thickness to 11.8 interaction lengths. To match the exterior muon systems the HO is divided into 5 rings in z, and resides in the first sensitive layer within the iron return yoke. In the central ring, where the overall depth of the HB is smallest, an additional layer of plastic scintillator is placed outside a steel absorber to increase the overall interaction length. The granularity of the scintillator tiles in the HO is the same as that in the HB, and they are arranged in such a way as to roughly map the layout of the HB tiles. This allows consistent clusters to be created throughout the sub-detectors of the HCAL.

The HCAL is completed by the HF calorimeters, cylindrical structures that begin at $z = \pm 11.1$ m and cover the forward pseudorapidity region $3.0 < |\eta| < 5.2$ [72]. This forward region is a very hostile environment; almost 90% of the energy deposited in the detector per proton-proton interaction is deposited in the HF. For this reason different technologies are required to ensure the radiation hardness of the HF. Quartz fibres are used as the active

medium in the HF due to their fast response and radiation hardness. Incoming particles above a given threshold generate Cherenkov radiation - radiation produced when a charged particle moves above the phase velocity of light in a medium - within this volume. This threshold is as low as 190keV for electrons in the quartz of the HF [73]. The calorimeter consists of steel plate absorbers, 5 mm thick, with grooves for the quartz fibres. The total absorber length is 165 cm, or 10 interaction lengths. The HF is equipped with two sets of fibres running parallel to the beamline, one of which extends the full length of the detector whilst the second begins a further 22 cm along the beamline. As photons and electrons deposit a large fraction of their energy in the first 22 cm of their shower depth, this second set of fibres allows discrimination between showers from these particles and those from hadrons. The HF has a spatial resolution of 0.175×0.175 in η and ϕ. The Cherenkov light from the fibres is channeled to photomultiplier tubes (PMTs) contained in Readout BoXes (RBXs) attached to the detector.

The energy resolution of the HCAL was measured using a test beam of electrons, pions, protons and muons [74]. As the ECAL will affect an incoming particle before it arrives an the HCAL, an ECAL crystal was included in the test beam setup. The hadronic energy resolution of the HCAL-ECAL system is parameterised using:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus C^2 \tag{3.4}$$

where E is the hadronic energy in GeV, S is the stochastic term, and C the constant term. The values of S and C were measured at $S = 0.847 \pm 0.016 \text{GeV}^{1/2}$ and $C = 0.074 \pm 0.008$ for both the barrel and endcap regions [75]. The HF was found to have values $S = 1.98 \text{GeV}^{1/2}$ and $C = 0.09$ [76].

3.2.5 Muon System

As implied by the name, the muon system is responsible for the identification, measurement and triggering of muons [77]. As muons are minimum ionizing particles they are expected to pass through the main bulk of the inner material of CMS and leave the detector without depositing most of their energy as other particles will. For this reason a series of tracking systems are placed outside the other detectors. In contrast to the uniform inner magnetic field, the magnetic field outside the solenoid varies greatly with η. In order to

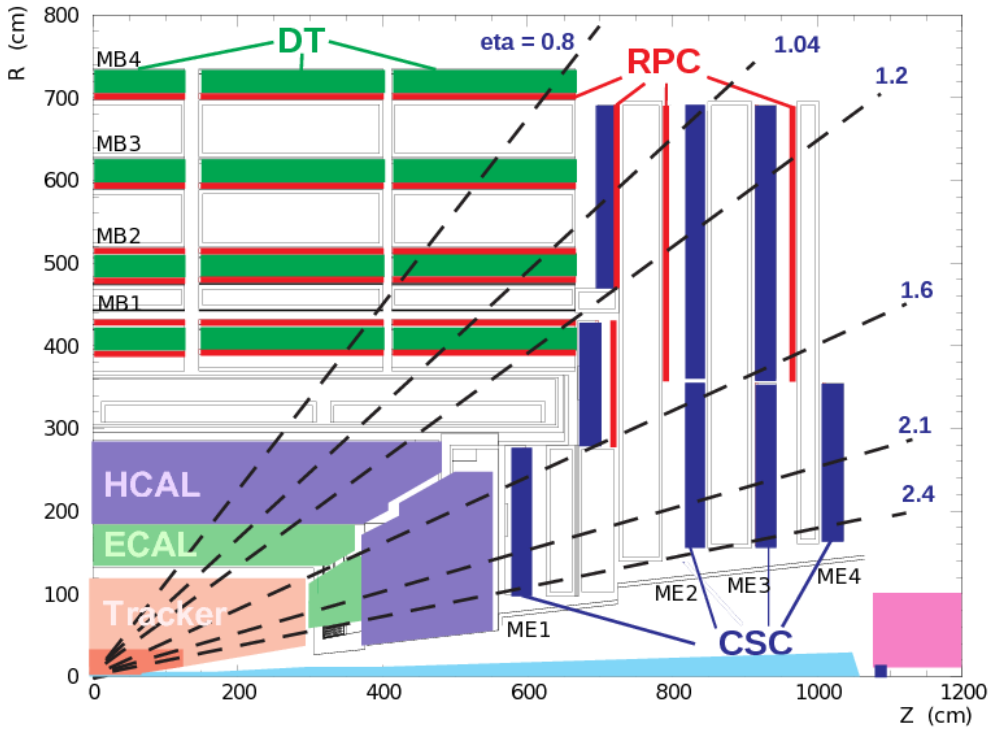


FIGURE 3.9: The location of the various CMS muon systems [78].

maintain sensitivity in as wide a geometric acceptance as possible, three different detector technologies are employed that suit these differing conditions.

As in the other sub-detectors, a barrel and endcap design is employed for the muon system. The barrel, covering the pseudorapidity range $|\eta| < 1.2$, contains a total of 250 Drift Tube (DT) chambers, which are suited to the low muon rates and low magnetic field strength of the barrel region. Each DT contains a mixture of Argon (15%) and CO_2 (85%) gas around a 2.4 m wire that is aligned in the $r-\eta$ direction. The cell width, or maximum drift path, was selected as 21 mm, giving a maximum drift time of 380 ns. The DTs are grouped into four stations that are interspersed amongst the layers of the iron return yoke. These four stations are divided into five wheels along the z-axis, similar to the layout of the HO. There are a total of sixty DTs in the first three stations, and seventy in the outermost station. The DT chambers are arranged in three superlayers (SLs), each of which contain 4 layers of drift cells staggered by half a cell in distance to improve angular resolution. The central SL contains a wire running orthogonal to the beamline, providing a measurement in z. The outer two SLs run parallel to the beamline providing measurements in the magnet-bending plane ($r-\phi$). The outermost station is constructed without the innermost SL, and so only measures the ϕ co-ordinate.

In the endcap region ($0.9 < |\eta| < 2.4$) Cathode Strip Chambers (CSCs) are used to perform muon track measurements. These are more suited to the higher rate of muons (due to their fast response time, which is around 4.5 ns) and the non-uniform magnetic field expected in this region, and are capable of achieving high precision measurements, providing a muon trigger. The CSCs are multi-wire proportional chambers containing 6 anode wires between 7 cathode panels. The wires run azimuthally to measure the polar angle of the muon track. The endcap has four stations of CSCs that are, as in the barrel, placed between the iron return yoke of the solenoid. The CSCs achieve of a spatial resolution, in the $r - \phi$ plane, of 75 μm for the inner chambers and 150 μm for the outer chambers.

The muon system is completed by Resistive Plate Chambers (RPCs), gaseous parallel-plate detectors used in both the barrel and endcap regions up to a pseudorapidity of $|\eta| < 1.6$. With a fast response time of 1.26 ns (far shorter than the design bunch-crossing time of 25 ns for the LHC), the RPCs are designed primarily for muon triggering but can also provide enough spatial resolution information to aid in track reconstruction. The RPCs are double-gap chambers of two parallel electrodes constructed from graphite coated plastic. The barrel region contains six layers of RPCs, two in each of the first two stations and one in each of the remaining. The extra inner RPCs allow triggering on lower p_T muons that may not reach the outer stations. The first three stations of the endcap also house RPCs, allowing triggering on coincidence between hits in the RPCs and CSCs.

The muon momentum resolution has been measured using two separate methods [79]. Firstly, the resolution has been measured in the range $20 < p_T < 100$ GeV/c, as within this range a reference momentum for the muon can be extracted from the inner tracker. For these muons, the relative p_T resolution is found to be between 1.3% and 2.0% in the barrel and better than 6% in the endcap. These measurements agree well with simulation. It was found that muons reconstructed solely from hits in the muon chambers have a p_T resolution of better than 10% in the barrel region. Secondly, cosmic rays are used to evaluate the momentum resolution for high energy muons in the barrel region. The resolution is found to be better than 10% up to an energy of 1 TeV/c^2.

3.2.6 Trigger System

The LHC provides bunch crossings in the four experiments every 50 ns[2], causing tens of millions of collisions every second. Each event within the detector produces up to 2 MBytes of data, meaning that at the nominal collision rate CMS will be producing data at a rate of over 1 Terabyte per second. It is clearly impossible to extract this phenomenal amount of data from the detector.

Many of the events produced will be uninteresting from a physics perspective; the total proton-proton cross section of 101.7 ± 1.4 mb [80] is many orders of magnitude higher than the corresponding cross sections for any of the processes of interest at the LHC. These scattering events, dominated by low-p_T beam remnants, are by far the most common process observed at the LHC, but provide little in the way of useful information for a prospective physics analysis. It is therefore desirable, as well as necessary from a read-out and storage perspective, to select and read out only the interesting events occurring within the detector. CMS has the ability to record events at a rate of $\sim O(100)$Hz, so a reduction of 5 orders of magnitude in the event rate is required. The system that carries out this selection is called the 'Trigger' system, and in CMS is divided into two separate steps: the Level 1 trigger (L1) and the High Level Trigger (HLT) [81].

The L1 trigger is comprised of custom electronics built into the CMS front-end electronics and designed to reduce the event rate to 100 kHz. The high operation speed means that only information from the calorimeter and muon systems can be used in the L1 trigger. Readout information from these sub-detectors is stored in front-end pipelines that can store details from up to 128 bunch crossings. Corresponding to a real-time of 3.2 μs, this is the maximum allowable latency in the system. To make the decision about which events to pass to the high level trigger, the L1 looks at the detector with a coarse granularity. Calorimeter towers with large, isolated energy deposits are considered interesting at this stage, and initial estimates of the event's missing transverse energy are calculated. Within 25ns a decision is made on whether to continue processing any given event: if successful the event is passed through the CMS Data Acquisition (DAQ) system [82] to the HLT computer farm stored in the service cabin of CMS. If the event does not pass, it is removed from the pipeline buffers and permanently lost.

[2]The gap between bunch crossings is expected to drop to 25 ns during Run 2 high energy operation.

There are around 512 front-end buffers that store the information waiting for a L1 decision, and these must all be collated into a single location for further processing. This is achieved by a system of switch fabrics - a combination of physical switching modules and software directing the data to its destination - connecting various sub-detector readout units to the further filtering modules. The data flow is controlled by an event management system.

High Level Trigger filters are then applied to the data in a sequential process to optimise the data flow. Initially, a further decision is made based solely on calorimeter and muon system information. This is done to avoid saturating the system's bandwidth with the large volume of tracker information; these filters reduce the number of events passed to the final HLT filters by at least an order of magnitude. After this initial step the full event data is transferred to the HLT system so that the filter algorithms can be applied.

The HLT is able to achieve excellent performance because it has access to all the information associated with a given event. The L1 trigger only has access to its local information, whereas the HLT can see the information from the entire detector simultaneously. Additionally it gains access to information not available on the timescale of the L1, such as tracker information and the full granularity of the calorimeters. Finally, the HLT carries out basic physics object reconstruction and filters events containing different signatures into separate, self-contained datasets. These are stored on readily accessible media to allow quick and easy identification of appropriate data for any studied process. The HLT saves these selected events at a rate of $\sim$ 100 Hz.

Chapter 4

Certification of Tracker Data

Before it can be used for any sort of analysis, the quality of the data coming from the detector must be checked. A large number of quality tests are applied to the data, both as it is read off from the detector and after event reconstruction (described in Chapter 5). If it passes these tests, the data is made available for use in analysis. This process is referred to as *data certification*, and the first half of this chapter will give an overview of the process. The second part of the chapter will discuss improvements made to the certification procedure for the pixel tracker implemented during the long shutdown of the LHC (LS1).

4.1 Data Certification and Quality Monitoring

4.1.1 System Overview

In order to ensure the accuracy of physics analyses carried out by CMS there must be rigorous checks carried out on the quality and validity of the recorded data. CMS has a common, standard system for Data Quality Monitoring (DQM) that provides a workflow for each sub-detector and physics group to establish the integrity of the data for their own needs, and for the overall performance of the detector. Monitoring of both 'online' (real-time during run taking) and 'offline' (during prompt event reconstruction) quantities is available. The system comprises several key stages that will be discussed briefly in this chapter:

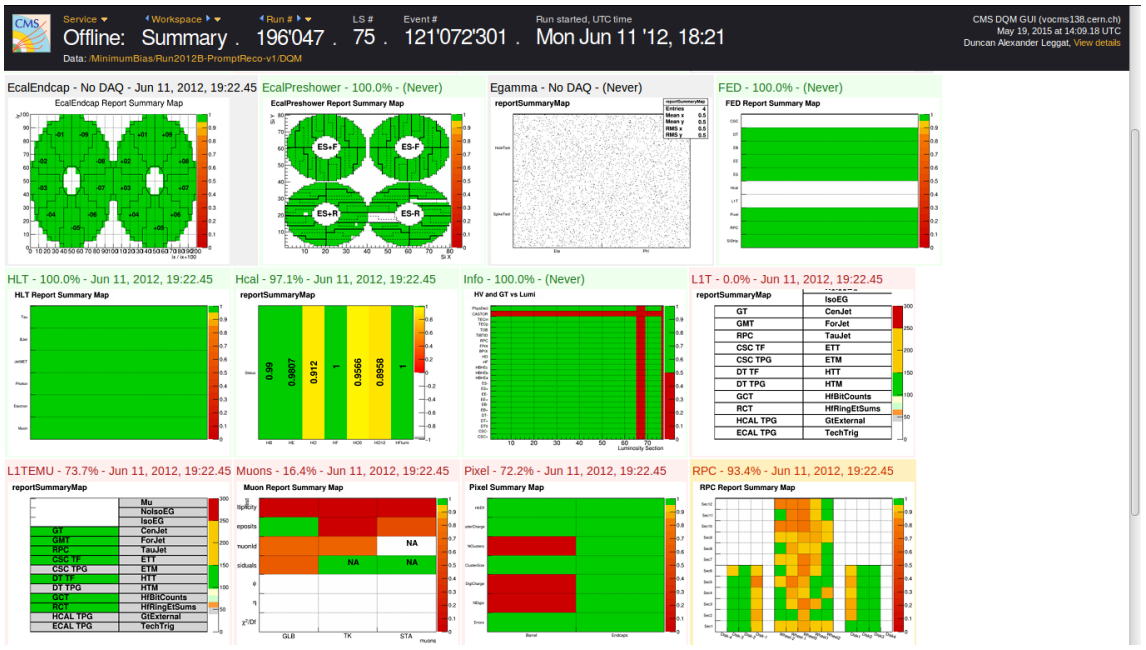


FIGURE 4.1: A screenshot of the CMS DQM GUI showing the offline summary page for a run taken during 2012. The plots on this screen are links to more detailed front pages for each CMS subsystem. As explained in the text, the DQM shifter is responsible for monitoring these distributions and informing the relevant expert should a problem arise.

- The creation, filling and archiving of 'Monitor Elements': customisable ROOT [83] objects containing user-defined distributions of monitorable information from the detector;
- Automated quality testing of monitor elements;
- Visualisation of monitor elements via a Graphical User Interface (GUI);
- Certification of datasets for physics analysis via the Run Registry website.

The ultimate goal of the DQM system is to quickly and accurately find problems with the detector - in either hardware or reconstruction code - to ensure good detector and operation efficiency. The GUI is structured such that there is a front page containing select monitor elements that are known to be sensitive to problems, whilst storing all monitor elements in a series of directories. During operations there is someone monitoring this front page (seen in Figure 4.1) to identify emerging problems, who is known as the *DQM shifter*. If something is seen, they will inform an expert who can use the rest of the information stored in the system to diagnose the problem.

Both the online and offline DQM systems are divided into four sequential stages: Data processing, visualisation via the GUI, certification (both automatically and by the shifter) and sign-off by subsystem experts.

4.1.2 Online DQM

The online DQM system creates two distinct sets of monitoring histograms, one at the point of the high level trigger filters (see Chapter 3.2.6 for details of the trigger system), and one based on the data that is read out of the detector.

The high-level trigger filters (as used in regular event processing) operate at 100 kHz and produce a largely reduced selection of histograms every 23.3 seconds of operation (known as a *luminosity section*). These histograms are then summed across the different filter streams and sent to a storage manager proxy server, which is used to distribute events (along with these histograms) to the different DQM applications.

Each DQM application, of which there is generally one per subsystem, receives event and trigger histogram information from the storage manager proxy at a rate of 10-15 Hz. The subsystems in charge of each application select trigger paths from which to receive events relevant to their needs. Upon receiving the raw data each application runs its choice of reconstruction and analysis modules to create the monitor elements it needs, and conducts any associated quality tests. Each subsystem team provides an Extensible Markup Language (XML) configuration file that details which tests should be run, along with the parameters that define when warning and error flags should be set. XML is a markup language designed to be simple and generic, offering an excellent existing standard for the composition of these files.

The monitor elements are then displayed on the central DQM GUI in real time. Nominal 'reference runs' (earlier runs during which optimal detector performance was observed) may be displayed as an overlay on any monitor element to show differences in distributions and to isolate problems. During the run the distributions are periodically saved to a ROOT file before being fully archived at the end of each run. This archiving includes combining all information from the different applications into a single file, uploading that file to the central GUI server and also backing it up to tape. Automatic certification information from each subsystem is harvested from these files, and uploaded to the Run Registry

conditions database described in Section 4.1.4. Information is stored on disk for several months for browsing and web-based monitoring (WBM).

4.1.3 Offline DQM

As seen in Figure 4.2, many offline tasks produce DQM information, including prompt and re-reconstruction of data along with validation of alignment, calibration and software releases. Prompt reconstruction occurs immediately after read-out from the detector, and relies on a priori estimates of the detector conditions. Re-reconstruction occurs much later once the conditions of the detector during the run have been studied and fully understood, allowing a more accurate event reconstruction. The details of the reconstruction software will be discussed in Chapter 5. Although each task has widely varying requirements, the offline DQM workflow has been split into a CMS-wide two-step process.

During the first step the histogram monitor elements are created and filled using event data. As with event reconstruction, this step is parallelised across events, meaning many instances of the monitor element histograms exist at this stage. The resulting histograms are stored along with other output information in the output event data files. These files are merged by the CMS data processing system and the histograms they contain are summed together.

The second step, or harvesting step, is used to extract, sum and create efficiency and quality information across entire runs. The harvesting step is run as the last step in the event processing, and occurs after the parallelised events have been merged. As such, it has access to the monitor elements from all processed events. This includes status information from the data acquisition (DAQ) and Detector Control System (DCS), an automated system that controls the safe, correct and efficient operation of the detector [85]. All this information is combined within subsystem-specific algorithms to produce detector or physics object summaries, which, in turn, are used to calculate efficiency and quality information. The harvesting algorithms also produce a preliminary automatic data certification decision. All of the output from the harvesting step, including histograms, quality tests and certification results are saved in a ROOT file that is uploaded to the central DQM GUI.

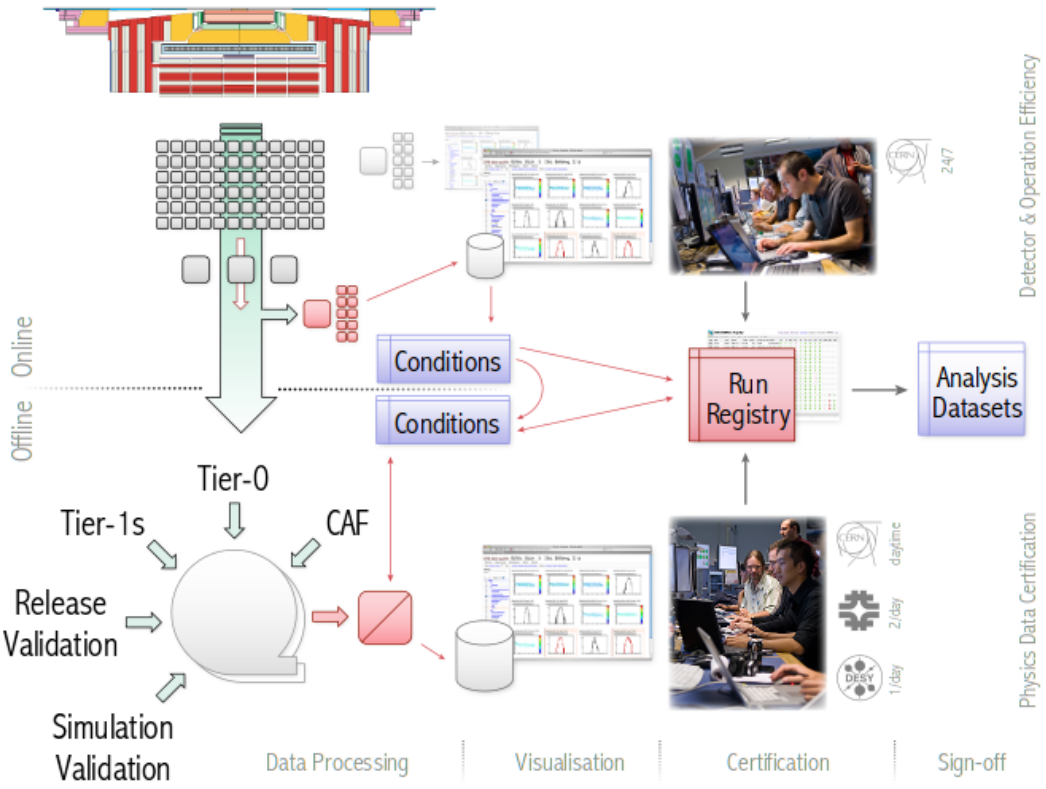


FIGURE 4.2: The structure of the online and offline DQM sequence. The first step, data processing, occurs during data readout (online) and in reconstruction and validation processes (offline) and produces DQM information for the considered events. The second step, visualisation, is carried out by the GUI shown in Figure 4.1. The certification and sign off steps are carried out by each subsystem using the run registry tool [84].

As with the online DQM, the output of the offline sequence is merged in the DQM GUI server, after which it is backed up to tape and cached on disk for several months. All the online and offline DQM applications are designed to be aesthetically similar and to sit alongside each other on the GUI, allowing all collaborators quick and easy access to all quality information. After uploading, the automatic certification decisions, in the form of quality flags, are extracted and uploaded to the run registry. As was the case in the online DQM, reference runs may be displayed alongside the distributions; this is especially important for data certification. The results of the quality flags are propagated to the Run Registry conditions database (discussed in Section 4.1.4) and Data Bookkeeping Service (DBS), an online database that tracks all datasets available within CMS.

4.1.4 Run Registry

As previously mentioned, the Run Registry [86] is a central part of the CMS data certification workflow. Its roles include the collection, management and tracking of data certification information, and it is used to display this information to the entire collaboration. The run registry consists of a web-based user interface and the conditions database, a persistent store of the condition of the detector over time.

The data certification sign-off process begins with online and offline shifters: physicists who monitor output from the different subsystems in real-time and after re-processing, respectively. They fill the run registry with basic run information, along with any information relating to any unexpected occurrences during the run. This information, along with the automatically-generated flags from each subsystem, is used to generate an initial binary quality flag for a run.

Once the initial automated decision has been made, the shifter examines a variety of distributions that have been specified for this procedure for each subsystem. These distributions are carefully chosen to identify foreseeable potential problems within each part of the detector. The shifter then adds any further notes that are needed to the run registry, and may override the automatic certification flag depending on what they have observed.

This combined quality flag is then referred to detector and physics object groups for confirmation. These subgroups have regular sign-off meetings to discuss the verdict and communicate the final result to the experiment. This final outcome is recorded in the offline conditions database and DBS, where they are also accessible through the CMS data discovery interface (DAS), a search-engine designed to interface with DBS.

4.2 Certification and Monitoring of the Pixel Detector

The pixel detector is at the heart of CMS. It is integral to the reconstruction of tracks used in the particle flow algorithm, determining the position of primary and secondary event vertices, and measuring, with high resolution, the path of all charged particles in the event. These roles are crucial for the majority of physics analyses performed by CMS, and as such, the detector must be constantly monitored in order to ensure optimal operation and performance. For this purpose when the detector is running there is always

a member of the shift crew monitoring the online DQM system, which includes monitoring the operation of the pixel detector. If problems arise that are visible in the DQM system they will contact experts, or Detector On-Calls (DOCs), who will be able to diagnose and solve the problem in a timely fashion to reduce detector down time as far as possible.

Before discussing certification of pixel data, it is important to understand the pixel read out system and the different hardware involved. A pixel module consists of 16 readout chips (ROCs), each of which contains 4160 pixels. The ROCs perform zero-suppression with a variable threshold in order to reduce the size of the data payload read out. In the barrel, the modules are mounted in groups of 8 on carbon fibre supports (called *ladders*), whilst in the endcap they are mounted on trapezoidal *blades*, of which there are 24 per disk. Information from the module is passed to an Analog Optohybrid: a chip that amplifies the module's output and converts it to a laser drive current with adjustable gain and threshold. The 1310 nm laser sends an amplitude modulated pulse along an optical fibre link, via patch panels, to off-detector Front-End Drivers (FEDs).

The FEDs digitise the analog input signal and subsequently process the resulting digital signal, before transferring the information to the CMS DAQ via a first-in first-out data-link. Each FED has optical inputs from 36 pixel modules, each of which has a dedicated analog-to-digital converter (ADC). The ADCs are mounted in groups of 4 across 9 daughter boards on the FED. Post digitisation the signals are sent to one of 4 field-programmable gate arrays (FPGAs) that process the data. This processing includes the initial decoding of the now digital hit information and adjustments to the data for changing pixel conditions. For example, even small changes of temperature in the pixel detector lead to different levels of background noise in the detector itself and the readout electronics, and can be accounted for here. These data (along with meta-data generated by the FED itself) are then sent on to the CMS DAQ, where trigger decisions and readout occur. A more complete description of the pixel readout electronics can be found in [87].

4.2.1 Pixel Data Certification

The primary tool for the certification of pixel data is a 2D summary histogram stored in the DQM GUI. This histogram (or summary map) highlights the results of quality tests performed using several key pixel distributions, deemed to be the most important for determining whether the pixel detector is running correctly. The summary map, shown

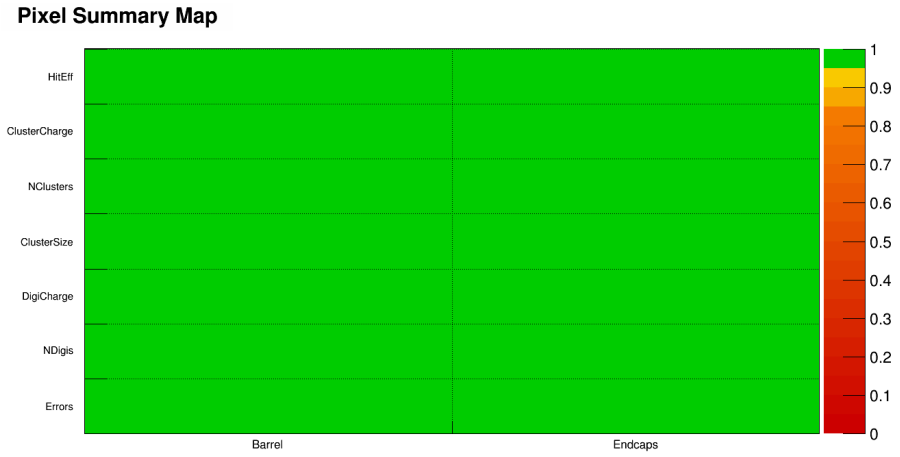


FIGURE 4.3: The summary map for the pixel detector as seen in the DQM GUI. The map is divided along the x-axis into barrel and endcap, and along the y-axis by the chosen quality tests. A green square indicates nominal values for the relevant quality test, whereas a red box would indicate an error flag set. An orange box indicates a warning flag.

in Figure 4.3, is split between the barrel and the endcap along the x-axis, with the y-axis divided between the different summary variables. When quality tests indicate that a particular part of the detector is running normally, the relevant square in the map is displayed green. When a problem is detected, the area appears red.

Automatic certification of recorded data is performed using the summary map; if an error appears the run is recorded in the run registry as bad. In reality, the pixel detector is too complex to base a decision on whether a run is good or bad exclusively on the summary map, thus human verification of each run is required. Once the reason that a run has been labelled bad has been identified, the run can be labelled as bad should the problem negatively affect the recorded data, and good otherwise.

Testing the quality of a given distribution means identifying a key parameter of the distribution, such as its mean, and checking whether it lies within the desired range. Further ranges can then be selected to indicate when a warning or error flag should be set. The DQM GUI displays warnings and errors as yellow and red highlights, respectively, on the distributions associated with the quality tests. There are a wide variety of available quality tests to meet the different requirements of the different subsystems. These tests are documented within the CMS software [84].

4.2.1.1 Summary Variables

The version of the summary map used during Run 1 was designed out of necessity before the first collisions occurred based on what was a priori considered important within the detector. During ongoing running it became apparent that the tests chosen for the pixel summary map did not accurately reflect the nature of problems that began to arise in the detector, and that a redesign would become necessary when possible. The original summary variables that were chosen are as follows:

- **Hit Efficiency**: This is defined as the fraction of reconstructed tracks that contain hits in the pixel detector. Typically, around 90% of reconstructed tracks originate within the pixel detector. If this number is significantly lower, there is likely to be a problem with the recording of data from the pixel detector. An error is shown on the DQM GUI if less than 1% of all reconstructed tracks include hits in the pixel detector.

- **Number of Digis**: The term 'digi', used throughout CMS reconstruction, refers to the number of Analogue-To-Digital Converter (ADC) counts from a given channel per event. In the pixel detector, this can be roughly translated as the number of incident electrons in each silicon pixel. If the number of incident particles is above a certain threshold, normally between 1000-2000 electrons for each individual pixel, then the channel is used for event reconstruction. In the pixel detector digis (and later, clusters) are referred to as 'on-track' if they are included in a reconstructed track, and 'off-track' if they are not. All observables are recorded for both on- and off-track digis and clusters, but only the on-track information is used for the summary map.

  The number of digis per ladder for the barrel (or per blade for the endcap) is used as a measure of the occupancy of the event; if there are an unusually high number of digis then this could indicate that there is a problem with the readout electronics or that the run is particularly noisy, and if there are too few digis it could imply problems with the high voltage supply or other readout problems. As can be seen in Figure 4.4, the expected number of digis varies greatly with distance from the interaction point and is much higher in the inner layers of the pixel detector; it also varies widely along the z range of the detector. This results in a wide range of acceptable values

for the number of digis per ladder, which is not necessarily indicative of a problem arising in any of them.

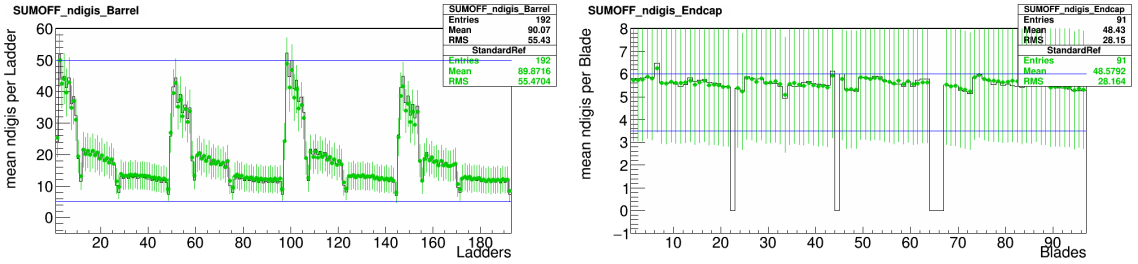


FIGURE 4.4: The average number of digis per ladder in the barrel (left), and per blade in the endcap (right). This figure also demonstrates the reference overlay that can be applied in the DQM GUI. The area within the blue lines is the expected range; if a large number of the ladders (blades) are outside this range an error flag is set. The layer structure of the pixel detector can be clearly seen in this distribution, with the ladders registering a larger number of digis being closer to the beamline. The dips in the endcap distribution are the result of blades excluded from data acquisition, and are not included in the error flag calculation.

- **Digi Charge**: This is defined as the mean ADC count per ladder (blade for the endcap), and is shown in Figure 4.5. The expected range is between 85 and 115 counts in the barrel (95 and 115 in the endcap). A warning is set if less than 90% of the ladders (blades) are within this range, and an error flag is set for less than 75%. A value outside of this expected range can be caused by read out problems, high voltage supply failure or other electronics problems.

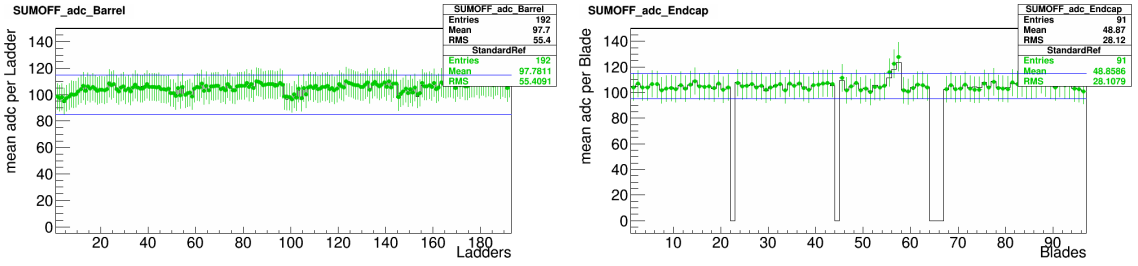


FIGURE 4.5: The average number of ADC counts per ladder in the barrel (left), and per blade in the endcap (right). The area within the blue lines is the expected range; if a large number of the ladders (blades) are outside this range an error flag is set.

- **Number of Clusters**: The tracking algorithm for CMS (discussed in Chapter 5) takes digis from the pixel and strip detectors as its input and, in a similar manner to clustering jets from calorimeter energy deposits, collects neighbouring digis into 'clusters'. These clusters are then used in the iterative track reconstruction described in Chapter 5. From a DQM perspective, the distribution associated with the clusters that is monitored is the mean number of clusters per ladder (or blade). As seen in

Figure 4.6, the expected number of clusters varies between 1.2 and 5 (1.1 and 1.7), and, like the number of digis, varies greatly with distance from the interaction point.

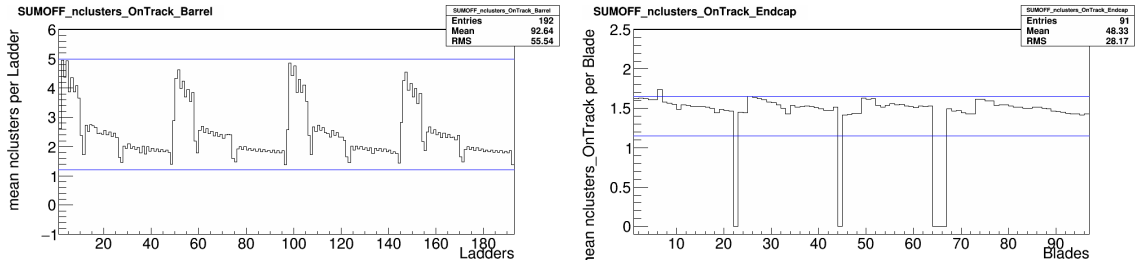


FIGURE 4.6: The average number of cluster per ladder in the barrel (left), and per blade in the endcap (right). The area within the blue lines is the expected range; if a large number of the ladders (blades) are outside this range an error flag is set.

- **Cluster Charge**: Similar to the digi charge, the cluster charge (the sum of the charge of the digis associated with a given cluster), gives a measure of the energy of the particle that created the track. Unlike digi charge, the number of ADC counts in the cluster is converted to the equivalent charge deposit for the cluster charge measurement. As seen in Figure 4.7, typical values for the cluster charge lie in the range 23 - 35 in the barrel (19 - 27 in the endcap), and increase the closer to the interaction point the ladder (or blade) is located. The shape of the cluster charge distribution (shown in Figure 4.8) can be used to diagnose timing related problems with the readout, and the mean varies greatly with the shape.

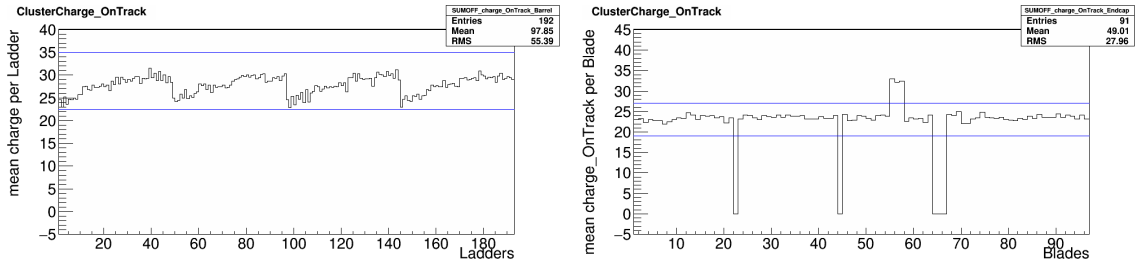


FIGURE 4.7: The average charge per on-track cluster for each ladder in the barrel (left), and blade in the endcap (right). The area within the blue lines is the expected range; if a large number of the ladders (blades) are outside this range an error flag is set.

- **Cluster Size**: The number of digis associated with a cluster is a measure of the physical extent of the cluster. The clustering algorithm requires digis to be adjacent in order to cluster them, so the clusters tend to remain small - the average cluster size is a little over 4 digis per cluster for the barrel, as can be seen in Figure 4.9. In the DQM, the distribution of the average size of clusters per ladder is monitored,

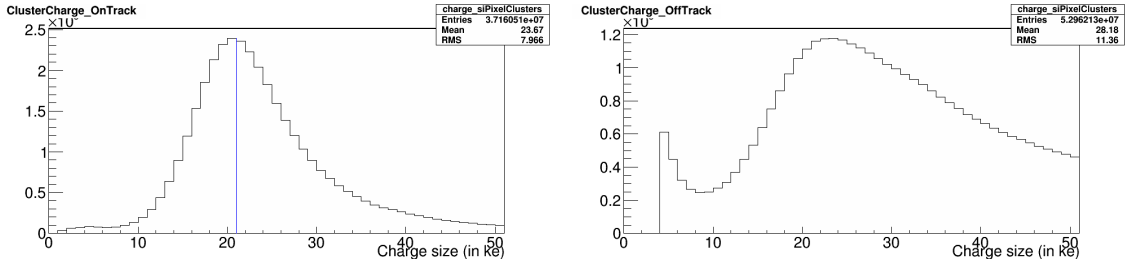


FIGURE 4.8: The cluster charge of every on-track (left) and off-track (right) cluster. Although all distributions are saved for both on- and off-track clusters, only the on-track information is used in the summary map. The shape of this distribution is very sensitive to timing problems; if the on-track cluster size peaks much lower than usual it is an indication that the timing within the pixel detector is incorrect.

shown in Figure 4.10. If less than 90% of ladders have an average cluster size within the expected range of 3 - 6 digis for the barrel (1.7 - 2.2 for the blades) a warning flag is set. An error flag if less than 75% of ladders fulfil this requirement.

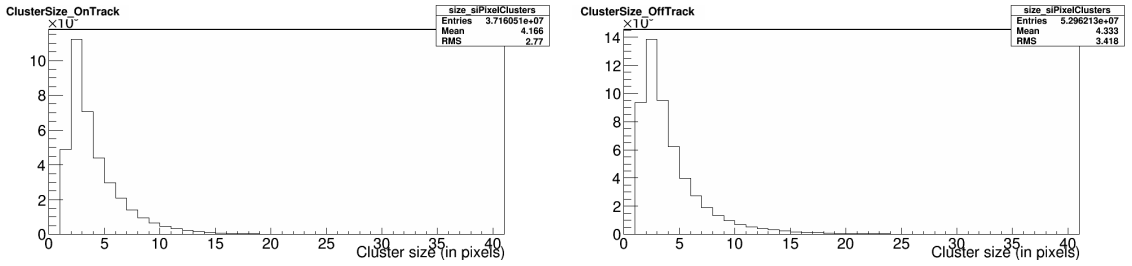


FIGURE 4.9: The number of digis in each on-track (left) and off-track (right) cluster. Digis are required to be adjacent to form a cluster, so the size is usually small.

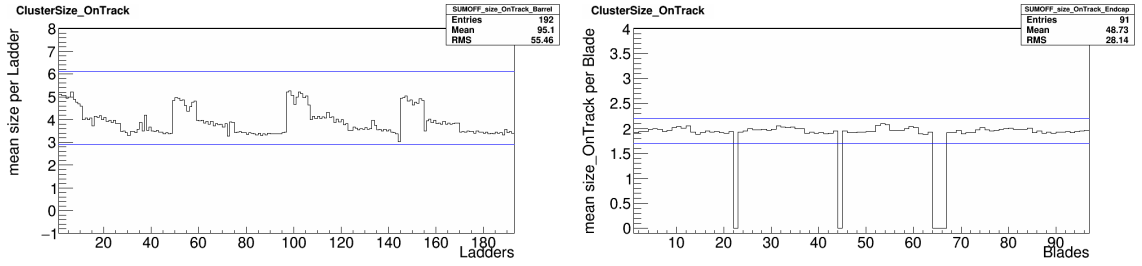


FIGURE 4.10: The average number of digis per cluster for each ladder in the barrel (left), and blade in the endcap (right). The area within the blue lines is the expected range; if a large number of the ladders (blades) are outside this range an error flag is set.

- **Number of Errors**: The number of errors reported by the FEDs. Unlike the other quality flags in the Run 1 Summary Map (which are saved as either 1 for nominal or 0 if there is an error) the number of errors is saved as a fraction. This fraction is the number of FEDs that have not reported an error divided by the total number of active FEDs. FEDs return errors for many reasons: overflow errors and Single-Event Upsets (SEUs) (changes in semi-conductor state caused by high energy

incident particles) are typical examples. Several errors are expected to occur during a standard run, and it is only when many errors appear during a short period of time that this is deemed a problem. Figure 4.11 shows the a map of the FEDs and the number of errors reported during one run in 2012.

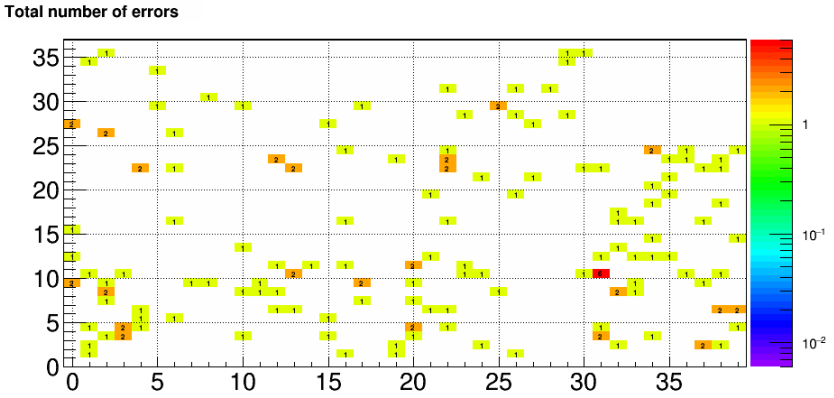


FIGURE 4.11: The number of errors reported by each FED. Over the course of a run there are expected to be some errors, and it is only if a very large number appear that an error flag is set.

The cluster and digi variables contain an additional check that enough information has been recorded throughout the pixel detector to draw valid conclusions from these checks. At least half of the ladders or blades in the barrel or endcap, respectively, are required to have over 25 entries. Typical values for the number of digis or clusters are in the hundreds of thousands for a standard run, and this requirement removes events with zero occupancy from consideration.

4.2.1.2 Summary Variables post Run 1

During ongoing running, it became apparent that the summary map was not reflecting the problems that actually arose in the detector. The reasons for redesigning the summary map include:

- The number of clusters and the number of digis represent very similar sets of information. Comparisons of these two distributions show that it is only the normalisation that is dramatically different between the two, and so it is not necessary to include both.

- Several distributions vary greatly with distance from the interaction point, which makes it difficult to define universal ranges for good and problematic issues. Poorly defined thresholds could result in problems not being detected properly.

- The ranges used for the quantities associated with digis in the barrel detectors during Run 1 were known to be incorrect, resulting in errors appearing for the digi charge and number of digi distributions throughout large portions of Run 1.

Based on the experience of data-taking during Run 1, the new summary map that has been designed for Run 2 includes the following variables:

- **Hit Efficiency**: During Run 1 layer by layer hit efficiency maps of the pixel detector were developed for display in the DQM GUI, subdivided into ladders and blades for the barrel and endcap, respectively. The pixel hit efficiency is the fraction of reconstructed tracks that contain hits in all layers of the pixel detector. The efficiency of a layer is calculated as $1 - \frac{nTracks_{missingHits}}{nTracks_{Total}}$ where the numerator is the number of tracks missing hits in the layer being considered. The maps for the barrel are shown in Figure 4.12, and the endcap in Figure 4.13.

  Throughout Run 1 the pixel detector operated at over 99.5% efficiency (if only fully functioning ROCs are counted in the efficiency calculation), but the level may go as low as 97% during a given run. This allows warning and error flag thresholds to be set, as even slight drops in efficiency could have serious consequences for the event reconstruction and hence physics analyses. This test replaces the previous hit efficiency quality test, which was deemed unnecessary and not informative, given the low threshold requirement that was placed on such a consistently high efficiency.

- **FED Occupancy**: An alternate approach to monitoring the average number of digis per ladder, this test is based on the number of digis in an individual FED divided by the average number of digis across all FEDs. This gives a distribution centred around 1, varying usually between 0.5 and 1.5. It was determined during Run 1 that problems in a limited number of modules did not mean that the run had to be registered as bad. However, it was observed that problems with entire FEDs were much more problematic for data-taking. Each FED reads in data from 36 modules, which means that individual module problems are insignificant next to wider issues

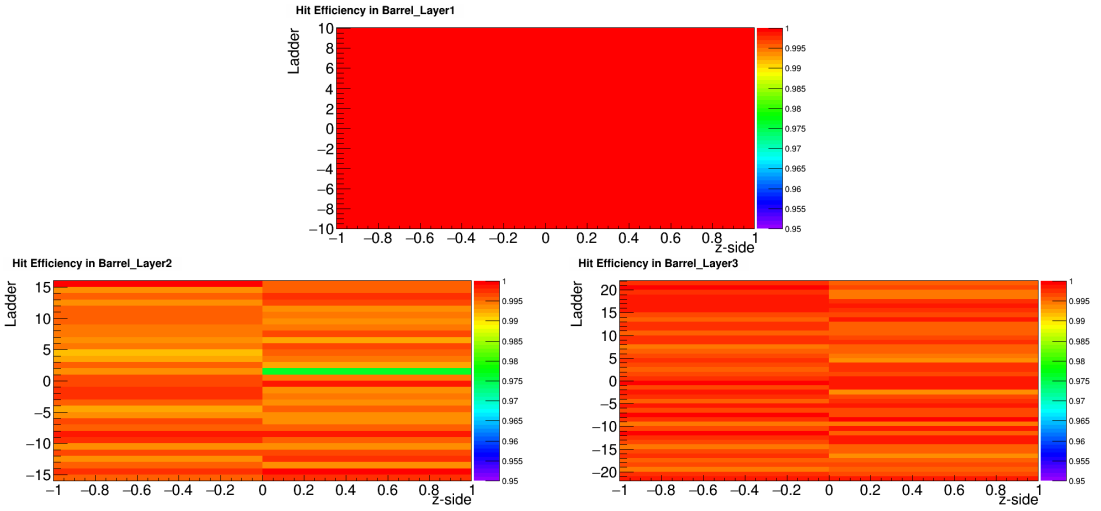


FIGURE 4.12: Hit efficiencies of the ladders in each layer in the barrel of the pixel detector. Layer 1 is set to 100% due to the ambiguity of defining the hit efficiency for the first layer of the pixel detector, in that it is unknown if the track began in Layer 2 or is missing a hit in Layer 1. The efficiency of the remaining layers is at least 99%.

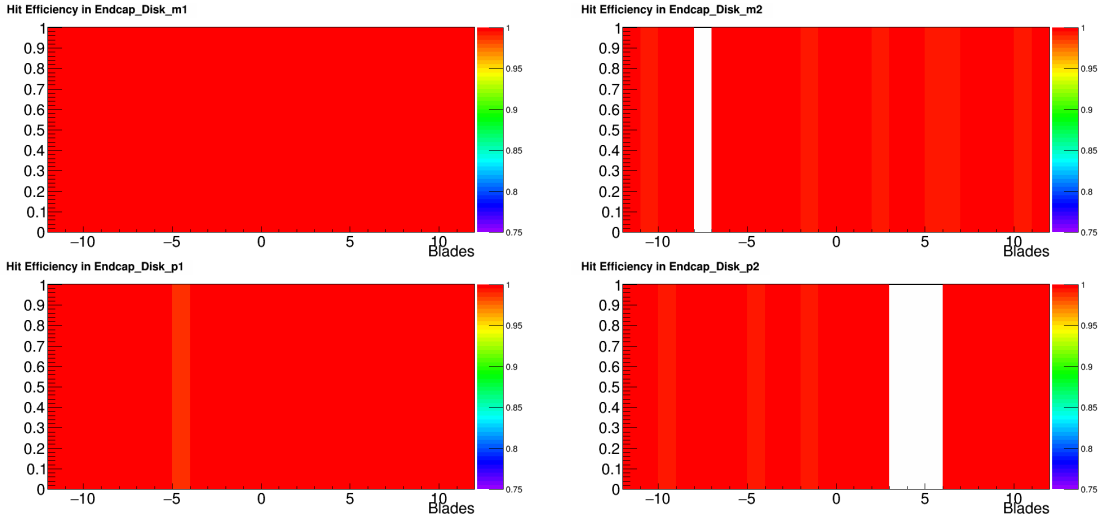


FIGURE 4.13: The hit efficiencies of the blades for each disk in the endcap of the pixel detector. Empty columns are blades that have been removed from data acquisition.

with entire FEDs, a fact that is reflected in this distribution. An example of this distribution from 2012 is shown in Figure 4.14.

- **Track Residual**: The track residual is the difference between the cluster position and the position of the reconstructed track as it passes through the layer in which the cluster has been detected. The distribution of track residuals, shown for both x and y for one ladder in Figure 4.15, is a Gaussian distribution centred around 0, variations in which indicate problems with track reconstruction.

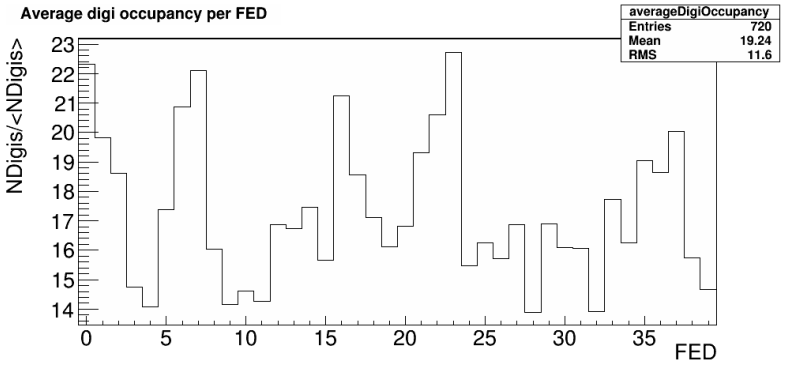


FIGURE 4.14: The average number of digis per FED divided by the average number of digis across all FEDs from a run taken in 2012. During Run 1, this distribution would increase with the number of luminosity sections collected during running, meaning that the values do not centre around 1 as described in the text. During LS1 normalisation to the number of luminosity sections was added, allowing the distribution to be used as part of the Run 2 summary map.

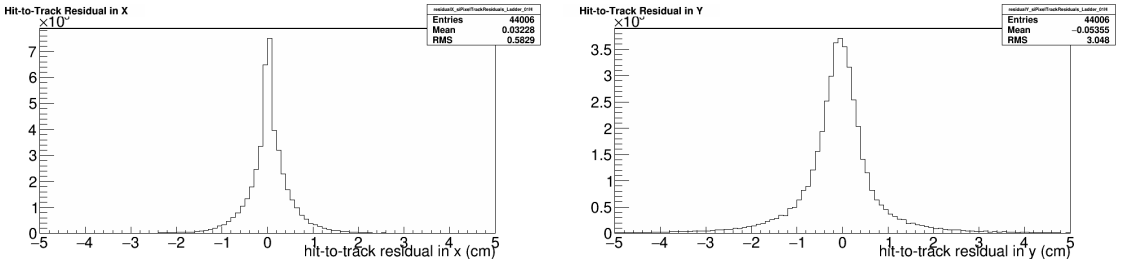


FIGURE 4.15: The hit-to-track residuals in both the x- (left) and y-axis (right) of one ladder in the barrel of the pixel detector. The width of the Gaussian distribution formed is a measure of the resolution of the detector.

The use of the number of errors and cluster charge distributions remain unchanged with respect to the Run 1 summary map. The thresholds for the quality tests were determined using long-term information from the historic DQM, which is described in Section 4.2.2, In addition to reworking the quality tests used to create the summary map, the ability to display warning flags has also been added. Warning flags are useful as they allow shifters to spot a potential problem as it is developing and to intervene before more serious problems appear. The summary map can also now display information on a layer by layer basis instead of simply displaying barrel and endcap information, alleviating the problems with distributions varying greatly over radial and z range.

4.2.2 Trend Plots

In addition to monitoring events and DQM status on a run-by-run basis, it is also important that shifters and experts can see how key operating parameters develop over time; for this reason historical data quality monitoring (hDQM) was developed. hDQM takes the root

output of the DQM process as its input, harvests information on a run-by-run basis and stores it in a pixel detector database object. This database can be accessed directly by the shifter, and various tools exist for analysing the stored data. In addition, a variety of pre-selected trend distributions are compiled nightly during running and displayed via a web server on an internal CMS machine. An example trend plot is shown in Figure 4.16.

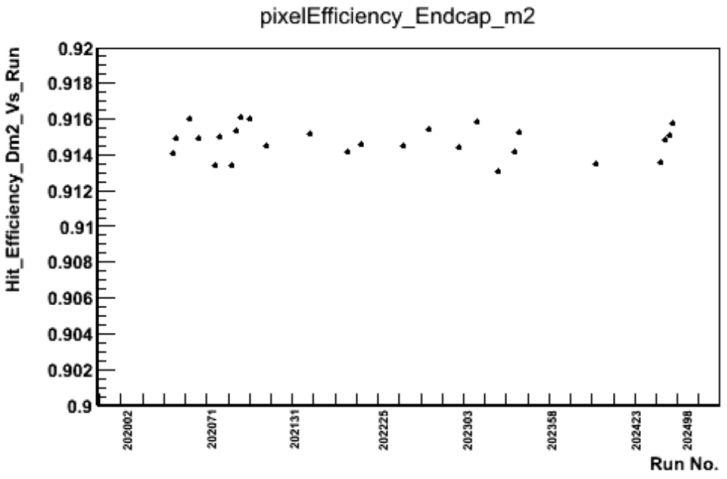


FIGURE 4.16: An example trend plot showing the distribution of hit efficiency (as defined in Section 4.2.1.2) for the second layer of pixel endcap detector as a function of run number. This tool allows the identification of trends in the data not obvious in individual runs.

During LS1 new tools were developed for the hDQM. Methods for plotting multiple distributions together in an aesthetic and meaningful manner were developed, along with automatic axis scaling for ease of comparison. The choice of distributions displayed on the web server has also been reviewed during LS1, and the web page itself is currently in ongoing developement. These developments will make it easy to identify changes in the detector over time and to quickly identify emerging problems during Run 2.

With these new tools in place the pixel detector is in a good state to begin data-taking for Run 2, and has excellent diagnosis tools to quickly identify and solve problems with the detector during early running.

Chapter 5

Event Reconstruction and Simulation

Monte Carlo (MC) simulation is an integral part of modern high energy physics analyses. It is used to develop the methods for extracting signal from the collected data and for the statistical analysis of the final results. For these reasons it is essential that the simulation be as detailed and realistic as possible. The first half of this chapter will discuss the methods used for generating events and their interpretation in the scheme of the CMS detector.

The event reconstruction at CMS can be broadly divided into three steps. In the first step, information from each sub-detector is used to construct 'low-level' objects within each system: clustered energy deposits in the calorimeters and tracks in the tracking detectors. These objects are then passed to the Particle Flow (PF) algorithm [88, 89] which uses information from all the sub-detectors simultaneously to identify and reconstruct all the particles present in the event. Finally, these reconstructed particles are used to reconstruct additional 'high-level objects' such as jets and missing transverse momentum. The second half of this chapter will discuss the object reconstruction techniques used by CMS, with particular emphasis on the PF algorithm.

5.1 Event Simulation

The aim of event simulation at CMS is to create realistic simulated events for the numerous different interactions expected at the LHC, in the same format as data as recorded by the detector. These events can then undergo the same reconstruction process as recorded data.

The simulation is composed of several distinct stages, each of which takes the output of the previous step as input. Simulation begins with the hard interaction, which is calculated using perturbative methods. These events are then typically passed to another program that carries out the parton showering and hadronisation. At this point each event is a list of particles produced in the interaction, their origins and momenta. These are then passed through a simulation of the CMS detector. A full reconstruction of the CMS detector has been created using GEANT4 for this purpose [90]. This simulation includes track propagation through the magnetic field, particle decays and interaction between the particles and the detector material.

The complicated nature of event simulation typically leads generators to contain a vast number of free parameters that can be independently altered to change the output. These parameters mostly govern the non-perturbative aspects of the event, such as the underlying event, which cannot be directly calculated in perturbative QCD. Although many of the parameters have a minimal impact on the final simulation, it is important to attempt to make them all as realistic as possible. To this end, generator 'tunes' are produced based on global fits of recorded data. The *Perugia tunes* [91], in particular the Z2 tune, are used for the simulated datasets used in the analyses presented in this thesis.

In addition to the desired process, a realistic event simulation will include inelastic proton-proton interactions that occur within the same or adjacent bunch crossings, known as pileup. In-time pileup, the interaction of protons in the same bunch crossing as the nominal interaction, cause additional low p_T particles and additional primary vertices. Out-of-time pileup, the interaction of protons from neighbouring bunch crossings (i.e. 50ns before or after the nominal bunch crossing), causes additional deposits of energy within the detector. As out-of-time interactions take place outside of the tracker, they do not contribute to reconstructed primary vertices for the event. In-time pileup is included in event simulation by generating minimum bias events using the PYTHIA generator [92], and mixing them with other simulated events from the process of interest.

In practice, the number of simulated pileup interactions does not match that observed in data, requiring the simulated datasets to be reweighted before use. Throughout the analyses presented in this thesis the simulated events will be reweighted based on a comparison of the distribution of primary vertices in the simulated samples and the recorded data. Changes in the instantaneous luminosity of the beam over time affect the pileup distribution in data, requiring a robust reweighting method that can be used to compare simulated samples with data from different running conditions. The reweighting process has an intrinsic uncertainty associated with it, which is considered when evaluating the systematic uncertainties of the analyses. Section 7.2.2 details the treatment of this uncertainty.

In order to compare MC events to data, a weight must be applied to correctly normalise the sample to its production cross section and the measured luminosity. The event weight is

$$w = \frac{\mathcal{L}\sigma}{N_{sim}} \tag{5.1}$$

where $\mathcal{L}$ is the integrated luminosity of the data, σ is the production cross section of the simulated process, and N_{sim} is the number of generated events. The cross sections are calculated using the methods explained in Section 5.1.1.

5.1.1 Simulated Samples

Single top events are generated using the POWHEG v1.0r1380 generator [93]. POWHEG is a framework designed to interface next-to-leading order (NLO) hard interaction calculations with parton shower generators. After generation, the events are passed to PYTHIA v6.426 [92] for parton showering and hadronisation. The three single top channels described in Section 2.2.1 are treated individually, with separate simulated samples for the top and antitop quarks.

The MadGraph v5.1.3.30 NLO matrix-element based generator [94] is used to generate inclusive $t\bar{t}$ and single boson production samples (Z+jets and W+jets), as well as the signal tZq sample for the second analysis. Samples containing $t\bar{t}$ with additional bosons, i.e. $t\bar{t}Z$ and $t\bar{t}W$, are also produced in this way. As with the POWHEG samples, parton showering and hadronisation are performed using PYTHIA6. The matrix element calculation is matched to the parton shower using the Kt-MLM algorithm [95]. This algorithm requires that all generated showers above a given k_T threshold - 20 GeV by default - be matched to

one of the particles produced at the hard scatter. In the case of additional or unmatched jets the generated event is discarded. The matching introduces a systematic uncertainty that is taken into account by varying the matching thresholds during the simulation process, detailed further in Section 7.3.2. The Z+jets sample generated in this way, a major background to the tZq search, proved to have limited statistics in the trilepton final state. Data driven background estimates were therefore required, and are discussed in Section 6.2.2.

The remaining background samples are treated differently for the two analyses. For the tW analysis inclusive diboson samples were produced and hadronised using PYTHIA6. These samples also proved to contain too few events for a trilepton analysis such as the tZq search, so samples with enriched leptonic decays of the bosons were produced using the MadGraph generator. The WW sample contains decays to 2 leptons and 2 neutrinos. WZ samples were produced for the two final states containing 3 leptons and 1 neutrino, and 2 leptons and 2 quarks. The hadronic decay of the W boson can still be a background for a tri-lepton search when one of the jets fakes a lepton. ZZ samples with final states of 4 leptons or 2 leptons and 2 quarks are produced. As with the WZ sample these could potentially produce a signal if one of the leptons is misidentified or a jet is misidentified as a lepton, respectively.

Lepton-enriched multijet samples are produced using PYTHIA. The predicted inclusive cross section for multijet production is too large to make generating a single inclusive sample practical or effective. For this reason, many multijet samples are produced with different generator-level filters to increase the likelihood that events will pass the analysis cuts. For the tW analysis, multijet samples are produced that are enriched in heavy flavour production and contain filters such that the event must contain at least one muon of $p_T^{\mu} > 15$ GeV/c. These samples are known as 'muon-enriched QCD'.

The decay of τ leptons is handled independently by the TAUOLA package [96], which includes all the different τ decay modes. This is used in conjunction with all the MC generators mentioned above to handle the tau decays in each simulated sample, with the exception of the multijet background samples. The CTEQ6.6M parton distribution functions are used for all samples [21]. In all the samples that include the top quark the top quark mass is taken as $m_t = 172.5$ GeV/c^2.

Table 5.1 summarises the cross sections and numbers of events for the simulated samples. The predicted $t\bar{t}$ production cross section is $\sigma = 252.9^{+6.4}_{-8.6}$ pb as calculated by the Top++2.0 program [97] to next-to-next-to-leading order (NNLO) in perturbative QCD, including soft-gluon resummation to next-to-next-to-leading-log order (NNLL). The top and anti-top single top samples are merged for the purposes of the analysis, and the total cross sections are calculated to next-to-leading order (NLO) in QCD using HATHOR v2.1 [98]. The inclusive single boson cross sections are calculated to NNLO using FEWZ 3.1 [99]. The production cross section of the diboson samples WW, WZ and ZZ are calculated using MCFM 6.6 [100].

5.2 Preliminary Object Reconstruction

CMS uses the Particle Flow (PF) algorithm to reconstruct the full event topology using information from all the sub-detectors. It is important to note that the PF algorithm is just one part in the chain of event reconstruction; low-level object reconstruction such as tracking and energy clustering occurs separately in each sub-detector and are taken as inputs to the PF algorithm. The PF algorithm then produces a list of particles that are used to construct further high-level objects such as jets and missing transverse energy. This section will discuss the preliminary reconstruction that occurs before the particle flow algorithm is used.

5.2.1 Charged Particle Tracking

All charged particles passing through the CMS detector leave hits in the central silicon detector, and the iterative tracking algorithm combines these hits in order to reconstruct their tracks [67]. The resulting tracks provide estimates for many parameters of the outgoing charged particles including the momentum of each particle at the interaction vertex, before any deflection from the magnetic field and the distance of closest approach to this point, known as the impact parameter.

As charged particles make up approximately two thirds of the energy in any given jet, high purity and well reconstructed tracks can greatly improve the accuracy of reconstructed jets; indeed, tracker information proves to be the most important input in the PF algorithm. This is also partly due to the p_T resolution in the tracker which, for charged hadrons with

Process	Notes	$\sigma \cdot$ BR $[pb]$	Number of events	tW	tZq
tZq		0.00783	1038665		✓
tW	tW	11.1	497658	✓	✓
	$\bar{t}W$	11.1	493460	✓	✓
s-channel	t	3.79	259961	✓	✓
	$\bar{t}$	1.76	139974	✓	✓
t-channel	t	56.4	3758227	✓	✓
	$\bar{t}$	30.7	1935072	✓	✓
$t\bar{t}$	Inclusive sample	245.8	6873750	✓	
$t\bar{t}W$		0.232	196046		✓
$t\bar{t}Z$		0.2057	229952		✓
WZ	Inclusive	22.44	10000283	✓	
	$\rightarrow$ 3 lepton 1 neutrino	1.09	2017979		✓
ZZ	Inclusive	9.03	9799908	✓	
	$\rightarrow$ 4 lepton	0.02	4807893		✓
WW	Inclusive	57.07	10000431	✓	
	$\rightarrow$ 2 lepton 2 neutrino	4.93	1933120		✓
W+jets		36257.2	577099051	✓	✓
Z+jets	Low-mass (10-50 GeV)	860.5	7132223	✓	
	High-mass (>50 GeV)	3532.8	30459503	✓	

TABLE 5.1: The simulated samples used in the analyses, including information on the cross section and the number of generated events. The ticks indicate which of the analyses the samples were used in.

p_T up to several hundred GeV, is better than the energy resolution of the calorimeter system. To be fully exploited by the PF algorithm the reconstructed track collection must have found as many of the actual tracks as possible (have a high tracking efficiency), whilst keeping the number of additional, 'fake' tracks to a minimum (have a low 'fake rate'). An iterative tracking algorithm was selected because it fulfilled these criteria [67].

The tracking algorithm can be divided into five important steps:

- **Local Reconstruction** As described in Section 4.2 signals in the strip and pixel tracker are converted to digis and collected into clusters, or hits. The position of each digi and the associated uncertainty are estimated during local reconstruction.
- **Track Seeding** Two or three hits, predominantly from the pixel detector, are combined into 'pairs' and 'triplets', which are used as seeds for the full track identification procedure (pattern recognition).
- **Pattern Recognition** Using a Combined Kalman Filter (a combinatorial variation on a global Kalman Filter) [101], tracker hits are grouped together into potential particle trajectories. This is conducted on a layer by layer basis, beginning in the centre of the detector with the seeds and working outwards. Each additional hit is combined into the 'proto-track' that is then used to estimate the position and uncertainty of the location of the hit in the next layer, taking into account the expected energy loss of the particles. If there is more than one compatible hit per layer, several track candidates are created. The algorithm also allows the propagation of tracks across layers with no hits.
- **Fitting** Each track identified during the pattern recognition step is refitted twice using the Kalman filter: firstly from the innermost seed outwards to reduce bias from the seeding step, and secondly from the outermost layers inwards to avoid bias in the track building stage.
- **Quality Check** Ambiguities in the track finding, introduced if one seed gives rise to multiple tracks or multiple seeds form the same track, are removed with a quality check. Tracks flagged as low quality (typically with few hits and a high χ^2) that share more than half their hits with high quality tracks are removed.

The track finding algorithm is iterative, and repeats these steps six times, beginning with the zeroth iteration. At the end of each iteration the hits associated with high quality

tracks are removed and the next iteration is performed [67]. With each iteration looser cuts are applied, to increase the tracking efficiency whilst keeping the fake rate low by having a reduced pool of hits. The first four iterations use seeds from the pixel tracker, whilst the final two use seeds from the strip tracker. In this way objects that are produced outside of the pixel volume, such as photon conversions or decays of long-lived particles, are still reconstructed.

Using this iterative technique particles with p_T as low as 150 GeV, a production vertex greater than 50 cm from the beam axis and at least 3 hits are reconstructed with a fake rate of the order of 1% [67].

5.2.2 Primary Vertex Reconstruction

The reconstructed charged particle tracks are then used to reconstruct the interaction vertices in the event. For tracks to be eligible for primary vertex reconstruction they must pass several criteria: a low impact parameter with respect to the LHC beam axis; more than a minimum number of pixel and strip hits and a low normalised χ^2 - the χ^2 divided by the number of degrees of freedom of the fit. The tracks that meet these requirements are clustered along the z-axis at their point of closest approach to the beamline, forming a list of primary vertex candidates. An adaptive vertex fitter [102] is used to perform a three dimensional fit on all the vertex candidates and reconstructs all possible vertices for the candidate tracks, along with their associated uncertainties.

Pile-up vertices are expected to contain mostly low-p_T tracks, so the primary vertices are ranked in decreasing order of sum of associated track transverse momenta. The first vertex from this list is then used as the location of the primary interaction for the purposes of analysis, and the rest are interpreted as pile-up interactions.

A similar method is used to locate secondary vertices originating from the decay of b-quarks, which will be discussed in Section 5.4.3.

5.2.3 Calorimeter Clustering Algorithm

The calorimeter clustering algorithm is used to measure the energy and location of particles incident to the CMS calorimeter system [103]. The algorithm works independently of the

PF event reconstruction, with the energy deposits found in this manner linked with charged particle tracks by the PF algorithm to improve the energy measurement of the resulting objects. This combination also helps to resolve high-p_T or closely spaced tracks and can be used to determine the energy deposits coming from neutral hadrons and photons. Clustering is carried out using the energy deposition information from each sub-detector independently. No clustering occurs for the Forward HCAL due to the large cell size; here each cell is considered as an individual cluster in its own right.

The clustering algorithm starts by finding cells with energy in excess of a sub-detector specific threshold and labels them as seeds. These seeds are summed together with adjacent cells, forming topological clusters. The clusters are kept if they contain a total energy over two standard deviations above the electronic noise expected in that part of the calorimeter. These thresholds are 80 MeV for the ECAL barrel, 300 MeV for the ECAL endcap and 800 MeV in the HCAL [103]. Clustering in the ECAL is complicated by photon conversions and Bremsstrahlung radiation; although the charged particle of interest is deflected by the magnetic field, the radiated photons continue in their original direction, causing energy deposits stretched in ϕ. The clustering algorithm is designed to compensate for this effect by allowing extended clusters in the ECAL. These 'superclusters' are used as the ECAL input to the PF algorithm.

5.2.4 Muon Identification

The reconstruction of muons is carried out using two types of tracks: those reconstructed from the inner tracker as described in Section 5.2.1 and those reconstructed within the muon system (known as standalone-muon tracks). Muon tracks begin as reconstructed hits in either the DT or CSC that have been formed into short track segments. They are then combined into tracks using a Kalman filter, as described in Section 5.2.1. From these two types of tracks muons are reconstructed in two ways [79]:

- **Tracker Muon** Every track from the inner tracker with $p_T > 0.5$ GeV and $|p| > 2.5$ GeV is considered a potential muon candidate, and is therefore extrapolated out to the muon system, taking into account loss of energy and uncertainties from multiple scattering. If any muon segments exist in the muon system within the extrapolated track position the track in the inner tracker qualifies as a *tracker muon*. This is also known as the 'inside-out' method.

- **Global Muon** For every standalone-muon track, a search for a corresponding inner track is conducted. The best-matching track from the inner detector is then fitted along with the standalone-muon track to form a *global muon track*. This is also known as the 'outside-in' method.

Prompt muons (muons that were produced directly in a hard interaction) should be well reconstructed by both algorithms, and it is muons that satisfy this criterion that are passed to the PF algorithm for further validation.

5.2.5 Electron Reconstruction

As charged particles pass through the tracker volume, they lose energy through interactions with the material. The majority of charged particles are heavy enough that this energy loss occurs in the form of multiple Coulomb scattering when passing between materials, but in electrons the dominant effect is Bremsstrahlung radiation. Kalman filters are used for track fitting in CMS because they can incorporate noise and other inconsistencies (in the case of track fitting, they are caused by multiple scattering) as Gaussian fluctuations. Bremsstrahlung radiation is highly non-Gaussian, and, as such, electron tracks are not well reconstructed when using a standard Kalman filter. As a result, dedicated electron track reconstruction is used in CMS [104]. Electron tracks are reconstructed with a relaxed Kalman filter to find a complete electron trajectory, and then refitted using a Gaussian-Sum Filter (GSF). A GSF is an altered form of the Kalman filter that interprets uncertainties as the sum of multiple Gaussians instead of individual Gaussians. This allows the GSF to handle the changes in the electron's trajectory far more comprehensively, although at the cost of additional CPU time.

Two different forms of 'electron identification' [105] are used by the PF algorithm as seeds for reconstructing electrons: the ECAL- and tracker-driven approaches. The ECAL-driven approach uses ECAL superclusters as the seeds, projecting back from the centre of the supercluster to the innermost layer of the pixel detector. Hits and general track seeds from within the resulting region are selected as the matching track for the electron. This method is best suited for isolated, high-p_T electrons where the potential track seeds are limited and the ECAL clusters do not overlap with other jet deposits.

The tracker-driven approach is more appropriate for reconstructing low-p_T and non isolated electrons. A high-purity Kalman filter is used for track finding, similar to the iterative tracking described in Section 5.2.1. For low-p_T electrons, which will emit negligible amounts of Bremsstrahlung radiation, these tracks can be directly extrapolated to superclusters in the ECAL. For more substantial amounts of Bremsstrahlung energy loss the tracks are refitted using a GSF, before being characterised with a multivariate estimator [106]. The energy loss for the track can then be estimated.

The seeds generated by the two methods are merged into a single collection on which a GSF is run to determine the final electron track properties. Using the GSF at this stage results in more hits being included in the electron track reconstruction, giving a better estimate of the electron's momentum and the energy lost in its path through the tracker. The GSF tracks are used for the PF reconstruction of electrons.

5.3 The Particle Flow Algorithm

Any given particle is expected to generate a signal in several of the CMS sub-detectors: energy deposits in the ECAL and HCAL and charged particle tracks in the inner tracker and muon systems. The Particle Flow algorithm aims to combine these disparate pieces of information to efficiently identify and reconstruct all particles present in each event. The algorithm begins with a 'link' step that groups together information from the sub-detectors in combinations likely to have come from particles. The algorithm eventually classifies these blocks into reconstructed particles of five types: electron; muon; photon or charged/neutral hadron.

5.3.1 The Link Algorithm

The link algorithm considers pairs of elements from different sub-detectors and rates the quality of the link by their relative distance from each other. In general, links are made by extrapolating the outermost hit of a track in the inner tracker to the calorimeter system. If this position falls within the bounds of a calorimeter cluster a link is made between the two. The size of the acceptance region may be increased by up to a cell to account for non-uniformity in the calorimeters. The distance between the shower maximum of

the cluster and the projected track is characterised by ΔR in the (η, ϕ) plane, where $\Delta R(\eta, \phi) = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

If an ECAL cluster lies within the boundary of an HCAL cluster they are linked, with the distance again defined using the ΔR between the clusters centres. Bremsstrahlung photons are identified by extrapolating tangents from a track's intersection point at each layer in the inner tracker to the ECAL. If this extrapolated path encounters an ECAL cluster it is linked with the track as a possible Bremsstrahlung electron. Global muons already have a link between an inner track and a muon system track, and these are sent directly to the reconstruction algorithm. Typically between one and three items are linked into 'blocks' in this way. By linking as few objects into blocks as possible the event complexity does not greatly affect the performance of the algorithm.

5.3.2 Particle Flow Algorithm

The blocks created by the link algorithm are then passed to the PF algorithm, which carries out the reconstruction and identification of the particles in the event. As particles are identified, the constituent tracks and clusters from the block are removed from the available pool, in a similar manner to the approach used in the iterative tracking.

Muons are considered first because they are expected to be the best reconstructed of the physics objects. The momentum of each muon candidate is measured from the global and inner tracks associated to it. If the measurements are within three standard deviations of each other the muon is classed as a 'particle flow muon' and the tracks from the block are removed from further consideration.

As most of the reconstruction has already occurred, electrons are considered next. In order to discriminate between electrons and charged hadrons, GSF tracks are required to be associated with clusters in the ECAL. The shape of the ECAL supercluster must be laterally thin in order to be consistent with an electromagnetic shower. Various parameters of the tracks, such as the number of hits and χ^2 value, are used to distinguish between electrons and pions. These and other parameters go into a multivariate estimator which, if satisfied, will classify the object as a 'particle flow electron'.

What remains must therefore be objects associated with charged/neutral hadrons and photons. Any remaining tracks with links to an HCAL cluster are classified as charged

hadrons. Care is taken here to avoid double counting by discarding multiple links from tracks to keep only the closest one. The momentum of the track and the energy deposited in the cluster are compared, and any excess in the HCAL is attributed to a neutral hadron or photon, depending on the shape of the deposit. Any remaining clusters with no associated tracks in the ECAL and HCAL are identified as 'PF photons' and 'PF neutral hadrons', respectively.

5.4 High-Level Object Reconstruction

The rest of this chapter covers the methods used by CMS to reconstruct high level objects such as jets and missing transverse energy, and the way these objects are corrected for detector effects.

5.4.1 Jets

Jets are the collimated stream of particles originating from the hadronisation of partons from a hard interaction. Whilst several jet reconstruction algorithms are available, CMS uses the anti-k_T algorithm [107] for its ability to create approximately cone-shaped jets whilst retaining infrared and collinear safety in the face of soft radiation. The algorithm produces jets from a list of object positions and transverse momenta, including each particle in exactly one jet [108]. The algorithm calculates the value $d_{i,j}$ for every pair of objects and d_{iB} for every object, defined as:

$$d_{i,j} = min(\frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2})\frac{\Delta_{i,j}^2}{R^2} \tag{5.2}$$

$$d_{iB} = \frac{1}{k_{T,i}^2} \tag{5.3}$$

where k_T is the transverse momentum of each particle, $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ (where y is rapidity) and R is the size parameter of the jet, set as 0.5 as standard in CMS.

The calculated values of $d_{i,j}$ and d_{iB} are then compared to find the minimum value, d_{min}. If d_{min} is from a single object, it cannot be merged further and is labelled as a complete jet. This jet is then removed from the collection and is not considered further. If d_{min} comes from a pair of objects, these objects are merged into a 'protojet', with momentum

and position calculated by;

$$\begin{aligned} k_T &= k_{T,i} + k_{T,j} \\ y &= [k_{T,i}.y_i + k_{T,j}.y_j]/k_T \\ \phi &= [k_{T,i}.\phi_i + k_{T,j}.\phi_j]/k_T \end{aligned} \tag{5.4}$$

This process is repeated until there are no protojets remaining, and all objects have been clustered into a jet.

The anti-k_T algorithm is used throughout CMS for jet reconstruction, producing distinct jet collections based on the input particle collection. The two sets of jets used in the analyses presented in this thesis are *PF-jets* and *GenJets*.

PF-jets are produced using the list of particles reconstructed by the PF algorithm as input. Although jets are also produced from calorimeter and tracker energy deposits, PF-jets are found to be reconstructed with the most accurate momentum estimates. This is largely due to the energy composition of a typical jet, which comprises approximately 65% of its energy in charged particles, 25% in photons and 10% in neutral hadrons. The relatively poor hadron calorimeter resolution means that the already small portion of energy carried in neutral hadrons becomes less important when reconstructing the jets. The most prominent information therefore comes from the tracker and ECAL deposits. As the particle flow algorithm fully reconstructs the particles produced in the event, not just the tracks and energy deposits, the list of particles used as input to the anti-k_T algorithm, and therefore the produced jets, are reconstructed more accurately. For this reason PFJets are used throughout the analyses presented in this thesis.

GenJets take the energy deposits of MC generator-level particles as their input. They are used for simulation studies, and to examine the generator properties of different processes.

5.4.1.1 Jet Energy Corrections

A detector-level jet, reconstructed from energy deposits in the detector, will generally have a different energy from the corresponding generator-level jet, even when reconstructed using the same jet algorithm. This discrepancy is usually attributed to the non-uniform performance of the CMS calorimeter, and mis-modelling between CMS simulation and

actual performance in areas such as electronic noise and pileup levels. For these reasons the energy of every reconstructed jet must be adjusted to correct for these effects. The ultimate aim of these corrections is to achieve a detector response (the average signal per unit of deposited energy) that is linear and uniform in η.

Discrepancies are characterised into different levels, and the corresponding corrections are applied as a scale factor to the four-momentum of the jet. The levels of correction used in the following analyses are:

- **L1 Pile-up** Corrects the jet energy for additional energy that does not belong to the initial hard scattering, i.e. electronics noise and pile-up.
- **L2 Relative Jet Correction** An η-dependent scale factor that corrects for variations in jet response due to detector non-uniformity. It is derived using the transverse momentum of dijet events: events with one jet in the barrel region $|\eta| < 1.3$ are used to derive the η-dependent correction for the other jet. The aim of this correction is to make the jet response flat in η.
- **L3 Absolute Jet Correction** A p_T-dependent correction used to correct the p_T dependence of the jet reconstruction. This is determined using dileptonic decays in γ^*/Z+jets samples with the goal of giving a flat response in jet p_T .

As the L2 and L3 corrections are derived from simulation events, there is an additional correction that must be applied to data jets to account for the small discrepancy between data and simulated jet response. This is known as the **L2L3 Residual** and is applied to data only. There are additional, high order jet corrections, such as flavour-dependent corrections, but these are not included in these analyses. The final correction applied to each jet is the produce of the individual corrections from each level, and the uncertainties associated with each are considered as source of systematic uncertainty. More detailed descriptions of all the jet energy corrections can be found in [109].

5.4.2 Missing Transverse Energy

Certain weakly interacting particles, most notably the neutrino, escape the detector without leaving any detectable signal in any of the sub-systems. However, these particles are

indirectly observable because of their effect on the overall transverse momentum of the event: in the initial state the proton-proton interaction has zero net momentum, and so this must also be true in the final state. As a large portion of the underlying event and proton remnants escape the detector along the beamline, CMS can only reliably consider the conservation of momentum transverse to the beamline. The missing transverse momentum, denoted as $\vec{E}_T^{miss}$ or $\vec{\not{E}}_T$, is calculated from the vector sum of the transverse momentum of all objects in the event. The $\vec{\not{E}}_T$ is then the negative of this vector, which brings the total transverse momentum of the event to zero. The scalar $\not{E}_T$ is given by the magnitude of this vector.

There are several methods of calculating the $\not{E}_T$ of an event at CMS which, as with jet reconstruction, take different particle collections as input [110]. The $PF\,\not{E}_T$, which will be used through the analyses presented in this thesis, estimates the missing transverse momentum from the list of particles reconstructed by the PF algorithm. Studies have found this to be the most accurate method of estimating $\not{E}_T$ currently used by CMS [110].

The performance of missing transverse momentum energy reconstruction at CMS has been studied using dileptonic Z+jets events and dijet events [111]. These events contain no neutrinos in the final state, and so the $\not{E}_T$ resolution and scale is dominated by detector effects. Through these studies, different levels of corrections have been determined, which can be applied to the calculated $\not{E}_T$ to improve the accuracy of the calculation.

Two levels of correction may be applied: Type 1 Jet Energy Scale (JES) corrections, and Type 2 unclustered corrections. Type 1 corrections compensate for the difference between the jet energy measured at generator and detector levels. Below a certain threshold, typically around 20 GeV, the Type 1 corrections become less reliable, causing inconsistencies in the corrected $\not{E}_T$ distribution. For this reason a Type 2 correction is made based on unclustered calorimeter energy deposits and jets with low p_T . The improved energy resolution obtained using PF reconstruction means that Type 2 corrections are only required if calculating the $\not{E}_T$ without PF. When using PF reconstruction the energy scale of low p_T particles is improved to a point where the Type 2 correction is no longer required. The corrections are provided by a dedicated Jet Energy Scale group within CMS.

The analyses presented in this thesis will use Type 1 corrected PF $\not{E}_T$ to aid identification of the leptonic decay of the $W^{\pm}$ boson, which includes a neutrino.

5.4.3 b-Jet Identification

The lifetime of mesons produced from the hadronisation of quarks produced from hard interactions varies inversely with their mass: mesons formed of light quarks can travel long distances through the CMS detector without decaying whilst those involving heavy quarks decay before reaching the detector volume at all. The b quark's decays to the up and charm quarks via the weak interaction are suppressed by the CKM matrix, meaning that mesons containing b quarks defy this trend and may travel several centimetres from the beam-line before decaying. These unique properties make it possible to specifically identify jets originating from b quarks.

Almost 100% of the top quark branching ratio is to a W boson and a b quark, making it vitally important that b-quarks can be correctly identified when studying processes involving the top quark. By accurately identifying and reconstructing the jets from b-quarks it is possible to significantly reduce the background from other channels that would otherwise look very similar to top decays: W+jets, Z+jets and QCD multijet events are all dominated by jets from light-flavour quarks, c-quark fragmentation and gluons.

CMS employs several 'b-tagging' algorithms of varying complexity to discriminate between jets originating from b-quarks versus light quarks or gluons [112]. These typically use reconstructed objects from the event, for example vertices, tracks and leptons to produce a single discriminator value for each jet in the event; the higher the value of the discriminant the more likely that the jet originated from a b quark. Designated 'working points' for each of the algorithms (along with the algorithms themselves) are defined by the B-Tagging Physics Object Group within CMS. The 'loose', 'medium' and 'tight' working points are defined such that the probability of incorrectly tagging a jet originating from a light quark or gluon (the *mis-tag rate*) is 10%, 1% and 0.1%, respectively [113]. The tagging algorithms can then be evaluated at their different working points by their efficiency - the amount of jets originating from b quarks that are successfully tagged - and their purity - the fraction of jets passing the tagging that are genuine b-jets.

The algorithms can broadly be defined as those relying on the impact parameter of particle tracks, and those that use reconstructed secondary vertices. The impact parameter (IP) is the distance of closest approach of a track in a jet to the reconstructed primary vertex. Thanks to the excellent resolution of the CMS pixel detector, this quantity can be calculated very accurately in three dimensions. The sign of the IP is given by the sign

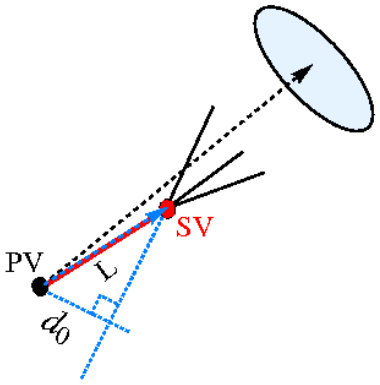


FIGURE 5.1: An event containing a secondary vertex (SV) originating from a B hadron, visible in red, decay. The dashed black line shows the jet axis, whilst L represents the decay length of the B hadron. The diagram also demonstrates the impact parameter, d_0 of a track associated with the jet [114].

of the scalar product of the IP and the jet direction; this tends to give positive IPs to tracks originating from the decay of particles along the jet axis. A related value is the IP significance; defined as IP/σ_{IP} where σ_{IP} is the uncertainty of the IP calculation. In a very general sense, a jet originating from a b quark will contain large, positive values of the IP significance.

The other class of b-tagging algorithm centre around the identification of *secondary vertices*, the point in flight at which the B hadron decays to a jet, as seen in Figure 5.1. Secondary vertices are identified similarly to primary vertices as described in Section 5.2.2, but take already reconstructed jets as the starting point. High-purity tracks (described in Section 5.2.1) in a cone $\Delta R = 0.3$ around the jet axis are grouped together, and an Adaptive Vertex Fitter is used to identify jet vertex candidates. The secondary vertex is considered separate from the primary vertex and originating from a b-quark decay if the following criteria are satisfied:

- they share less than 65% of their associated tracks;
- the significance of the radial separation between the primary and secondary vertices is greater than 3σ;
- the flight direction of the candidate is within $\Delta R < 0.5$ of the jet direction.

In addition, candidates that have a radial distance of greater than 2.5 cm to the primary vertex, or a jet mass close to the K^0 meson mass are rejected. This helps to reduce the

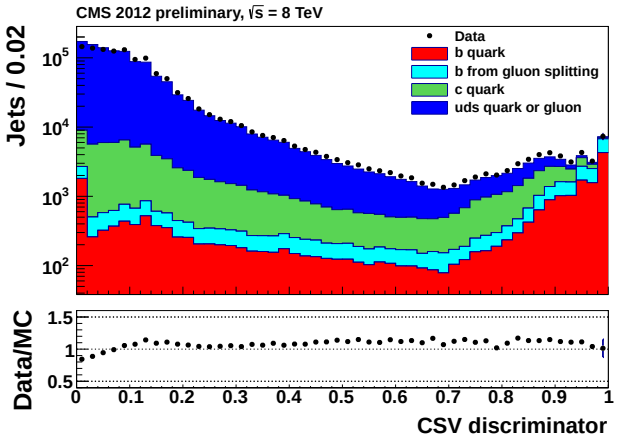


FIGURE 5.2: The distribution of the CSV discriminant for jets of all flavour in data and simulation. Multijet and $t\bar{t}$ events were used for the study. The medium working point, used throughout this thesis for the selection of b-tagged jets, corresponds to a value of the discriminant of 0.679 [115].

contamination of vertices originating from the interaction of particles with the detector material and decays of long-lived particles.

The analyses presented in this thesis use the *Combined Secondary Vertex (CSV)* b-tagging algorithm at the medium working point [115]. The CSV tagger combines information from all variables known to discriminate between b jets and non-b jets. By combining information from secondary vertices with impact parameter information, b-jets with no measurable secondary vertex can also be correctly identified. The discriminant uses a combination of variables including the reconstructed mass, track multiplicity and three dimensional flight distance between the primary and secondary vertices to calculate two ratios: one to discriminate between b and c jets, and one to distinguish between b and light jets. These two discriminants are then combined to give the overall CSV discriminant, which is shown for $t\bar{t}$ and multijet events in Figure 5.2. The chosen medium working point corresponds to a b-tagging efficiency of $62\pm2\%$ and a misidentification rate of $1.51\pm0.02\%$.

Differences in the mis-tag rate and tagging efficiency between simulation and data are accounted for by reweighting the simulated events. Variations in these weights are included as a systematic uncertainty in both studies, discussed in more detail in Section 7.2.5.

Chapter 6

Event Selection and Background Estimation

Once all objects have been identified and reconstructed (as described in the previous chapter) the next step in a physics analysis is to define requirements on the kinematics and topology of the event that will create a signal-enriched subset of the data. Simulated samples are used to optimise the requirements, rejecting as much background as possible whilst maintaining signal in the sample. CMS provides filtered datasets, where the inclusion of any given event in a particular dataset depends on whether the event was selected by a specific trigger path, as discussed in Section 3.2.6. The event selections of both analyses are discussed in the first part of this chapter.

Whilst simulated samples are vital in establishing an understanding of the process under study and its relevant backgrounds, limited sample statistics and poor description of some distributions in the data by simulation often mean that additional methods must be employed to correctly estimate the background contribution in data. These methods include reweighting simulated data to correct for improper modelling, as discussed in Section 6.1.3.1 and 'data-driven' background estimations, as discussed in Section 6.2.2. Additional methods to discriminate between signal and background may then be employed to further enrich the signal sample, such as the multivariate analysis discussed in Section 6.2.3.

6.1 Event Selection

Several stages of selection requirements must be passed before an event is accepted as a signal candidate. Firstly, the event must pass at least one of the trigger paths chosen for the analysis. Data events then have a series of 'cleaning filters' applied that remove those containing anomalous detector effects that compromise the integrity of the recorded data and that are either not present in or are impossible to include in simulation. Only then are requirements placed on the events based on the objects expected in the final state of the process being studied.

As described in Chapter 2, the tW channel (see Figure 2.3 (c)) is characterised by an on-shell top quark and a W boson. The top quark decays $\sim 100\%$ of the time to a b-quark and a W boson. Only leptonic decays of the W boson are considered, in which each W boson decays to a lepton and its associated neutrino. This means that the final state around which kinematic requirements will be developed is two oppositely charged leptons, a large amount of missing transverse energy and exactly one jet originating from a b-quark.

At leading order the tZq channel (see Figure 2.8) contains a top quark, a recoil quark (which can be of any flavour) and a radiated Z boson. As with the tW channel, only top decays to a b-quark and a leptonically decaying W boson are considered. In addition, only charged leptonic decays of the Z boson are considered. This means that the tZq final state contains three leptons, two of which must be of the same flavour but oppositely charged, and have a dilepton invariant mass consistent with the Z mass. There will also be measurable missing transverse energy, and two jets, at least one of which originates from a b-quark. The top quark originates from a b-quark in the initial state, which may be a sea quark itself or could come from from a gluon splitting. In the latter, there may be an additional b-jet present in the event.

6.1.1 Trigger and Event Cleaning

6.1.1.1 Trigger Selection

The studies presented in this thesis rely on the identification of multiple leptons in the final state, and accordingly datasets that have been identified by the trigger system (as described in Section 3.2.6) as containing two leptons are used. These 'dilepton' triggers

Final State		HLT Path(s)
tW	tZq	
ee	eee	HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*
$e\mu$	$ee\mu$ $e\mu\mu$	HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v* HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*
$\mu\mu\mu$	$\mu\mu\mu$	HLT_Mu17_Mu8_v* HLT_Mu17_TkMu8_v

TABLE 6.1: Triggers used per decay channel

are divided into three groups depending on the expected final state leptons: the electron-electron (ee), the electron-muon ($e\mu$) and muon-muon ($\mu\mu$) datasets. Table 6.1 gives the HLT trigger paths that were used in these analyses.

In these trigger path names, Mu refers to a reconstructed muon and Ele to an electron. The number that follows indicates the energy threshold of the particle. The version number of the trigger path at the end of the name, here indicated as v*, changes with the version of the trigger table being used; a different version does not indicate an intrinsically different trigger path. Due to the large numbers of energy deposits in the calorimeter system, identifying large deposits is not enough to keep the trigger rate at a usable value; additional restraints must be added to the selection requirements within the trigger logic.

A lower rate can be achieved by raising the energy threshold of the physics object in question, but this has knock-on repercussions for physics analyses that must then apply more stringent selection requirements in offline analysis. An alternative method, which is used in the electron triggers of CMS, is to adopt more complicated triggers that take isolation (Iso in the trigger path's name) and simple identification criteria (Id) into account. The HLT extracts this information from both the calorimeter, (Calo), and the tracker (Trk), by applying basic requirements on the shape of the cluster, total energy depositions and angular separation between tracker and ECAL deposits. Each Id and Iso requirement contains a Tight (T), Loose (L), or Very Loose (VL) suffix that indicates the severity of the kinematic cuts applied. For example, the first of the HLT paths used to select $ee\mu$ and $e\mu\mu$ events given in Table 6.1 requires the presence of a muon with a p_T greater than 17 GeV and an electron with p_T greater than 8 GeV. In addition the electron must have passed tight identification requirements in the calorimeter, very loose identification criteria in the tracker, and very loose isolation requirements in both the tracker and calorimeter.

The tW analysis, based around a dileptonic final state, required events to pass the relevant trigger path for the final state. This meant that only events passing the dielectron trigger

would be considered for the *ee* channel, and so forth. The publication of a timely result was considered of high importance for the tW analysis, and hence the analysis was begun before the full $\sqrt{s} = 8$ TeV dataset was available. A sub-set of the data corresponding to 12.2 fb^{-1} was initially considered with the intention of analysing the remaining data when it became available. It later became apparent that the uncertainty in the result was not dominated by a limited availability of statistics in the data sample, and so the remaining data was not added as planned.

As the tZq analysis is searching for a tri-leptonic final state, a different approach must be taken with triggers to avoid double counting between the channels. The two final states containing a combination of electrons and muons, the $ee\mu$ and $e\mu\mu$ channels, require events that pass at least one of the MuEG triggers with no further restrictions applied. Events with three same flavour leptons, the *eee* and $\mu\mu\mu$ channels, are required to pass their associated dilepton triggers, but are vetoed if they also pass the muEG or different flavour dilepton triggers. As the tZq analysis had less time constraints, the full available 8 TeV dataset was used, corresponding to an integrated luminosity of 19.7 fb^{-1}.

6.1.1.2 Filtering

After trigger selection but before further requirements are applied, several 'cleaning' filters are applied to data events to remove known anomalies from detector and accelerator effects.

Applied first are the **CSC Beam Halo Filter** and **beam scraping veto**. Despite the high performance of the beam focusing in the LHC, it is unavoidable that particles will migrate radially from the nominal bunches, causing a 'beam halo' to circulate with the beams. It was found during early data-taking that these halo particles could be picked up inside the detector and mistakenly added to the "real" event. As the muon detectors are the most sensitive part of the detector to beam halo effects, a filter based around muon tracking kinematics is used to remove troublesome events.

In order to reduce the number of beam halo particles originating from the LHC bunches, collimator blocks are placed around the accelerator. Although this helps to clean the beam, particles from the beam halo that collide with the collimator blocks can cause particle showers that go on to interact with the detectors. These 'beam scraping' events

are removed by requiring that at least 25% of tracks reconstructed in the inner detector pass the high purity threshold as described in Section 5.2.1.

The **HCAL noise filter** is applied to remove events with anomalous noise in the hadronic calorimeter. A certain amount of noise from the electronics is expected[1], but anomalous noise in the HCAL is found to originate in the hybrid PhotoTridoes (HPTs) and read-out boxes (RBXs). Although there are various sources for this noise, they all manifest as large, isolated energy deposits up to the TeV scale. Noise events are most efficiently identified by the pulse shape and isolation of the HCAL readout, and multiplicity in the RBXs. Events are rejected if they have very high multiplicities in the RBXs, or if the pulse shape exhibits little development over time.

During 2011 data taking, events were observed in which the HCAL calibration laser fired during collision bunch crossings. The events are identified as containing over 5000 reconstructed hits in the HCAL; a number far higher than expected during nominal running. This resulted in the creation and application of a **HCAL laser filter** during 2012 data taking.

The ECAL contains a number of noisy crystals that are masked during event reconstruction, and several noisy towers with dead data links. Although these constitute <1% of the detector, if significant energy falls within one of these regions it will result in the affected event being recorded as having large missing E_T. Several methods are used to minimise the impact of these events: energy deposition information read by the trigger for affected regions is often still valid, and can provide a flag if large amounts of energy have been deposited in otherwise dead regions. An event may be rejected if a dead region has a small angular separation from a reconstructed jet. Finally, an event may be vetoed if large amounts of energy is found in the cells bordering any of the masked crystals. These requirements are combined in the **ECAL dead cell filter**, which is applied to the data.

It has also been found that two specific supercrystals in the ECAL occasionally give anomalously high energy readings. The **ECAL bad supercrystal filter** removes events in data that have a total supercrystal energy greater than 1 TeV with few associated 'good' reconstructed hits.

As with the HCAL, the laser calibration system in the ECAL has also been the source of unphysical events, although the cause, in this case, is different. As the ECAL crystals are

[1]This is also referred to as the 'pedestal' noise.

expected to lose transparency over time, laser calibration is used to provide a scaling factor for each crystal to correct for the loss. Occasionally, very large, unphysical values for this scale factor are calculated making the crystals appear highly energetic. The **ECAL laser correction filter** therefore rejects in data events in which this scale factor is over 3.0 in the barrel, and over 8.0 in the endcap.

Events with normal calorimeter performance but with little to no tracks have been observed in CMS. These events were found to be caused by two separate phenomena: the tracking algorithms automatically abort when there is a very large occupancy in the tracker to avoid excess CPU and memory usage. This results in a lack of tracks in an otherwise reconstructed event. The second class of these events comes from the interaction of satellite RF buckets when no hard collision in the nominal bunch has occured. Both types of event can be removed by placing a requirement on a single variable; the sum of the p_T of the tracks belonging to good vertices divided by the H_T - the scalar sum of the transverse momentum - of all the jets in the event. This quantity must exceed a threshold of 10% to indicate a real and good quality event. This requirement is applied by the **tracking failure filter**.

Finally, events are occasionally observed with large coherent noise throughout the detector. Although the trigger usually rejects these events, a small number are read off from the detector and reconstructed. The noise in the tracker creates a large number of fake tracks, usually only in the strip tracker. These are removed using the **tracking odd event filter**.

Once these filters have been applied the remaining events are considered good events for use in the analysis.

6.1.2 Lepton Selection and Vetoes

In both analyses, lepton candidates from the list of PF reconstructed objects are considered and must pass several additional criteria to be selected for further use in the analysis. In addition to these selection requirements, which are referred to as the standard selection criteria, a set of less stringent requirements are applied to the PF objects to identify any 'loose' leptons within any given event. Events with additional loose leptons that do not pass the standard selection are vetoed from being possible signal events.

Although the leptons are selected using the same set of criteria, the two analyses' differing final states mean that separate requirements on the number of leptons and the charge configuration must be implemented. In the tW analysis, exactly two oppositely-charged leptons are required. For the tZq analysis, two oppositely charged, same flavour leptons, consistent with a leptonic Z decay, are required. A third lepton, originating from the leptonic decay of the W boson coming from the decaying top quark, is also selected, but has no further charge or flavour requirements applied.

6.1.2.1 Electrons

Electron candidates identified by the PF algorithm are only considered further if they have been identified using the GSF scheme (see Section 5.2.5). They must then pass the following additional selection criteria:

- the transverse momentum must satisfy $p_T > 20$ GeV;
- the pseudorapidity must be within the range $|\eta| < 2.5$;
- the distance in z from the primary vertex identified in the event must be less than 1.0 cm;
- the transverse impact parameter of the electron with respect to the beam spot (the luminous region produced by the collision of the two beams) must be less than 0.04 cm.

Electrons also undergo an additional identification step, using a multivariate analysis technique (MVA) to combine various properties of an electron candidate into a single discriminating value. The value of the discriminant ranges between -1 and +1; the greater the number the more electron-like the candidate is. Properties of the candidate used in the MVA include kinematic variables such as the η and ϕ positions of the track and associated superclusters, and other quality information, such as the presence of pre-shower hits and χ^2 of the GSF fits. Different versions of the discriminant exist depending on whether the HLT path requires the presence of an electron or not. For the analyses presented in this thesis, an electron passes the MVA identification if its triggering discriminant is between 0.5 and 1.0.

The relative isolation of a particle is calculated, in general, by summing the p_T of PF reconstructed objects within a cone of fixed radius surrounding the lepton and dividing by the lepton's p_T . If the sum of the energy in this cone is small, the lepton is considered isolated. In the presence of pileup interactions, the isolation requires a correction to account for the additional energy deposited in the detector. Charged hadrons have associated tracks with them, and can be removed from the isolation calculation if they do not originate from the same primary vertex as the electron. The effect of neutral hadrons and photons originating from pileup events is accounted for by subtracting an average energy over the effective area of the electron, which is extended in the ECAL as a result of Bremsstrahlung radiation. The adjusted formula used is:

$$I_\rho = \frac{I_{chargedHadron} + max(I_{neutralHadron} + I_\gamma - \rho \cdot Area_{electron}, 0.)}{p_T} \tag{6.1}$$

where $I_{chargedHadron}$ $I_{neutralHadron}$ and I_γ are the energy deposits within a cone of fixed radius of charged hadrons, neutral hadrons and photons, respectively, ρ is the energy density of the event and $Area_{electron}$ is the effective area of the electron, calculated from the supercluster η and reconstructed electron p_T . For an electron to be considered isolated, it must have a value of $I_\rho < 0.15$ for a cone $\Delta R = 0.3$.

Electrons originating from photon conversions within the tracker represent a significant source of fake electrons. Two techniques are used to distinguish these from prompt electrons created in a hard collision: identifying missing hits in the tracker volume and secondary track association. The first technique identifies any layers within the tracker that are missing hits in the electron's associated track. If there are any empty layers the electron is identified as a conversion electron and is not considered further. The second technique searches for a second track such that the two are compatible with a photon conversion to an electron-positron pair. The electron is not considered further if a second track is identified within 0.02cm in the $r - \phi$ plane and if the cotangent of the polar angle of each track differs by less than 0.02.

Loose electrons are identified from GSF electrons using the same pseudorapidity, MVA ID and isolation requirements as the standard electron selection but require a less stringent requirement on the transverse momentum: p_T >10 GeV. The loose selection places no requirement on the distance from the interaction point, either along the z axis or transversely, and no photon conversion veto is applied.

6.1.2.2 Muons

Muons that have undergone PF reconstruction have already been identified as both Global and Tracker muons. In addition, muon candidates are required to pass the following criteria:

- the transverse momentum, p_T, must exceed 20 GeV;
- the pseudorapidity must be within the range $|\eta| < 2.4$;
- the transverse impact parameter with respect to the beam spot must be less than 0.02 cm;
- the distance in z between the primary vertex and the muon vertex must be less than 0.5 cm;
- the normalised χ^2 of the muon track fit must be less than 10.

The pileup correction for muon isolation is simpler than for electrons, and involves removing the neutral hadron and photon contribution from the fixed radius cone. This is referred to as the $\Delta\beta$ correction, and is calculated using the formula:

$$I_{\Delta\beta} = \frac{I_{chargedHadron} + max(I_{neutralHadron} + I_{\gamma} - 0.5 \cdot I_{pileup}, 0.)}{p_T} \tag{6.2}$$

where I_{pileup} is the neutral energy deposition within the cone of fixed radius. The 0.5 in this formula is a crude averaging of neutral to charged particles in the pileup interactions. For a muon to be considered isolated,it must have a value of $I_{\Delta\beta} < 0.2$ using a cone of $\Delta R = 0.4$.

Muons originating from decays in flight are observed to be more common in data than expected from simulation, resulting in an excess in the number of fake muons. In order to differentiate between these and prompt muons several additional identification criteria are applied. At least one hit is required in both the pixel and muon tracking detectors. There must be at least 6 hits found in the inner tracker, with at least two matched stations in the outer muon system.

Loose muons are selected from the PF muon list using the same pseudorapidity and isolation parameters as the standard muon selection, but with a less stringent $p_T > 10$ GeV requirement. No additional identification requirements are applied.

6.1.3 Lepton Invariant Mass Selection

6.1.3.1 tW Analysis

In order to reject low mass Z+jets and multijet backgrounds, events where the invariant mass of the lepton pair, m_{ll}, is less than 20 GeV/c^2 are rejected. To suppress the background from Z boson decays, a further dilepton invariant mass requirements is placed on the same flavour lepton channels (ee and $\mu\mu$). Events where the dilepton invariant mass falls within $81 < m_{ll} < 101$ GeV/c^2 (corresponding to the Z mass) are vetoed. Although this selection reduces the background contribution from events containing a Z decay by design, it also helps to reduce the contribution from ZZ and WZ diboson backgrounds that also peak in this region.

Due to the increased level of pileup in the 2012 dataset, the missing transverse energy resolution is degraded compared to that observed 2011, particularly at low values of measured $\not{E}_T$. This leads to poor agreement between data and simulation when the expected $\not{E}_T$ of the process is small. This is a problem for the simulated Z+jets samples, for which mis-identified jets are the only source of possible $\not{E}_T$, the discrepancy for which can be clearly seen in Figure 6.1. The events that were rejected because the dilepton invariant mass was within the Z mass window are used as a Z+jets enriched control region to study this effect. The $\not{E}_T$ distribution for the control region is compared between data and simulation; it is also used to calculate correction factors to be applied to the Z+jet simulated data in the signal region. The uncertainty associated with this process is considered as a systematic uncertainty, as described in Section 7.2.4. As no requirement on the dilepton invariant mass is necessary in the $e\mu$ final state, the scale factors cannot be calculated directly for this channel; an average of the scale factors in the ee and $\mu\mu$ channels is used in this case. Table 6.2 shows the scale factors per channel, as a function of the $\not{E}_T$ of the event. Figure 6.2 shows the distribution of the $\not{E}_T$ in the ee and $e\mu$ final states after the reweighting has been applied.

Although a large proportion of the Z+jets background is removed with the m_{ll} veto, the high production cross section means that there is still a significant contribution of Z decay events in the ee and $\mu\mu$ samples at this stage. Z+jet events, unlike tW events, which contain two neutrinos, are not expected to exhibit a large missing transverse energy. For

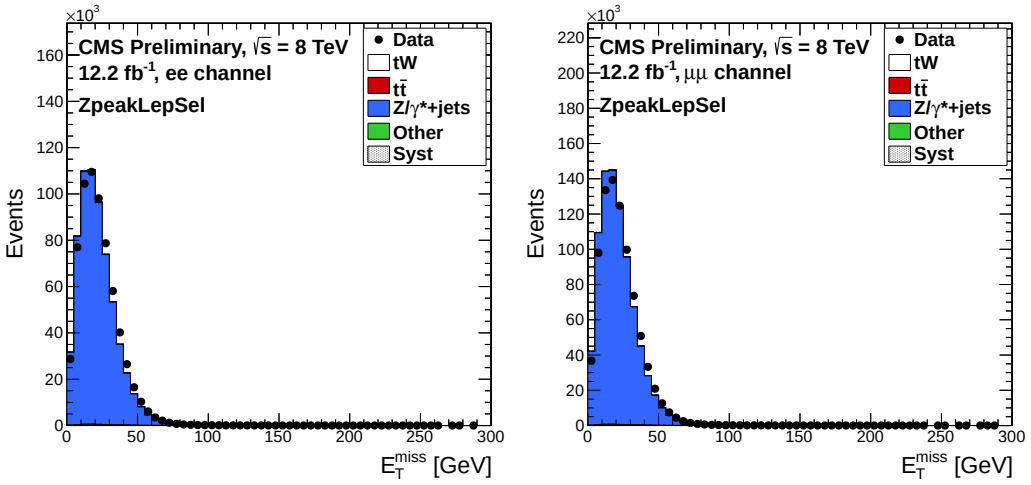


FIGURE 6.1: Distribution of $\not\!\!E_T$ in data and simulation in the ee (left) and $\mu\mu$ (right) final states in the control region selected by requiring that the dilepton invariant mass fall within the Z mass window.

TABLE 6.2: Scale factors for Z+jets simulated samples as a function of $\not\!\!E_T$ in the three final states.

$\not\!\!E_T$	$\mu\mu$ Scale Factor	ee Scale Factor	$e\mu$ Scale Factor
$<$ 10 GeV	0.8841	0.9215	0.9028
10 to 20 GeV	0.9386	0.9608	0.9497
20 to 30 GeV	1.0131	1.0247	1.0189
30 to 40 GeV	1.1012	1.0964	1.0988
40 to 50 GeV	1.1850	1.1633	1.17415
50 to 60 GeV	1.2500	1.2529	1.25145
$>$ 60 GeV	1.3071	1.2194	1.26325

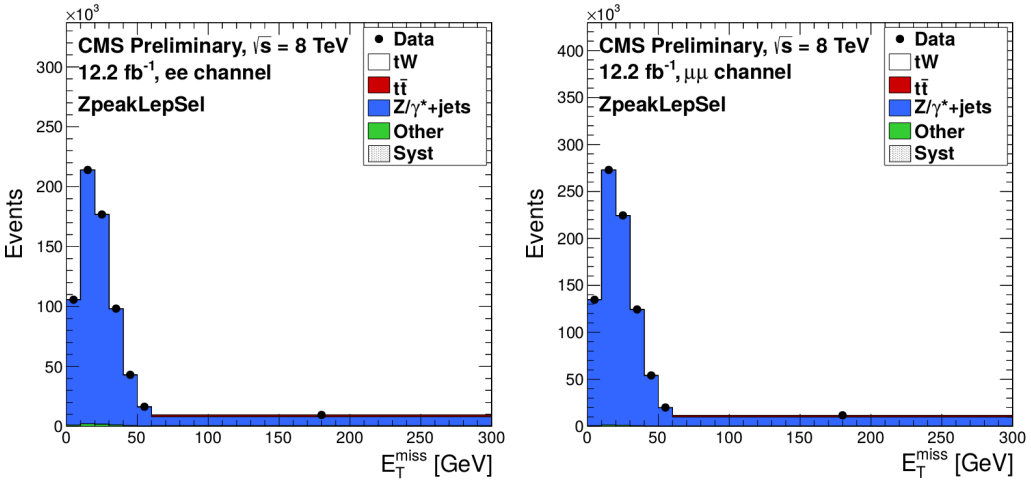


FIGURE 6.2: Distribution of $\not\!\!E_T$ in data and simulation in the ee (left) and $\mu\mu$ (right) final states after the reweighting has been applied.

this reason an additional requirement is placed on the $\not{E}_T$ in the ee and $\mu\mu$ channels, vetoing events where $\not{E}_T < 30$ GeV/c^2.

Unlike the same-flavour final states, the majority of Z+jets contamination in the $e\mu$ final state originates from $Z \to \tau\tau$ events. To remove as much of this contribution as possible, the H_T kinematic variable is defined. H_T is defined, for the purpose of the tW analysis, as the scalar sum of the transverse momenta of the leptons, jet and $\not{E}_T$ in the event and is required to be greater than 160 GeV for an event to be considered signal in the $e\mu$ final state. This variable exploits the different topology of the Z+jets final state, namely the lack of missing transverse energy, without having to place a requirement directly on the $\not{E}_T$ itself, which would remove a greater fraction of the total signal than background. The distribution of H_T for each final state before the selection is applied can be seen in Figure 6.3.

6.1.3.2 tZq Analysis

In the tZq analysis, the selection requirements are changed to reflect the presence of a real Z boson in the final state. Two leptons that are consistent with a Z boson decay must be selected: the leptons must be of the same flavour, but opposite charge, with an invariant mass within the Z mass window $76 < m_{ll} < 106$ GeV. If two lepton pairs satisfy these criteria, the one with the dilepton invariant mass closest to the Z mass of 91 GeV is chosen as the Z candidate. An additional requirement on the angular separation between the lepton pair was considered, such that $\Delta R > 0.5$ but it was found that the isolation requirement on the leptons effectively made this redundant.

6.1.4 Jet Selection and b-tag Requirements

Jets are reconstructed using the anti-k_T algorithm and a jet size parameter, R, of 5 in the PF reconstruction scheme. The jet energy corrections, as described in Section 5.4.1.1, are applied after the jets have been identified. Jets are only considered in the analysis if they satisfy the following criteria: $p_T > 30$ GeV and $|\eta| < 2.4$.

Lepton candidates that fall within a cone of radius $\Delta R = 0.3$ of a jet in the $\eta - \phi$ plane are considered to be part of that jet and not a standalone lepton. Additional jet identification criteria have been developed within CMS to check the quality of the reconstructed jets. By

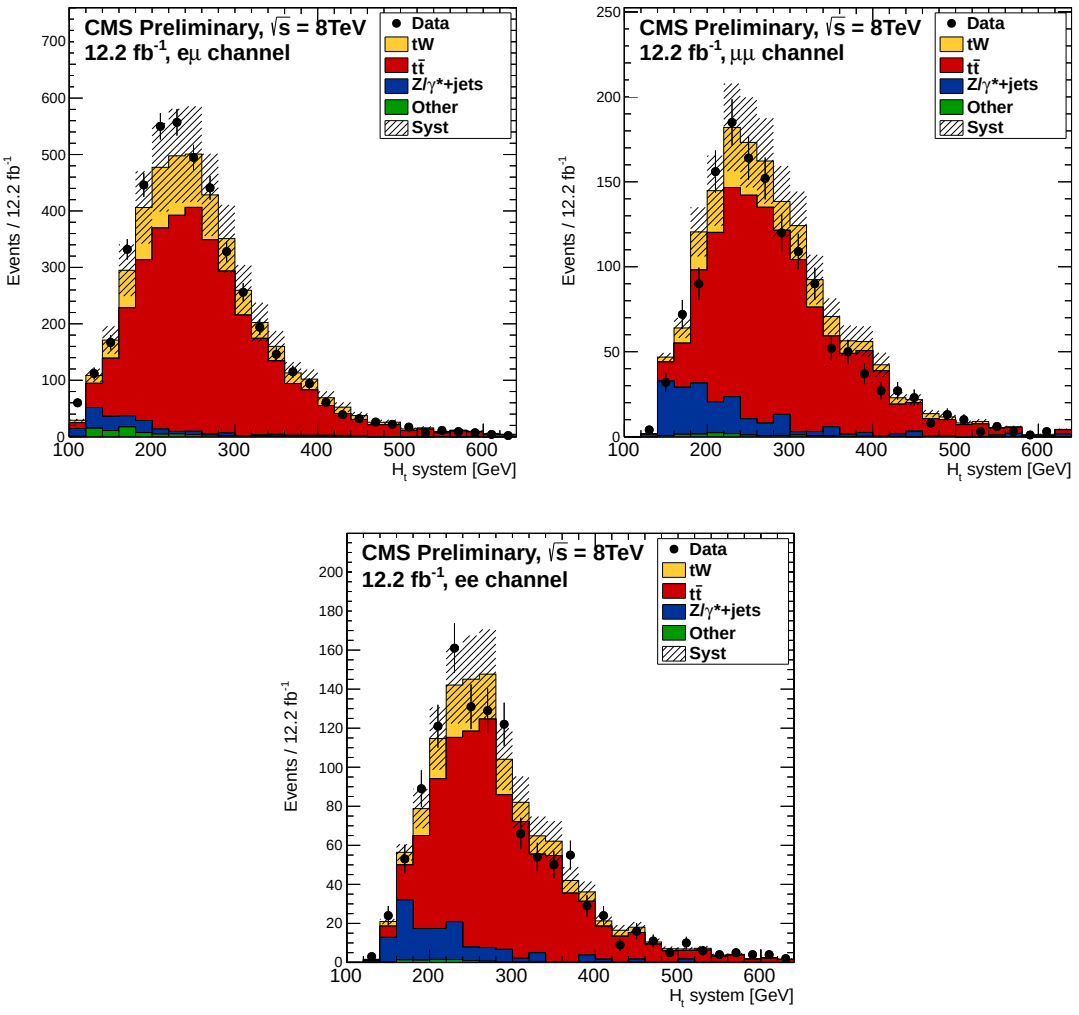


FIGURE 6.3: H_T distributions before the selection in the $e\mu$ (left), $\mu\mu$ (centre), and ee (right) final states.

requiring that the jet comes from multiple energy deposits across the ECAL and HCAL sub-detectors, jets originating from anomalous deposits in a single detector can be removed from the sample [116]. Each jet must pass the following criteria:

- It must have a charged particle multiplicity of more than 0;
- It must have been constructed from more than one PF object;
- the fraction of the jet energy deposited in the ECAL by charged and neutral electromagnetic particles must be less than 0.99;

TABLE 6.3: Event yields after selection in the tW signal region.

Channel	ee	$e\mu$	$\mu\mu$	Combined
tW	192 ± 14	798 ± 28	250 ± 16	1239 ± 35
$t\bar{t}$	859 ± 29	3164 ± 56	1126 ± 34	5150 ± 72
Z + Jets	151 ± 12	98 ± 10	217 ± 15	465 ± 22
Other	10 ± 3	41 ± 6	13 ± 4	64 ± 8
Background	1020 ± 32	3303 ± 58	1356 ± 37	5679 ± 75
Data	1198 ± 35	4201 ± 65	1443 ± 38	6842 ± 83
Sum All MC	1212 ± 35	4101 ± 64	1605 ± 40	6918 ± 83

- the fraction of the jet energy deposited in the HCAL by neutral hadrons must be less than 0.99. The corresponding fraction for charged hadrons must be greater than 0.

The b-tagging discriminant is then determined for each jet to decide whether it should be considered a b-jet. For both analyses, the Combined Secondary Vertex tagging algorithm at the medium working point is used. This corresponds to a b-tagging efficiency of $62 \pm 2\%$ and a misidentification probability of $1.51 \pm 0.02\%$.

In the search for tW production, each event must contain exactly one b-tagged jet. In order to reduce contributions from the dominant $t\bar{t}$ background, events that contain additional loose jets passing the above b-tagging criterion are rejected. These are defined in the same way as the standard jet selection, but with a less stringent requirement on the transverse momentum, p_T >20 GeV.

In the search for tZq production, either two or three jets must be present in each event, at least one of which is b-tagged. Although no further requirements are placed on b-tag multiplicities, in practice it is found that only events with one or two b-tagged jets pass the selection. These correspond to the signal regions where the recoil quark is a light quark or a b-quark, respectively.

6.2 Background Estimation

For both analyses, the first attempt to estimate the number of expected events in both background and signal was made using simulated samples. The event yields for the tW analysis can be seen in Table 6.3, whilst the tZq yields are presented in Table 6.4.

TABLE 6.4: Event yields after event selection requirements have been applied for the tZq search.

Channel	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	Combined
tZq	1.7±0.0	1.7±0.0	1.8±0.0	2.1±0.0	7.2±0.0
$t\bar{t}Z$	2.3±0.2	2.7±0.2	2.4±0.2	3.1±0.2	10.6±0.4
WZ	2.9±0.2	5.6±0.3	3.9±0.2	4.0±0.2	16.4±0.5
Z+jets	8.9±3.0	4.5±2.0	4.2±1.8	5.1±2.3	22.7±4.6
$t\bar{t}W$	0.2±0.1	0.3±0.1	0.4±0.1	0.4±0.1	1.3±0.2
ZZ	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.2±0.0
Signal	1.7±0.0	1.7±0.0	1.8±0.0	2.1±0.0	7.2±0.0
Background	14.4±3.0	13.2±2.1	11.0±1.8	12.7±2.3	51.2±4.7
Data	11.0±3.3	13.0±3.6	24.0±4.9	15.0±3.9	63.0±7.9
Total MC	16.0±3.0	14.9±2.1	12.8±1.8	14.8±2.3	58.5±4.7

6.2.1 tW Analysis

After the full event selection has been applied, two processes are left as the dominant backgrounds in the tW search across all three signal channels: Z+jets and $t\bar{t}$ production. The Z+jets contribution is the smaller of the two: the expected yields are similar to those of the signal in the ee and $\mu\mu$ channels, but represent a smaller fraction of the background in the $e\mu$ channel. Although every effort is made to remove this background through $\not{E}_T$ and invariant mass requirements in the ee and $\mu\mu$ channels, the high production cross section and large $\not{E}_T$ tail in Z+jets production means that a significant number of events pass the selection requirements regardless. Although a more stringent requirement could be placed on the $\not{E}_T$ of prospective events to remove a larger proportion of the Z+jets background, this would start removing too much signal to be viable. The chosen limit of 30 GeV is, therefore, a compromise between background rejection and signal efficiency. As the $e\mu$ final state represents the most sensitive channel in the tW search, and contains relatively little Z+jets background, the level is considered acceptable for the rest of the analysis.

The main background source for the the tW analysis is, then, $t\bar{t}$ production. This happens for two reasons: the first is that the Standard Model cross section, 245.8 pb, is over ten times that of tW, where the cross section is only 22.2 pb. The second, more important factor, is that tW and $t\bar{t}$ production are topologically very similar. As explained in Section 2.2.2, the next-to-leading-order (NLO) diagrams for the two processes interfere, making a

TABLE 6.5: Event yields after selection in the tW two jet one tag control region.

Channel	ee	$e\mu$	$\mu\mu$	Combined
tW	118 ± 11	474 ± 22	157 ± 13	749 ± 27
$t\bar{t}$	1934 ± 44	7544 ± 87	2465 ± 50	11943 ± 109
Z + Jets	109 ± 10	73 ± 9	128 ± 11	310 ± 18
Other	7 ± 3	28 ± 5	8 ± 3	42 ± 7
Background	2050 ± 45	7645 ± 87	2602 ± 51	12296 ± 111
Data	2139 ± 46	7501 ± 87	2418 ± 49	12058 ± 110
Sum All MC	2167 ± 47	8118 ± 90	2759 ± 53	13045 ± 114

precise definition of the tW signal at higher orders difficult. The LO signals are also very similar, the only difference being a single, b-tagged jet originating from the second top quark in $t\bar{t}$ events. This means that the kinematic cuts used to isolate tW also strongly favour $t\bar{t}$ selection, resulting in a signal region populated by over four times more $t\bar{t}$ events than signal events.

In order to reduce the dependency of the analysis on the details of the simulation of $t\bar{t}$ events, additional 'control' regions are defined. These regions are chosen to be topologically similar but orthogonal to the tW signal, enriched in $t\bar{t}$ events but lacking in signal events. The two chosen regions, produced in both simulation and data, contain exactly two jets, with either one (2j1t) or both (2j2t) of the jets b-tagged. The other selection requirements are left unchanged. The 2j1t region is composed of ∼92% $t\bar{t}$ events, with ∼5% contamination from tW signal events, whilst the 2j2t region is ∼97% $t\bar{t}$ events with slightly less than 3% signal events. This compares to the signal (1j1t) region, which contains ∼18% tW signal events.

All three regions are defined for each of the dilepton final state, as well as for the combination of the three channels. The control regions are used to constrain the $t\bar{t}$ background by including them in the statistical fit, described in Chapter 8, simultaneously with the signal region. The background contribution from Z+jet events is estimated from simulation only. The final event yields for data and simulation for the signal and control regions, for each final state, are presented in Tables 6.5 and 6.6, and can be seen visually in Figure 6.4.

The contributions from all other background processes, including multijet and diboson production, is found to be very small compared to the signal and the contributions from the dominant backgrounds: they make up less than 1% of the final event yields. It was therefore decided simulation was sufficient to estimate these backgrounds.

TABLE 6.6: Event yields after selection in the tW two jet two tag control region.

Channel	ee	$e\mu$	$\mu\mu$	Combined
tW	36 ± 6	126 ± 11	45 ± 7	207 ± 14
$t\bar{t}$	1092 ± 33	4502 ± 67	1439 ± 38	7032 ± 84
Z + Jets	11 ± 3	4 ± 2	17 ± 4	31 ± 6
Other	1 ± 1	4 ± 2	2 ± 1	7 ± 3
Background	1103 ± 33	4510 ± 67	1457 ± 38	7070 ± 84
Data	1163 ± 34	4269 ± 65	1392 ± 37	6824 ± 82
Sum All MC	1139 ± 34	4636 ± 68	1502 ± 39	7277 ± 85

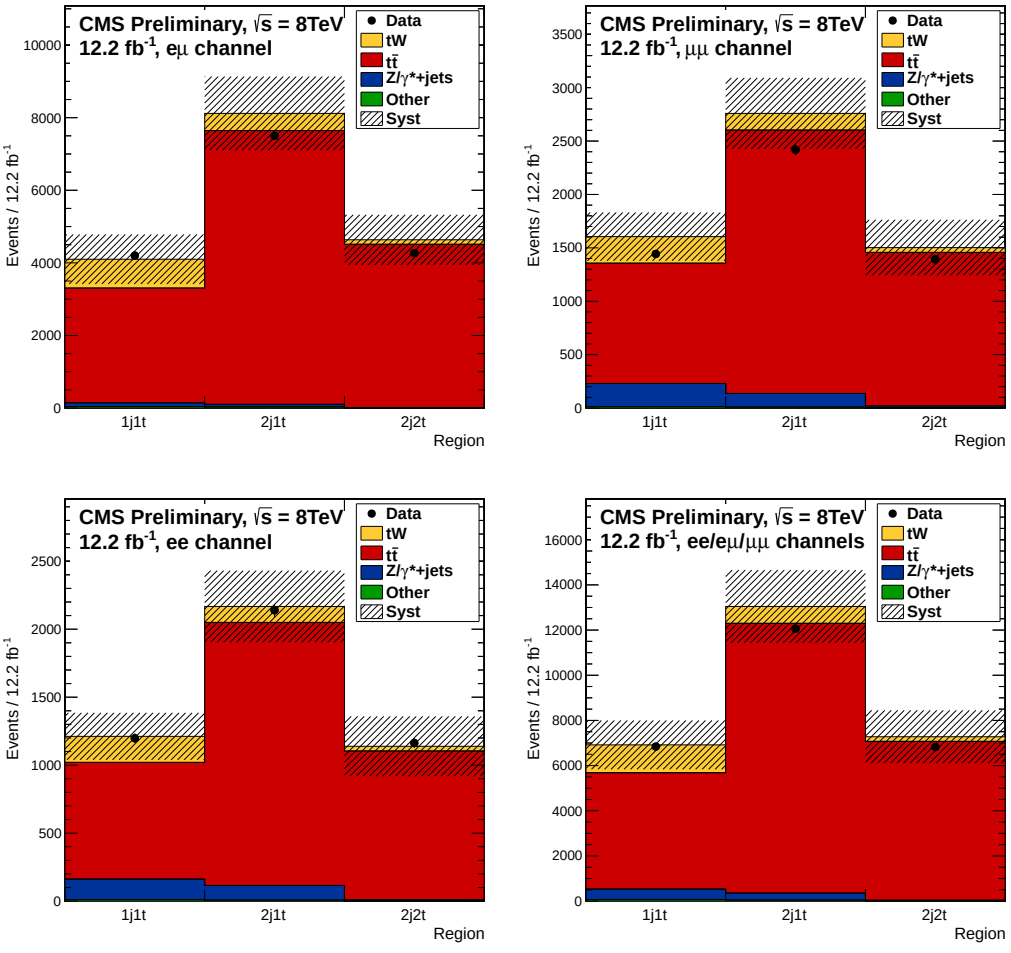


FIGURE 6.4: Event counts for signal and control regions with systematic uncertainties. From top-left to bottom-right, the plots show the $e\mu$, $\mu\mu$, ee and combined final states.

6.2.2 tZq Analysis

Similarly to the tW analysis, a first attempt at estimating the impact of background processes on the tZq signal region was carried out using simulated samples of the processes expected to contribute. Table 6.4 shows the event yields of the different processes after all event selection requirements have been applied. It is found that the dominant backgrounds come from two main sources: those containing a third lepton that comes from a genuine W decay, i.e. WZ diboson and $t\bar{t}Z$ production, and those whose third lepton is misidentified as the lepton from the W decay, that is Z+jets and $t\bar{t}$ production.

Whilst this approach is viable for the majority of background processes, it was found that the limited statistics of the Z+jets sample in the three lepton region results in an inaccurate background estimate. Thus, simulation cannot be used to determine the Z+jets background contribution. Of the ∼30,000,000 events in the Z+jets samples, only $O(1)$ event passes the full event selection for each channel; the impact this has can be clearly seen in Figure 6.5. A data-driven estimate for the Z+jets background was therefore implemented for this analysis.

A topological region similar to the signal region but enriched in Z+jets events is required. The selected control region for this analysis inverts the isolation of the lepton originating from the W decay, creating a sample enriched in the fake leptons of the Z+jets that pass the signal selection requirements. As an example of the distributions obtained in this region, the $\not{E}_T$ is shown in Figure 6.6.

Although this sample was originally conceived as an estimate for the Z+jets background, there is also a sizeable contribution from $t\bar{t}$ production. Having compared $t\bar{t}$ and Z+jets contributions in the signal and background-enriched samples, it was found that the ratio of the two was similar in both. It was decided that the background enriched sample should be used to model both Z+jets and $t\bar{t}$ contributions as a single fake lepton enriched sample. To avoid double counting simulated $t\bar{t}$ events are no longer considered in the signal sample when using the data-driven background estimate.

Although the data-driven background sample is very similar to the signal sample, after investigating different distributions it was discovered that a bias was introduced in the Z p_T spectrum: it was found to peak around 30 GeV higher in the background enriched sample than for the signal sample. This effect can be seen in Figures 6.7 and 6.8. Data

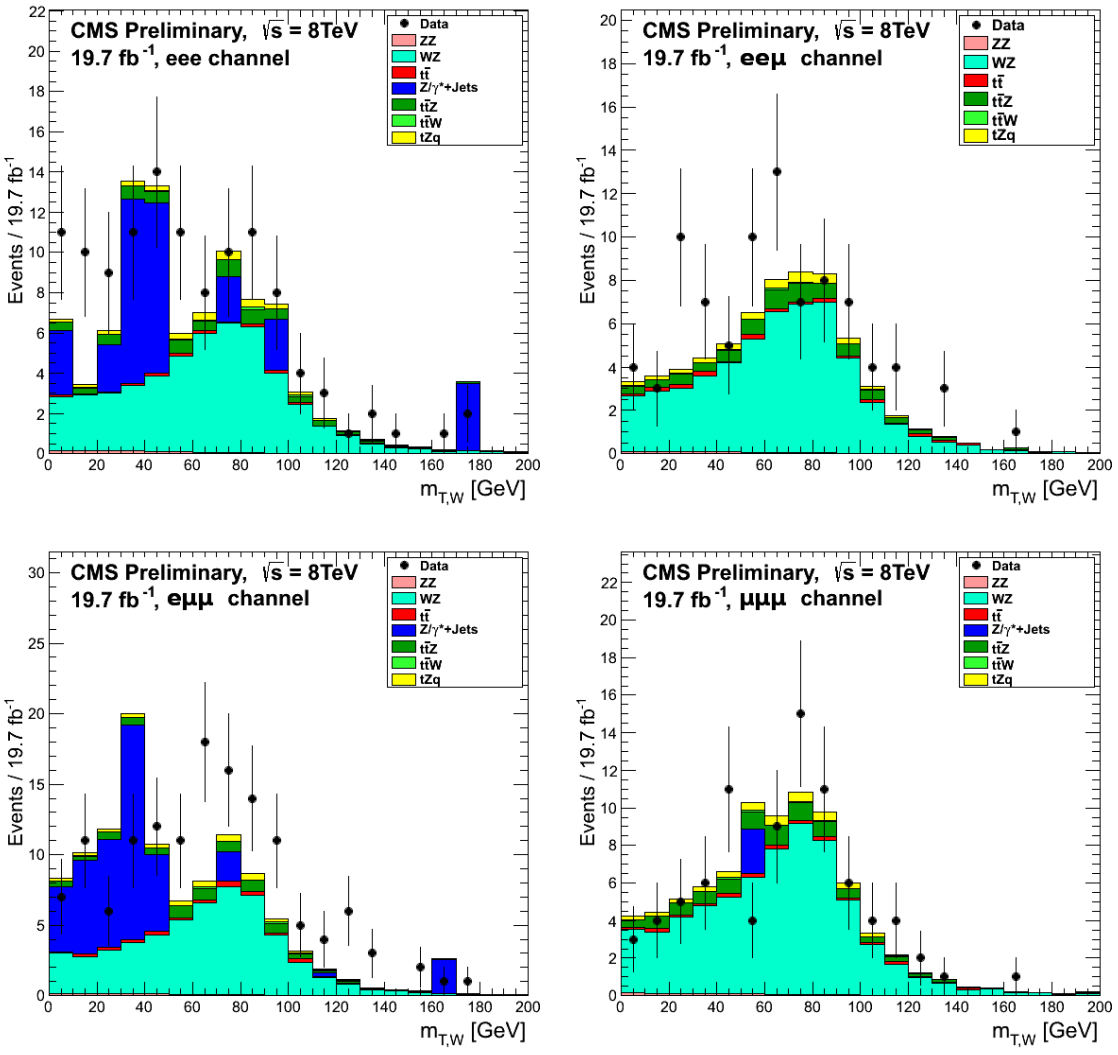


FIGURE 6.5: The transverse mass of the reconstructed W boson after jet selection requirements have been applied.

events selected in this region are reweighted before use at later stages of the analysis by scale factors determined by comparing the Z p_T distributions between the signal and background-enriched regions. The event weight is modified by the formula:

$$W = e^{(x_1 + x_2 \cdot p_{T,Z})} + x_3 \tag{6.3}$$

where the fit parameters x_1, x_2 and x_3 are channel specific and given in Table 6.7, and $p_{T,Z}$ is the p_T of the reconstructed Z boson.

Whilst the data-driven method can be used to determine the shape of distributions arising from fake lepton backgrounds, it does not give any estimation of the normalisation. One

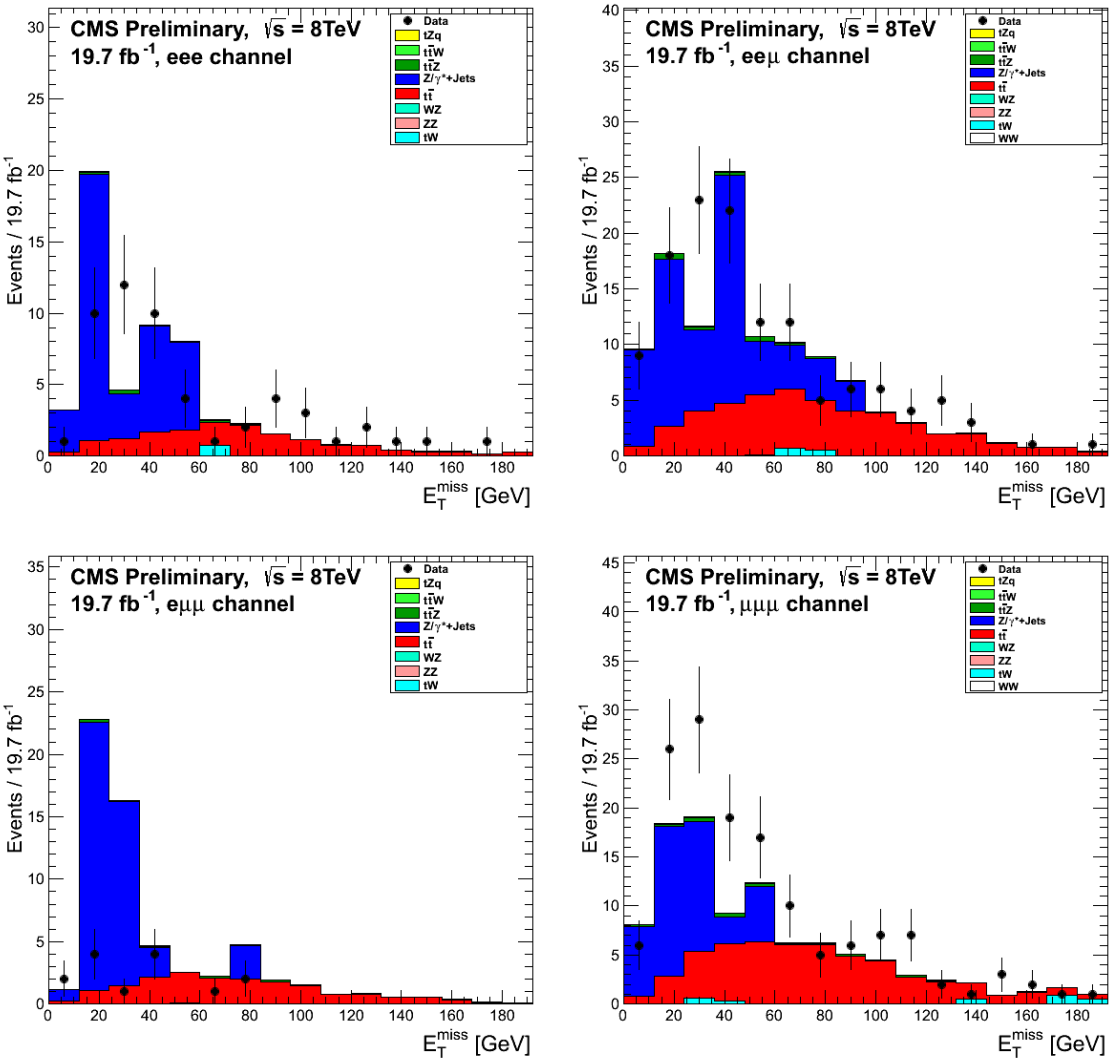


FIGURE 6.6: The $\not\!\!E_T$ distribution after full event selection in the Z+jets enriched sample (described in the text), for each of the three lepton final states. At higher values of the $\not\!\!E_T$, the control region is dominated by $t\bar{t}$ events, with the Z+jets events peaking much closer to 0.

TABLE 6.7: Parameters used in reweighting the data-driven fake lepton background according to equation 6.3, by channel.

Channel	x_1	x_2	x_3
eee	0.823	-0.0174	0.164
$ee\mu$	1.037	-0.0212	0.158
$e\mu\mu$	0.588	-0.0096	-0.032
$\mu\mu\mu$	0.912	-0.0213	0.233

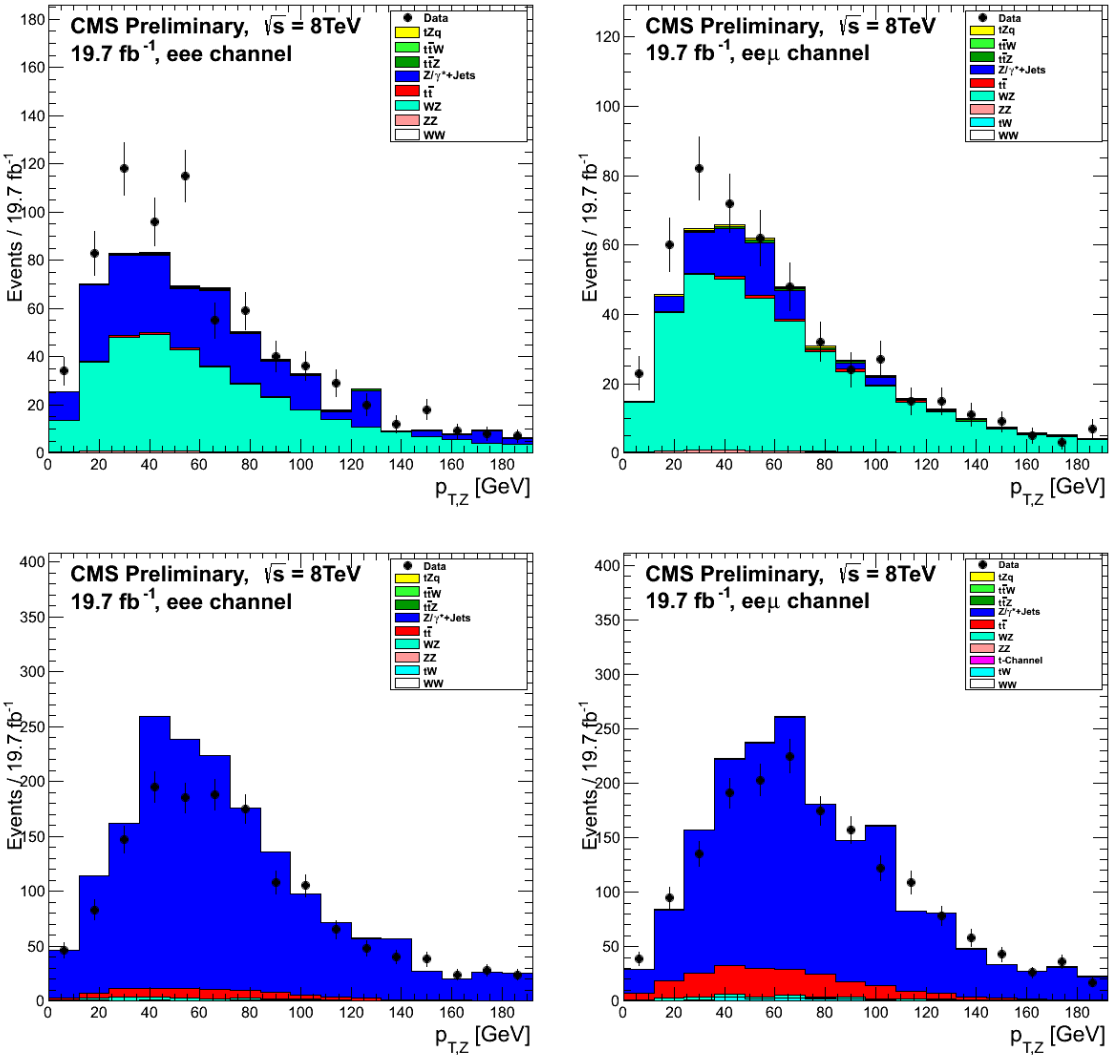


FIGURE 6.7: Reconstructed Z p_T distributions in the eee (left) and $ee\mu$ (right) final states, in the signal (top) and background-enriched (bottom) samples, after lepton selection requirements but before jet requirements. There is a bias introduced into the distribution by inverting the isolation cut on the third lepton, and this must be accounted for when using the selected data as a background estimate in the signal region.

method that was investigated to calculate this number was a profile likelihood fit of a variable recognised as having discriminating power between real and fake lepton backgrounds. The distribution chosen was the transverse mass of the reconstructed W in the event, given by the formula:

$$m_{T,W} = \sqrt{2 \cdot \not{E}_T \cdot p_{T,l} \left(1 - \cos(\phi_{\not{E}_T} - \phi_l)\right)} \tag{6.4}$$

where $p_{T,l}$ is the transverse momentum of the lepton originating from the W decay, and $\phi_{\not{E}_T}$ and ϕ_l are the polar angles of the $\not{E}_T$ and lepton originating from the W decay, respectively. Backgrounds containing a genuine W decay, such as WZ diboson events and $t\bar{t}Z$, are found to have a peak in the $m_{T,W}$ distribution at the W mass of 80 GeV. Z+jets

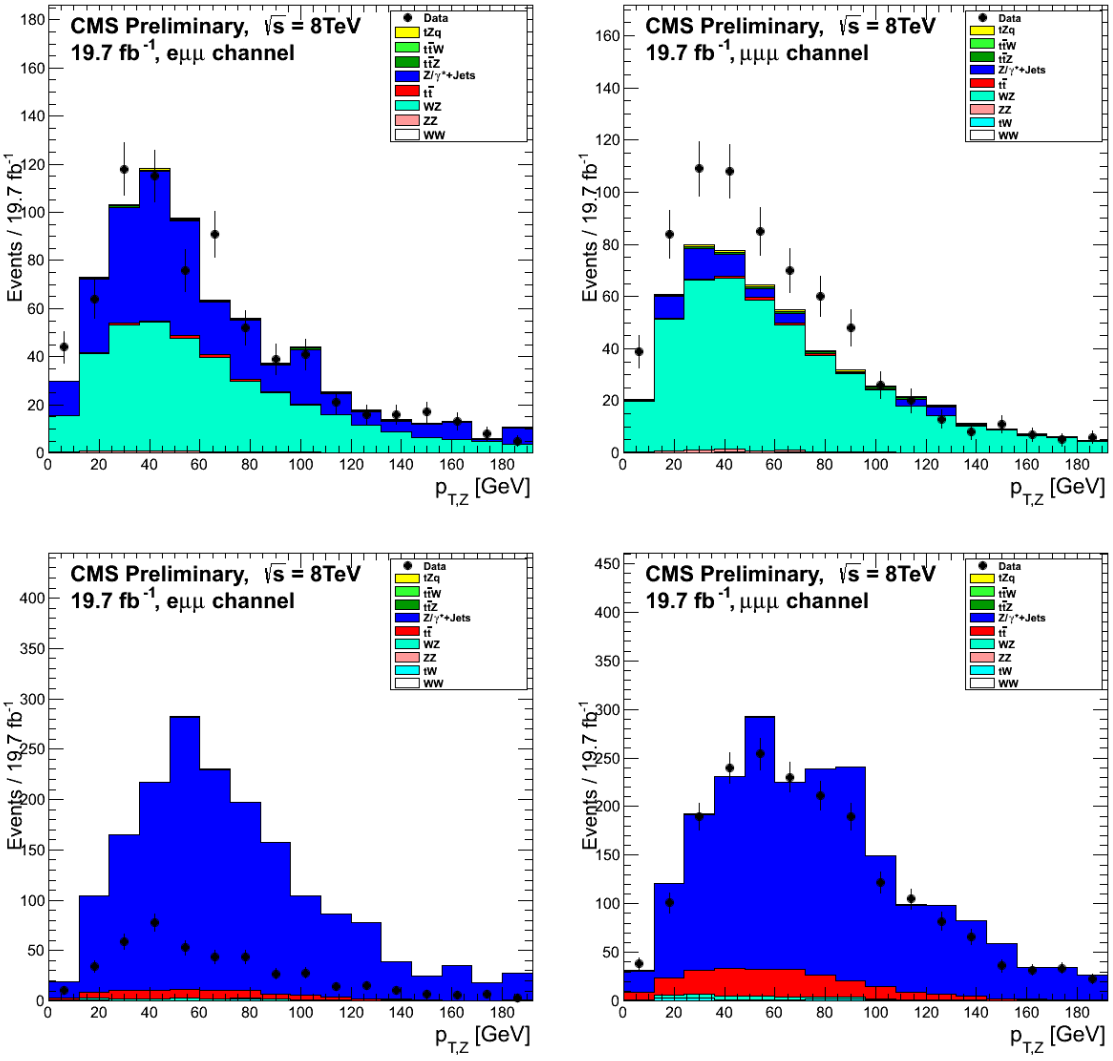


FIGURE 6.8: Reconstructed Z p_T distributions in the $e\mu\mu$ (left) and $\mu\mu\mu$ (right) final states, in the signal (top) and background-enriched (bottom) samples, after lepton selection requirements but before jet requirements.

events, on the other hand, contain no real W decay, and therefore a falling distribution peaking at 0 is observed. Fitting this distribution allowed a weight to be applied to the data-driven fake lepton background, and also allowed a scale factor to be inferred for the WZ background contribution, to ensure a more accurate representation of the data. A single scale factor was determined per channel for both the WZ and fake lepton backgrounds. The fit, the results of which can be seen per channel in Figure 6.9, was carried out after the selection of three good leptons and the identification of a Z candidate, but with no jet requirements applied. The increased statistics at this stage of the selection allowed a more accurate fit with lower uncertainties in the calculated scale factors. The fit was checked for consistency after jet requirements are applied but before b-tagging. The

low statistics of the final selection meant the fit could not be used at this point without large uncertainties.

The $m_{T,W}$ fit was, however, subsequently abandoned in favour of allowing the normalisation of the data-driven background to float in the fit used to extract the significance and cross section limits, which will be described in Chapter 8. This was found to streamline the analysis process and achieve comparable results.

6.2.3 tZq Boosted Decision Tree

After selection and background estimation, the signal region, as can be seen in Table 6.4, is dominated by WZ, $t\bar{t}Z$ and Z+jets backgrounds. The estimation of the Z+jets background has already been discussed, leaving only the WZ and $t\bar{t}Z$ backgrounds to consider. In order to discriminate between these backgrounds and signal events, a multivariate analysis (MVA) is performed. Although many different types of MVA exist, the Boosted Decision Tree (BDT) [117] was chosen for this analysis as it is widely used and well understood within the Single Top working group of CMS.

At its simplest, a decision tree is a set of questions (nodes), based on the parameters of the event being classified, the BDT input variables. Each node has two possible outcomes, which lead to independent further nodes. After a certain number of nodes, a final decision is made on whether the event is deemed signal or background (at a leaf). As the exact requirement of each node depends on all of the answers that have preceded it, the decision tree can achieve a much better separation between signal and background than an isolated requirement placed on any single variable could.

A single tree might not give very good performance, but by training many trees this performance can be significantly improved. Once a large number of trees with different configurations exist that have been trained on the data, they can be used to produce a single discriminant value based on how many classify the event as signal. Typically the output of a BDT ranges from -1, or completely background-like, to +1, which represents completely signal-like events. Two main methods exist for training the trees: bagging and boosting.

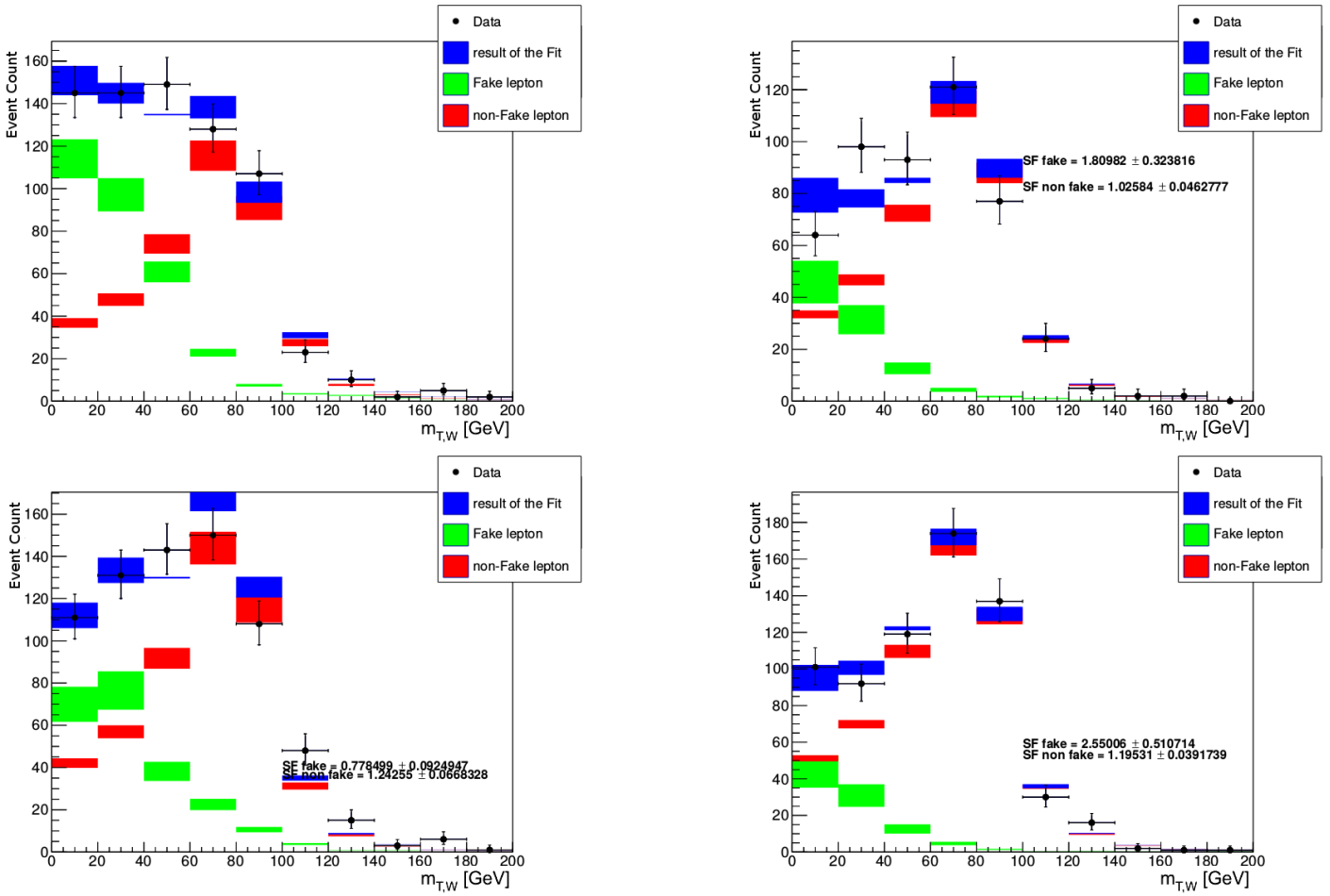


FIGURE 6.9: Fits of the transverse W mass to estimate the contributions of fake and real lepton backgrounds in the final state. The fit, carried out after lepton and Z candidate identification but before jet requirements, provides a weight for the data-driven background. The plots are, from top-left to bottom-right, for the eee, $ee\mu$, $e\mu\mu$ and $\mu\mu\mu$ channels, respectively.

Bagging takes a random subset of events from the training samples and uses them to train a tree. Each tree is trained on a different random subset of events, of which there may be overlap between trees.

The Adaptive Boost (AdaBoost) method [118] uses the complete training sample for each tree, assigning a higher weight to misclassified events, effectively modifying the next generation of trees to achieve better separation. AdaBoost is considered a robust choice of BDT training with which good results are easily achievable. It is therefore chosen as the training method for this analysis.

A known problem with BDTs is that they are easy to overtrain into interpreting statistical fluctuations as important features of signal events. This is particularly a problem with boosted training, as it can heavily weight misclassified events that should be considered outliers. In order to minimise the impact of overtraining, both signal and background samples are split in testing and training samples. The training is then carried out on the training sample only, with the test sample reserved to check the validity of the training. If the BDT's performance is better on the training sample than the test sample, it has been overtrained and will require the removal of some nodes (known as pruning).

Initially, separate BDTs were trained for the different background processes, but limited statistics led to overtraining effects. It was therefore decided that combining the two remaining dominant backgrounds, WZ diboson production and $t\bar{t}Z$ production, into a single background sample (and hence a single BDT), would create a more powerful and reliable discriminant.

6.2.3.1 BDT input variables

Many variables were tested as potential inputs for the BDT, a list of which can be found in Table 6.8. The variables that were chosen, indicated in the table with bold font, were selected because of their ability to separate the signal and background processes, and because good agreement was observed between data and simulation for each one. If variables were found to be highly correlated, the one with more separation power was chosen. Table 6.8 also shows the separating power of each of the chosen variables. The separating power, S, of a variable, y, is defined as:

Variable	Description	Separation
jjdelR	ΔR between the two leading jets	9.165e-02
bTagDiscri	The b-tag discriminant of the leading b-jet	8.182e-02
zjMinR	Minimum ΔR between the reconstructed Z and any jet	6.157e-02
TopMass	The mass of the reconstructed top	6.008e-02
lepPt	Magnitude of the vector sum of the p_T of the selected leptons	4.334e-02
leadJetEta	The η of the leading selected jet in the event	4.176e-02
secJetbTag	The b-tag discriminant of the second selected jet in the event	3.927e-02
lepMetPt	Magnitude of the vector sum of the p_T of the selected leptons and the met	3.708e-02
lbDelR	ΔR between third (w decay) lepton and b-quark	3.522e-02
lbDelPhi	$\Delta\phi$ between third (w decay) lepton and b-quark	3.522e-02
leptWPt	The p_T of the lepton coming from the W decay	3.135e-02
TotH_T	The scalar sum of the transverse momentum of the selected objects in the event	2.611e-02
NJets	The number of selected jets in the event	2.365e-02
Z p_T	The p_T of the reconstructed Z^0 boson	2.331e-02
secJetEta	The η of the second selected jet in the event	2.121e-02
totPt	Magnitude of the vector sum of p_T of selected objects	1.686e-02
zlb2DelR	ΔR between the second Z decay lepton and leading b-jet	1.664e-02
leptWEta	The η of the lepton coming from the W decay	1.502e-02
TopEta	The η of the reconstructed top	1.485e-02
zlb2DelPhi	$\Delta\phi$ between the second Z decay lepton and leading b-jet	1.423e-02
Z η	The η of the reconstructed Z^0 boson	1.380e-02
wzDelR	ΔR between the reconstructed W and Z bosons	1.342e-02
secJetPt	The p_T of the second selected jet in the event	1.280e-02
leadJetPt	The p_T of the leading selected jet in the event	1.263e-02
zlb1DelPhi	$\Delta\phi$ between the leading Z decay lepton and leading b-jet	1.241e-02
zlb1DelR	ΔR between the leading Z decay lepton and leading b-jet	1.236e-02
TopPt	The p_T of the reconstructed top	1.157e-02
totEta	η of the vector sum of p_T of selected objects	1.076e-02
H_T/totPt	Scalar sum divided by the magnitude of the vector sum of p_T of objects in the event	7.275e-03
NBJets	The number of selected b-jets in the event	1.431e-03

TABLE 6.8: The name and description of the variables considered as potential input variables to the BDT used to discriminate between the tZq signal candidates and the dominant ttZ and WZ backgrounds. Variables used in the BDT are indicated with bold text.

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy \tag{6.5}$$

where $\hat{y}_S$ and $t\hat{y}_B$ are the probability density functions of y in the signal and background samples, respectively. This variable is 0 for identical signal and background shapes of y, and 1 if there is no overlap. Figure 6.10 shows the distributions of the chosen variables in the signal and background samples. Figures 6.11-6.19 show the agreement between simulation and data for the chosen variables for the combination of all four channels. All figures demonstrate a reasonable level of agreement between data and simulation, and as such are good choices for the input variables.

TABLE 6.9: Variable importance in BDT discriminant calculation.

Variable	Importance
totHt	8.708e-02
jjdelR	8.528e-02
btagDiscri	7.814e-02
topMass	7.433e-02
wzdelR	6.888e-02
totPt	6.836e-02
leptWPt	6.459e-02
leadJetEta	6.122e-02
secJetbTag	5.411e-02
lbDelR	5.338e-02
zjminR	5.284e-02
secJetPt	4.755e-02
leptWEta	4.594e-02
Zeta	4.165e-02
zlb1DelR	3.991e-02
zlb1DelPhi	3.952e-02
zlb2DelPhi	3.725e-02

6.2.3.2 BDT Training

The BDT was created and trained using the AdaBoost method from the Toolkit for Multivariate Analysis (TMVA) [119]. The BDT is trained using 100 trees, using background and signal samples taken from simulated events that have passed the signal selection requirements. Each sample is split equally into a training and a testing sample. This results in four samples containing unique events, the signal training and testing samples with 33631 events each, and background training and testing samples each with 1426 events.

Figure 6.20 shows the distribution of the BDT discriminant for the training and testing samples for both signal and background. The BDT discriminant for testing and training samples match well, implying that no overtraining has occured in the training process. Table 6.9 shows the importance of each of the selected variables in determining the discriminant. The importance of a variable is derived from the number of nodes that use the variable, and weighted by the separation gained and the number of events present at each node. Figure 6.21 shows the correlation between the input variables in the signal and background regions. Variables that have a high correlation, appearing as a yellow or red square in the figure, give the BDT the same information, making them redundant in the training. Ideally all off-diagonal correlations would be zero.

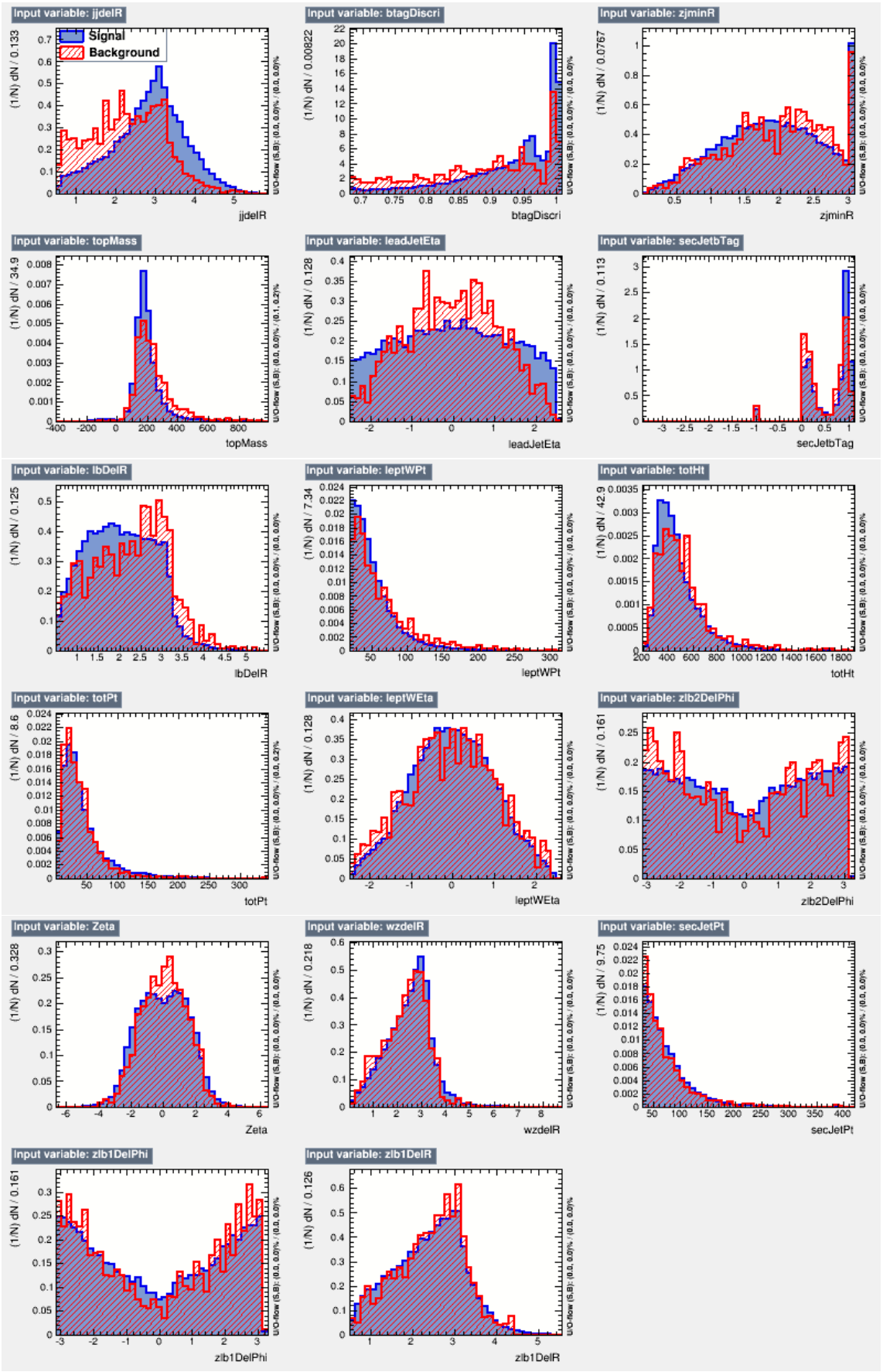


FIGURE 6.10: Distributions of the variables chosen for the BDT in the signal (blue) and background (red) samples.

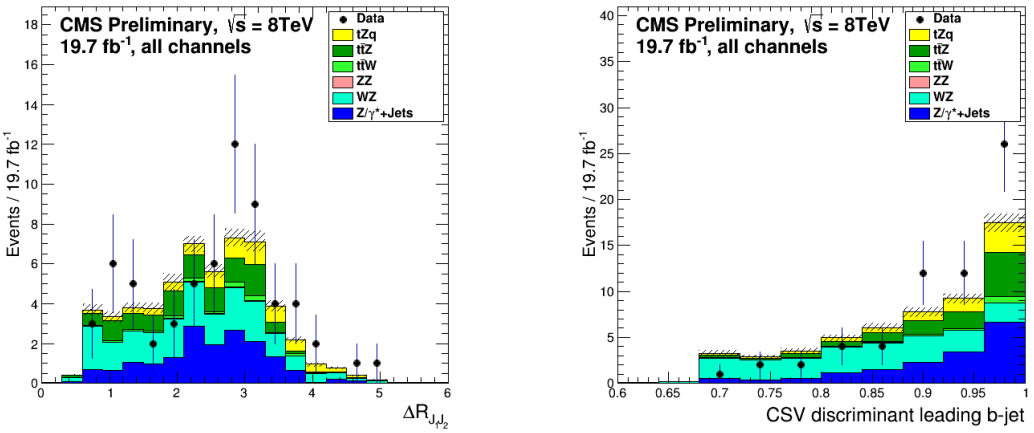


FIGURE 6.11: Distributions of the angular separation of the two leading jets in each event and the CSV b-tagging discriminant value of the leading jet after full selection for the combination of all channels.

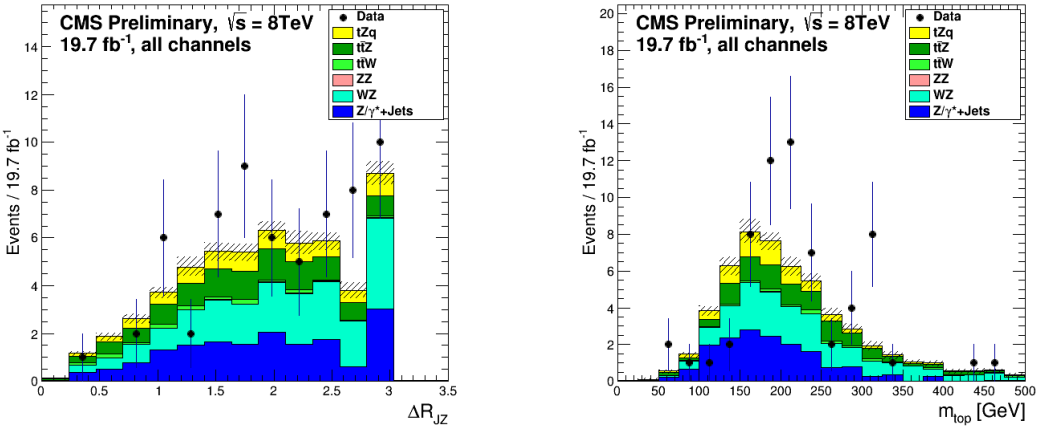


FIGURE 6.12: Distributions of the minimum ΔR between the reconstructed Z boson and any jet and the mass of the reconstructed top after full selection for the combination of all channels.

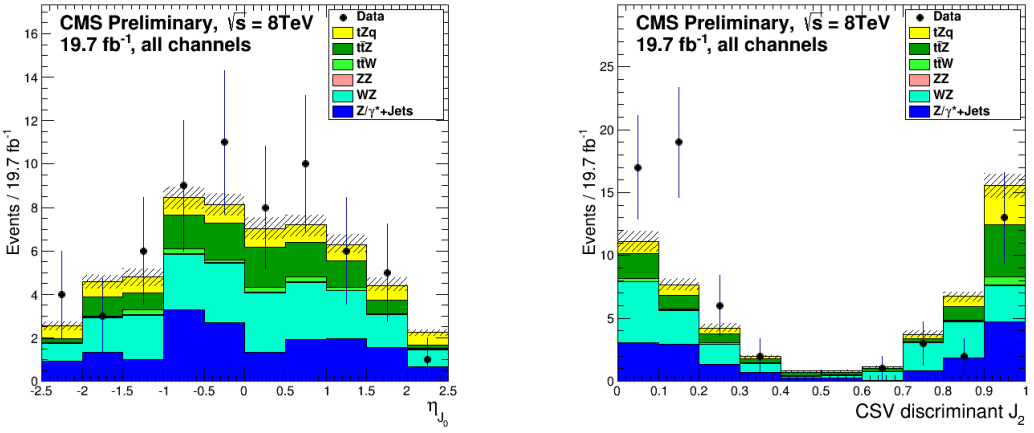


FIGURE 6.13: Distributions of the η of the leading jet and the CSV b-tagging discriminant of the second jet after full selection for the combination of all channels.

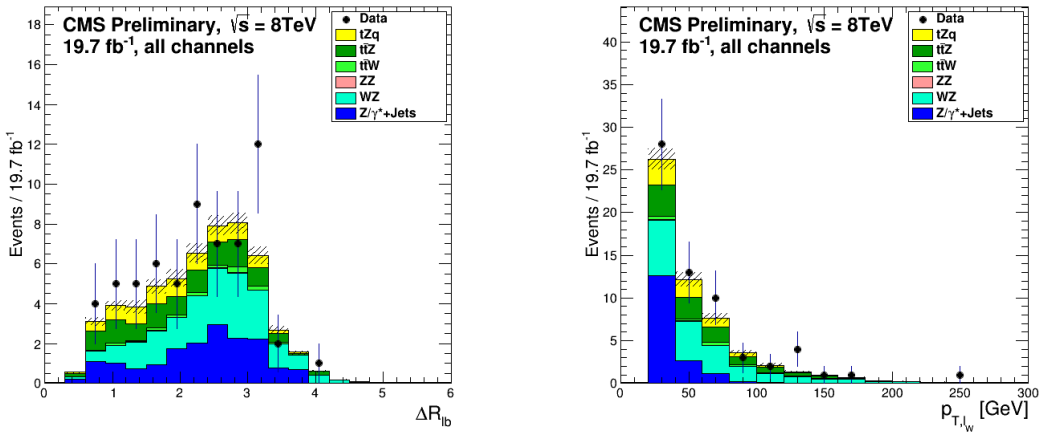


FIGURE 6.14: Distributions of the angular separation of the leading lepton and leading b-jet and p_T of the lepton originating from the W decay after full selection for the combination of all channels.

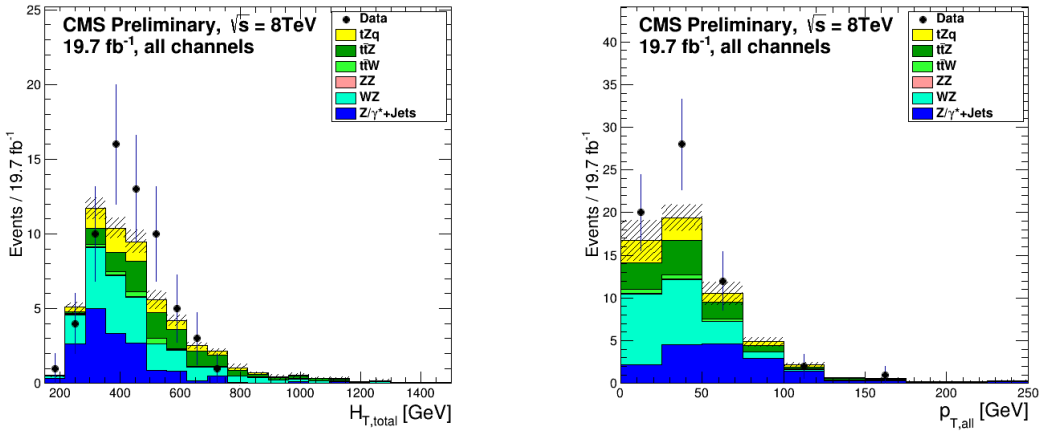


FIGURE 6.15: Distributions of the total H_T and the total p_T of the event after full selection for the combination of all channels.

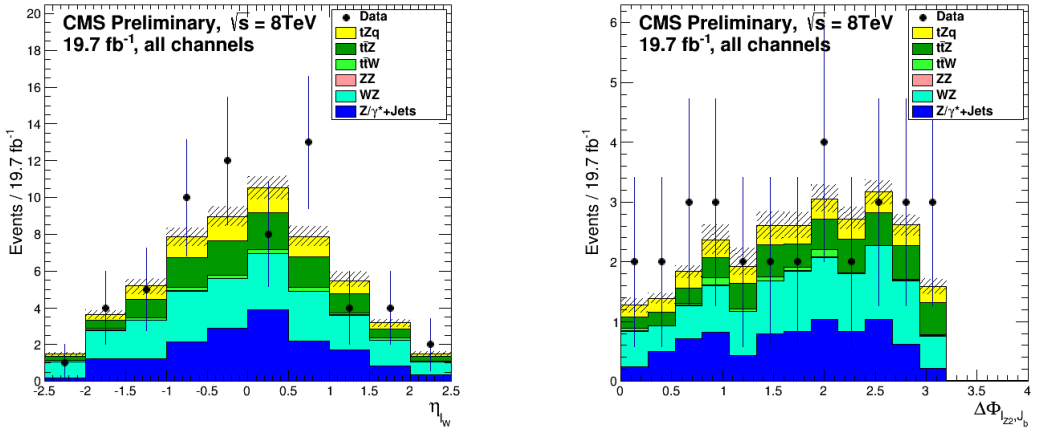


FIGURE 6.16: Distributions of the η of the lepton originating from the W decay and the separation in ϕ of the second lepton originating from the Z boson decay and the leading b-jet after full selection for the combination of all channels.

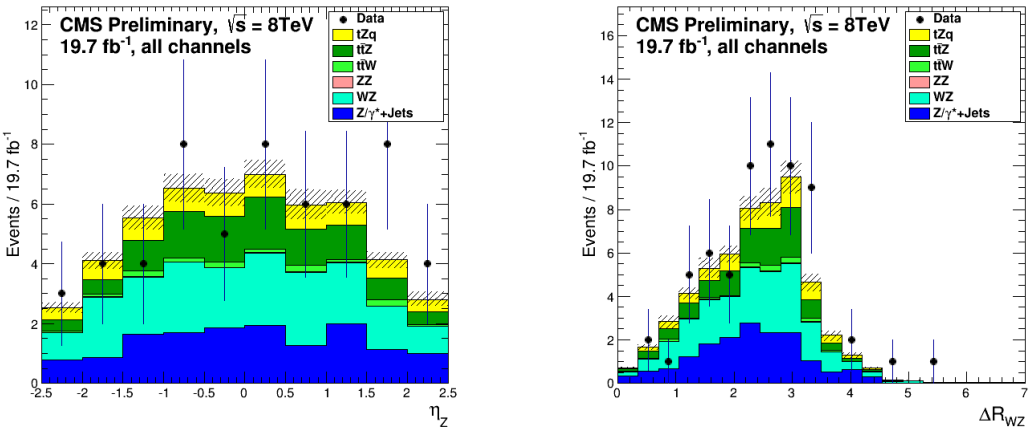


FIGURE 6.17: Distributions of the reconstructed Z η and the angular separation of the reconstructed W and Z bosons after full selection for the combination of all channels.

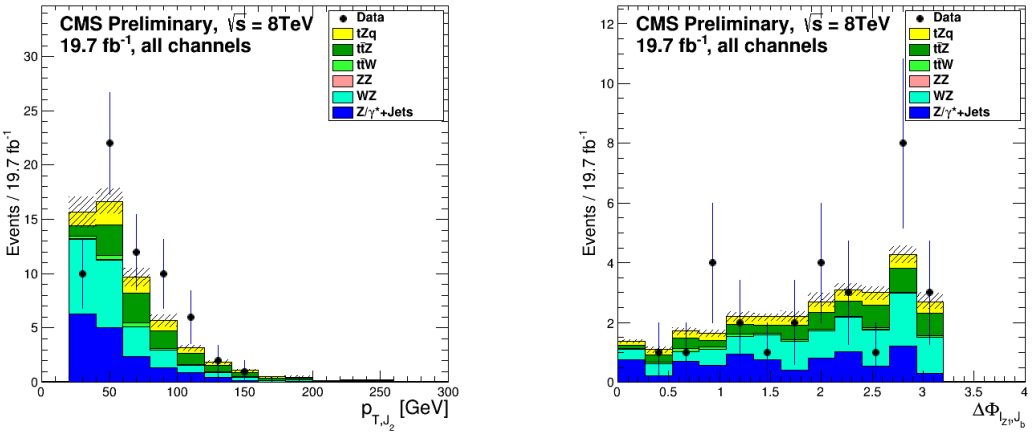


FIGURE 6.18: Distributions of the p_T of the second jet and the separation in ϕ of the second lepton originating from the Z boson decay after full selection for the combination of all channels.

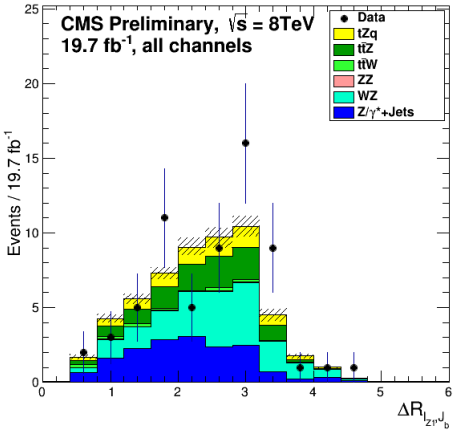


FIGURE 6.19: Distribution of the angular separation of the leading lepton originating from the Z boson decay and the leading b-jet after full selection for the combination of all channels.

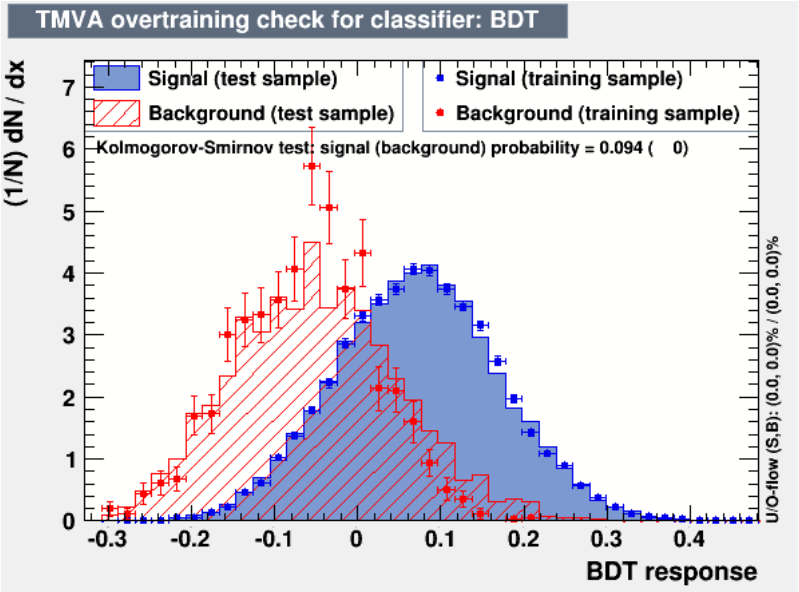


FIGURE 6.20: Distribution of the BDT discriminant for signal and background for training and test samples.

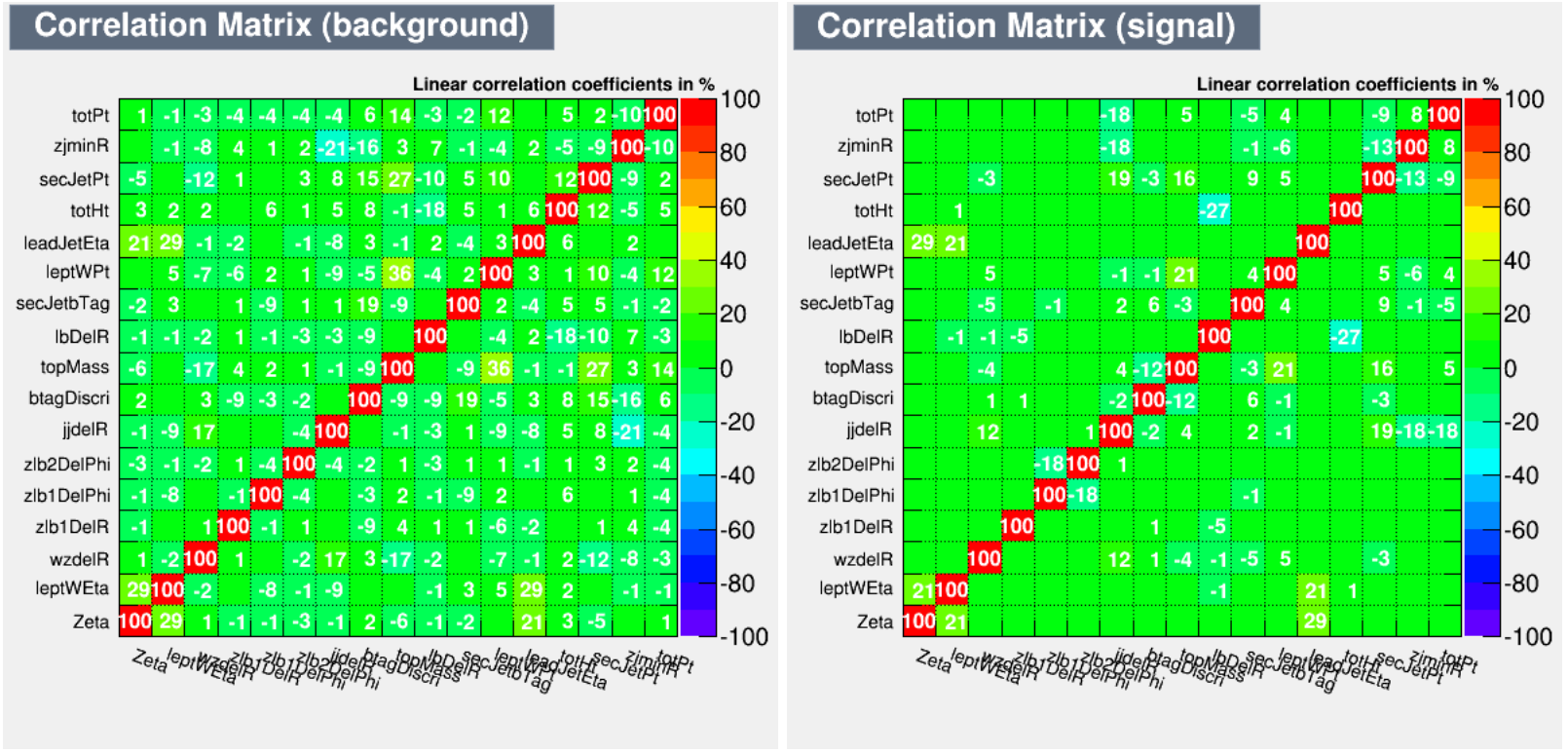


FIGURE 6.21: Correlation between BDT input variables for background and signal samples

6.2.3.3 BDT Reading

Once the BDT has been trained, it is used to 'read' the data and remaining background samples. A discriminant value is calculated for each event, the resulting distributions of which are used for the extraction of the cross section and signal strength as described in Chapter 8.

Chapter 7

Systematic Uncertainties

The cross sections of single top processes, particularly those produced in association with a vector boson, like those studied here, are relatively low. Consequently the statistical errors are of a comparable scale to the systematic uncertainties of the measurement. It is therefore vital that all sources of systematic uncertainty be understood and controlled as thoroughly as possible to allow robust measurements. There are, broadly speaking, two types of systematic uncertainty considered in these analyses;

- *Flat Rate Uncertainties* - Uncertainties in detector performance, reconstruction algorithms or theoretical cross section predictions that affect the overall rate of a process. These uncertainties are, in general, universal uncertainties relevant to all analyses and, as such, are calculated by external groups in CMS;
- *Scale Factor Uncertainties* - Sources of systematic uncertainty are introduced whenever scale factors are used to correct simulation for inconsistencies with data. As these scale factors occur on an event-by-event basis, these uncertainties affect not only the overall rate of a process but also the shape of distributions considered in the analysis. These uncertainties are usually assessed by altering the applied scale factor by one standard deviation up and down and evaluating its impact on the result.

The uncertainties are included as nuisance parameters in the statistical fit model, discussed in detail in Section 8. The statistical uncertainty arising from the limited size of the simulated events is the final type of uncertainty considered in the fit.

7.1 Flat Rate Uncertainties

7.1.1 Luminosity Measurements

CMS measures instantaneous and integrated luminosity in two ways; through a coincidence trigger in the HF sub-detector and by measuring the number of clusters found in the pixel detector. The HF measurement was used during early running of the LHC, but increasing uncertainty in the measurement due to event pileup and calibration shifts lead to the pixel based calculation method being developed during 2011. The *Pixel Cluster Counting Method* (PCC) [120], assumes that each pixel in the inner detector, of which there are $1\text{x}10^7$, has a very small probability of being a part of more than one track from any given bunch crossing. This implies that the number of pixel clusters in an event scales linearly with the number of interactions in a bunch crossing, and is, therefore, an excellent measure of luminosity. The measured rates are frequently calibrated using a Van der Meer scan [121].

The PCC method gives a value of 23.27 fb^{-1} for the integrated luminosity of the entire 2012 dataset, however a portion of this data is, for various reasons, not usable. From time to time a sub-detector will encounter problems, requiring re-calibration or rebooting before data can be taken again. Although this down time is kept to a minimum, the data taken during these times is labelled 'bad' and not usable for physics analyses. Run co-ordination provides a list of officially marked 'good' runs that should be considered by analysts; with the excluded runs the PCC method measures an integrated luminosity of 19.7 fb^{-1} for the 2012 dataset, with an associated uncertainty of 2.6% [122]. As every simulated sample is scaled to this luminosity, the associated uncertainty affects the overall normalisation of each process.

7.1.2 Lepton Efficiency

Lepton selection efficiencies and their associated uncertainties are calculated using the tag-and-probe method [123] to analyse $Z \to l^+l^-$ events, which provides a large, unbiased and highly pure lepton sample. The tag-and-probe method considers lepton pair events where one lepton, called the 'tag' lepton, is selected using very strict requirements whilst the second, or 'probe' lepton, is selected using very loose requirements. The selection

requirements whose efficiencies are to be measured, which are usually identification or isolation requirements that lie in a range somewhere between those used for the tag and the probe leptons, are then applied to the probe lepton collection. This creates two subsets of data: passing probe events and failing probe events. The efficiency of the selection is then the fraction of probe leptons that pass the considered selection criteria.

The tag-and-probe method was applied to muons and electrons independently, and for electrons separately in the barrel and endcap. To ensure the purity of the dilepton sample, the lepton pair are required to fall within the Z^0 boson mass window $70 < m_{ll} < 130$ GeV/c^2. The total lepton efficiency is then sub-divided into three separate components; the efficiency of the **trigger** to identify lepton candidates; the efficiency of the **reconstruction** algorithms to reconstruct leptons from detector information and the efficiency of the analysis **identification** and **isolation** selection requirements to select the leptons.

To measure the **trigger efficiency** of both electrons and muons the probes are selected based on the normal kinematic requirements, with the passing criterion being that it passes the HLT. The trigger efficiency is found to be greater than 99% for muons, and greater than 95% for electrons. The uncertainty on this value is of the order of 4%, which varies depending on the trigger in question.

The **reconstruction efficiencies** of electrons and muons are measured independently. The electron reconstruction efficiency is measured as the efficiency that an ECAL supercluster, the probe lepton, seeds an ECAL-driven electron candidate (passing probe), as described in Chapter 5. The probe definition also requires that the supercluster falls within the range of the inner tracker, and has a reconstructed energy of greater than 10 GeV. The reconstruction efficiency is found to be better than 85% at a supercluster energy of 10 GeV, rising to over 99% as the energy increases above 20 GeV [124].

For muons, the probe is a triggered muon candidate which passes if it satisfies the global and tracker muon criteria of the muon system (as described in Chapter 5). This gives a reconstruction efficiency of 95%-99% for all muon systems in data [125]. These uncertainties were calculated by the CMS EGamma and Muon Physics Objects Groups (POGs), respectively, for use in all analyses involving these objects.

The **identification and isolation efficiency** is defined as the efficiency that the reconstructed physics objects; GSF electrons and PF muons; pass the analyses' kinematic

requirements. The requirements used in the analysis, defined in Chapter 6, are based on recommendations from the Top PAG, and the quoted efficiencies are provided by the PAG for all recommended selections. For electrons, the efficiencies are found to be greater than 90% for electrons with p_T greater than 30 GeV, with a related uncertainty on the percent level. For muons with p_T greater than 20 GeV, the identification efficiency is better than 99%, and the isolation efficiency is better than 98%. The associated uncertainties are found to be of the sub-percent level for muons.

In practice, the uncertainties associated with the reconstruction, identification and isolation efficiencies are combined into a single uncertainty for each lepton flavour. For the tW analysis, this value was around 2% for each channel. The tZq search, containing an extra lepton, has a more conservative value of around 3% per channel. The trigger uncertainty, quoted as a separate uncertainty, varies between 2-5%, depending on the trigger in question.

7.2 Shape Uncertainties

7.2.1 Parton Density Function

Event generators assign momentum fractions and energies to the partons based on Parton Distribution Functions (PDFs) derived from data collected from many different experiments. Each experiment has its own associated uncertainties, and these must be propagated For this reason there is an uncertainty associated with the PDF measurements, and this must be propagated into physics analyses.

The PDFs are obtained using global fits on experimental data for deep inelastic scattering, Drell-Yan and jet processes. These are updated by the collaborations that perform the fits, such as CTEQ [21], every time new data and/or theoretical predictions become available. The functions used to generate the simulated samples for this analysis are from the CTEQ66 set, however CTEQ has since produced an improved set of uncertainties CT10 [126], that are used for the uncertainty calculations for the tZq search.

CT10 provides the nominal PDF weight along with 25 eigenvalues, which provide 50 alternative weights for each event. These are accessed using the LHAPDF (Les Houches Accord Parton Distribution Function) library [127]. The difference between each of these

weights and the nominal is added in quadrature and the result is used to calculate the associated systematic uncertainty. The PDF uncertainties prove to be one of the largest uncertainties in both analyses, altering the event yields of all simulated samples by at least 5%.

7.2.2 Pileup Reweighting

As mentioned briefly in Section 5.1, additional pileup interactions are included in the simulated samples, but the true number of primary vertices in the simulated events does not describe that observed in data well. This effect is compounded by the changing conditions at the LHC; even within a single data taking period the number of primary vertices in an event may change dramatically as the instantaneous luminosity of the beams changes. Poor description of the pileup in simulation can lead to incorrect estimation of the number of background and/or signal events, and as such additional corrections must be applied to all simulated samples. The number of primary vertices is reweighted to match the current running conditions of the LHC, and the associated uncertainty is incorporated into the systematic uncertainty on the results.

The number of primary vertices is extracted directly from minimum bias data over the course of the running period being examined. By varying the expected minimum bias cross section at the LHC by $\pm 5\%$ new primary vertex distributions are calculated and can be used to determine the impact on the analysis of more or less pileup in the data. Event pileup proves to be one of the smaller uncertainties in both analyses, with an impact on the event yields of simulated samples varying by less than 2% for its variations. It should be noted, however, that the $t\bar{t}Z$ sample in the tZq search is more susceptible to the pileup systematic, showing a shift in event yield of almost 5%.

7.2.3 Jet Energy Corrections

As described in Section 5.4.1.1, it is standard practice to apply corrections to reconstructed jet energies to compensate for discrepancies observed between generator and detector level jets. These corrections are designed to account for non-linearities in the calorimeter and to create a flat jet response in η and p_T. The Jet Energy Correction group carries out the various energy calibration studies that are required to calculate these corrections and

their associated uncertainties, and makes them available to the collaboration [109, 128]. Changes to the Jet Energy Scale (JES) affect the kinematics of each jet in the event, which can lead to a different number of jets passing or failing the selection requirements, altering the final event topology. This can have a serious knock on effect on the final result. To calculate the impact of the JES uncertainty, the correction factor is moved up and down by one standard deviation, with the effects also being propagated through to the calculated $\not{E}_T$. The JES uncertainty is found to be one of the more important uncertainties in the tW search. Although it does not largely impact the signal sample, it is found to affect the event yields of the dominant $t\bar{t}$ background by around 5%. The uncertainty is less pronounced in the tZq search.

The Jet Energy Resolution (JER) is defined as the standard deviation of a Gaussian fitted to the jet response of the detector. It has been found that the JER in data is ∼10% broader than found in simulation, with an associated uncertainty of comparable size [129]. To account for this, the 4-momentum of the simulated jets is smeared as a function of the true and reconstructed η and p_T. The smearing factor is applied twice or not at all to create scaled up and down systematic samples, which are included as nuisance parameters in the statistical analysis. The impact of the JER uncertainty is found to be small in both analyses, usually impacting event yields by less than a percent. There are notable exceptions to this, for example the tW Z+jets sample, but this is accounted for by the limited statistics in the simulated samples.

7.2.4 Modelling of the $\not{E}_T$

Events with neutrinos in their final state are affected by any uncertainties that originate from the modelling of the $\not{E}_T$ in simulation. The $\not{E}_T$ is calculated from the sum of the p_T of all PF-reconstructed objects along with so-called 'unclustered energy', meaning the uncertainties in these propagate into the $\not{E}_T$ uncertainty. Unclustered energy is energy that has not been included in a calorimeter cluster because of its isolation or low p_T. The energy of the PF particles has already been corrected during reconstruction, but the unclustered energy does not undergo any such correction, meaning that it effectively dominates the uncertainty in the $\not{E}_T$. To evaluate this uncertainty, the p_T of all reconstructed objects in the event are removed from the $\not{E}_T$ and the remaining energy is scaled up and down by 10%, the default uncertainty of the energy. Other uncertainties that impact the $\not{E}_T$, such

as JER and JES, are propagated at the time of calculation and the impact on the $\not{E}_T$ is included in their respective uncertainties. This uncertainty has a direct impact on the tW analysis, where there is an explicit requirement placed the $\not{E}_T$ in the ee and $\mu\mu$ channels and a requirement on the H_T in the $e\mu$ channel. Although there are no requirements placed on the $\not{E}_T$ in the tZq analysis, it still impacts on the result because it is used to construct several of the BDT input variables. The variations in the discriminant as a result of the change in shape of the input variables are used to estimate the associated systematic uncertainty.

As outlined in Section 6.1.3.1 the effects of pile-up on the $\not{E}_T$ distribution are corrected in simulation to match data using scale factors obtained from a Z+jets enriched control region. The difference between the original and scaled event yields is used as the uncertainty on the background normalisation arising from this reweighting. This uncertainty, which only affects the reweighted Z+jets sample in the tW search, is found to be very large, especially in the $t\bar{t}$ control regions where there are limited statistics in the simulated samples.

7.2.5 B-tagging uncertainty

The b-tag and vertexing (BTV) physics object group is responsible for the evaluation of efficiencies and misidentification rates of the available b-tagging algorithms in both data and simulation. When there is an observed discrepancy between the two, it publishes recommended scale factors to be applied to simulated events to ensure good agreement with the data. These studies have been carried out using $t\bar{t}$ and multijet samples from 8 TeV running [113]. These samples were chosen to ensure events with at least two jets, but a variable number of leptons. The b-tagging scale factors were applied differently in the two analyses, and the uncertainties were also, therefore, treated differently.

In the tW analysis the b-tag scale factor was applied as a probability that, in each simulated event, a successfully tagged jet would not be acknowledged as tagged. This probability was adjusted up and down by the relevant uncertainties as calculated by the BTV working group to give the estimate of the uncertainty.

In the tZq analysis the scale factors were replaced with a p_T and η dependent adjustment to the event weight for each jet, light or tagged. Although the two methods are expected to

give very similar results, this method was chosen because the reweighting does not change the b-jet topology in the same way as the probability method; as the signal region comprises different b-jet multiplicities the probability method could have a disproportionately large impact on the analysis. To calculate the uncertainty, the reweighting factor is adjusted up and down by the uncertainties calculated by the BTV POG. In both cases the scale factors were adjusted by a factor of a few percent per b-tagged jet. The impact of the b-tagging uncertainty is found to be more significant in the tZq analysis, which may have more b-tagged jets in its final state than the tW search.

7.2.6 Data-driven Reweighting Uncertainties

Similarly to the uncertainty introduced in Section 7.2.4, the reweighting of the reconstructed Z boson p_T in the data driven Z+jets background estimation described in Section 6.2.2 also introduces a systematic uncertainty. As with the $\not{E}_T$ modelling, the difference between the default and reweighted event yields is used as the uncertainty on the background normalisation.

7.3 Modelling Uncertainties

In addition to the uncertainties that change the normalisation and shape of simulated distributions, the origin of the simulated events themselves present further uncertainties for an analysis. Despite increasingly accurate measurements and interpretations of the many parameters of the Standard Model of particle physics, there are still gaps in our knowledge that could, in principle, lead to greatly altered simulated samples. Varying the parameters used when generating the simulated samples can produce radically different events both kinematically and topologically. The usual approach to account for our potential ignorance of these parameters is to produce simulated datasets where the parameters are varied to larger and smaller values. Applying the standard event selection requirements then produces newly shaped distributions that are included as nuisance parameters in the same way as the shape uncertainties.

7.3.1 QCD Renormalisation and factorisation scales

The PDFs are, for finite levels of perturbation theory, functions of the factorisation and normalisation scales. For the simulation of events these are parametrised together as Q^2, which for a hard scattering involving a top quark is chosen to be $Q^2 = m_{top}^2 + \sum p_T^2$. To investigate the impact of these scales on the analysis, additional simulated samples are produced where Q^2 is varied by factors of 0.5 and 2 for 'scale down' and 'scale up' samples, respectively. These samples, produced centrally, also include variations in the radiation of gluons from the incoming and outgoing partons, known as initial-state (ISR) and final-state radiation (FSR), respectively.

7.3.2 Parton Level Matching Thresholds

As explained in Section 5.1.1, several of the simulated samples, including the WZ sample, are produced using a hard scattering generated using the MadGraph matrix-element (ME) generator but the parton showering (PS) and hadronisation is performed by PYTHIA. These two stages must be matched in order to create a smooth transition between the two. This process is dependent on the Kt-MLM parton level matching threshold scale [95], which is typically set to 20 GeV. Dedicated systematic samples are produced where the threshold is set to 10 GeV and 40 GeV, respectively, to assess the impact of the threshold on the analysis result. As this systematic only applies to samples that require the matrix element to parton shower matching it cannot be evaluated for every simulated sample; in the tW analysis it is only relevant for the $t\bar{t}$ sample, in which it is one of the leading systematic uncertainties, and in the tZq analysis it only applies to the WZ samples in which it is similarly important.

7.3.3 Analysis Dependent Modelling Uncertainties

As described in Section 2.2.2, there are two separate schemes that can be used to calculate the cross section of tW production, the Diagram Removal and Diagram Subtraction schemes. The Diagram Removal was chosen as the scheme for use in the analysis, and as such the central value of the result is calculated using simulated samples with this scheme. An alternative sample was produced using the Diagram Subtraction scheme, and the uncertainty was calculated as the relative difference in event yield from the nominal sample.

It should also be noted that the two schemes were devised such that the difference in their results gives an estimate of the size of the tW/$t\bar{t}$ interference.

Reducing the uncertainty on the top quark mass has been one of the major successes of the LHC experiments; precision measurements from the pre-LHC era achieved sub-1% uncertainty [130], but combinations of the LHC and Tevatron results [11] along with full Run 1 dataset combinations within CMS [131] have improved precision to 0.38%. During early analyses the uncertainty on the top quark mass was included as a systematic uncertainty; this practice has since been abandoned due to the reduction of the uncertainty, although measured cross section values may be quoted as a function of the top mass if the precise measurement is an important factor in the result. Processes explicitly containing a top quark have additional samples generated with the value of the top quark mass shifted up or down from the nominal 172.5 GeV/c^2 by 6 GeV/c^2. As the uncertainty of the top quark mass is smaller than 1 GeV/c^2 , the difference in event yields from the nominal sample are interpreted as 3 σ uncertainties. This uncertainty, only evaluated for the tW search, is one of the main uncertainties in the $t\bar{t}$ and tW samples.

There are also additional corrections that are applied only to the $t\bar{t}$ simulation: **spin correlations** and **top** p_T **reweighting**. Simulated $t\bar{t}$ samples with and without spin correlations are generated; the difference in event yields is treated as a 1σ variation and symmetrised to be included as an uncertainty. It has been observed that the top quark p_T is softer in data than predicted by the generators used in the simulation of the samples [132]. For this reason an event-by-event reweighting is carried out based on the p_T of the top and anti-top quarks in the simulated sample. The uncertainty associated with this reweighting process is determined by applying the reweighting twice or not at all [133], but is not found to have a significant impact on the result. An additional uncertainty of $\pm$ 6.7% was assigned to the $t\bar{t}$ cross section to account for its measured uncertainty. As simulation-based $t\bar{t}$ background estimation only contributes a significant background to the tW search these uncertainties are not present in the tZq search. The combination of these effects are found to be small compared to other systematic contributions in the analysis.

One of the main backgrounds to the tZq analysis, ttZ production, is itself a rare process for which evidence has only recently been found [49]; the modelling of the simulated samples is, therefore, not well understood. In order to assess the validity of the simulated samples,

TABLE 7.1: The systematic samples used per channel in each analysis. Further explanation of each may be found in the text.

Analysis	Process	Systematic Samples
tW	tW	Q^2, m_{top}, DS/DR
	$t\bar{t}$	Q^2, m_{top}, ME/PS matching
tZq	tZq	MC@NLO
	WZ	Q^2, ME/PS matching

a second ttZ sample was created using the aMC@NLO generator in conjunction with HERWIG [134]. As the difference in event yields between the two is found to be small the modelling is considered consistent and valid for the analysis. The difference is used to estimate the contribution to the systematic uncertainty. Table 7.1 lists the relevant additional samples used per channel, for each analysis.

7.4 Impact of uncertainties

Tables 7.2-7.4 show the impact on the event rate, in percentage, for each of the considered uncertainties for each of the signal and control regions of the tW analysis. The two numbers for each entry refer to the up and down variation of the systematic, respectively. The uncertainties are evaluated for the $t\bar{t}$ enriched control regions along with the signal region as they are all included simultaneously in the fit described in Chapter 8, and are required in all regions to assess their impact on the significance and cross section measurements. The theoretical uncertainties, including the matching threshold, Q^2 scale and top quark mass are found to be the most important uncertainties for the tW search. The best way to improve the result would, therefore, come from a better theoretical understanding of the production of tW as well as the dominant $t\bar{t}$ background.

Tables 7.5 - 7.8 shows the impact on the event yield, in percentage, for each of the considered uncertainties after all event selection requirements have been applied for the tZq search for all considered final states. The statistical uncertainty due to the small sample size in the signal region is found to be the leading systematic uncertainty, particularly in the simulated $t\bar{t}Z$ sample. The result should improve substantially with additional simulated events and larger data samples.

TABLE 7.2: Rate impact of the systematics for the 1j1t signal region in the tW search. The rate impacts of both up and down variations of each considered systematic are shown for the tW signal sample, along with the dominant $t\bar{t}$ background and 'Other' backgrounds, which contains mostly Z+jets events.

Systematic Uncertainty (ee/$e\mu$/$\mu\mu$)	tW (%)	$t\bar{t}$ (%)	Other (%)
Luminosity	±4.4	±4.4	±4.4
Lepton identification	±2.33/±1.86/±2.28	±2.33/±1.86/±2.28	±2.33/±1.86/±2.28
JER	−0.70 /+0.56 /+1.61 −1.09/ −0.60/ −2.65	+0.10 /−0.36 /+1.13 +1.23/ −0.43/ +1.72	−8.98 /−5.23 /−4.75 +2.06/ −5.60/ +2.66
JES	+1.51 /+1.03 /+2.24 −3.87/ −1.04/ −1.58	+5.47 /+5.71 /+7.47 −7.26/ −4.77/ −3.91	−10.2 /−4.62 /−7.43 +12.6/ −4.00/ +7.33
MET modeling	+2.42 /+0.56 /+0.05 −0.31/ −0.23/ +0.11	+1.99 /−0.16 /+1.69 −1.24/ +0.51/ +0.61	-
Event pile up	+1.76 /+1.84 /+2.31 −3.19/ −1.73/ −2.90	+3.31 /+1.94 /+3.72 −1.23/ −2.26/ −0.65	+8.28 /+0.54 /+16.6 −10.5/ −5.58/ −4.46
B-tagging data/MC scale factor	+0.63 /−0.41 /+0.11 +0.98/ −0.15/ +0.02	+0.87 /+0.18 /+0.85 −0.63/ −0.44/ +2.63	−4.35 /−4.13 /+1.93 −1.52/ −3.65/ +1.65
Q^2 scale	+2.25 /−1.06 /+1.67 −1.00/ −3.04/ +0.31	+8.24 /+12.2 /−0.46 −8.93/ −6.74/ −10.0	-
ME/PS matching thresholds	-	−3.22 /+4.03 /−0.76 +1.20/ +5.73/ +2.80	-
tW DR/DS scheme	+0.00 /+0.00 /+0.00 +5.83/ +2.80/ +6.92	-	-
Top quark mass	+3.38 /−0.16 /+2.50 −1.58/ −5.34/ −3.01	+0.80 /+0.64 /−0.38 +5.50/ +6.20/ +2.05	-
Lepton energy scale	+0.76 /+0.06 /−0.58 −1.19/ +0.64/ +0.37	−0.63 /−0.06 /−0.17 +0.34/ +0.07/ +1.44	+0.09 /−4.05 /+6.65 −0.18/ −9.25/ +5.45
$t\bar{t}$ cross section	-	±6.705	-
PDF	+6.03 /+5.85 /+5.77 −6.03/ −5.85/ −5.77	+5.63 /+4.89 /+5.34 −5.63/ −4.89/ −5.34	-
Z+jet MET scale factor	-	-	+17.24 /+10.49 /+19.29 −17.24/ −10.49/ −19.29

TABLE 7.3: Rate impact of the systematics for the 2j1t control region in the tW search. The rate impacts of both up and down variations of each considered systematic are shown for the tW signal sample, along with the dominant $t\bar{t}$ background and 'Other' backgrounds, which contains mostly Z+jets events.

Systematic Uncertainty (ee/$e\mu$/$\mu\mu$)	tW (%)	$t\bar{t}$ (%)	Other (%)
Luminosity	±4.4	±4.4	±4.4
Lepton identification	±2.33/±1.86/±2.28	±2.33/±1.86/±2.28	±2.33/±1.86/±2.28
JER	−1.59 /+2.54 /+1.18 +0.38/ +1.46/ +1.49	+0.20 /+0.25 /−0.17 −0.06/ +0.16/ +0.15	−9.94 /−3.00 /−3.50 +11.4/ +0.26/ +8.96
JES	−6.21 /−4.70 /−3.76 +5.13/ +5.56/ +5.44	−0.63 /−0.30 /−0.21 −0.26/ −0.42/ −1.09	+0.93 /−2.07 /+1.24 +15.5/ +4.96/ +18.3
MET modeling	+1.04 /+1.09 /+2.30 +0.24/ +0.15/ +1.18	+0.69 /+0.41 /−0.32 −0.99/ −0.06/ −1.27	-
Event pile up	+2.16 /+1.79 /+1.81 −4.42/ −2.21/ −1.52	+2.11 /+1.86 /+0.93 −3.19/ −1.54/ −2.70	+5.95 /+5.24 /+15.0 −15.6/ −3.89/ −6.65
B-tagging data/MC scale factor	−1.63 /+1.06 /+1.78 −4.64/ +0.58/ −0.02	−1.49 /−0.53 /−0.56 −0.28/ −0.05/ −0.42	−3.05 /−2.78 /+1.19 −3.99 /−1.85/ +5.34
Q^2 scale	+8.47 /+6.23 /+4.01 +5.85/ +0.77/ −0.28	+4.25 /+6.96 /−2.08 −4.49/ −3.55/ −5.64	-
ME/PS matching thresholds	-	−2.20 /−0.05 /+1.09 +1.49/ +1.80/ +0.83	-
tW DR/DS scheme	+0.00 /+0.00 /+0.00 −2.63/ −5.19/ −7.64	-	-
Top quark mass	+12.7 /+7.14 /+5.17 +0.34/ −1.52/ −4.16	+4.47 /+2.58 /+4.08 −3.31/ −1.43/ −2.31	-
Lepton energy scale	+1.73 /+1.20 /+1.74 +0.03/ +1.35/ −0.24	+0.25 /+0.60 /−0.66 −0.28/ +0.28/ +0.16	−1.19 /+0.23 /+8.63 +0.72 /+1.97/ +0.92
$t\bar{t}$ cross section	-	±6.705	-
PDF	+7.27 /+6.12 /+4.47 −7.27/ −6.12/ −4.47	+4.73 /+4.97 /+5.06 −4.73/ −4.97/ −5.06	-
Z+jet MET scale factor	-	-	+17.70 /+8.13 /+19.98 −17.70/ −8.13/ −19.98

TABLE 7.4: Rate impact of the systematics for the 2j1t control region in the tW search. The rate impacts of both up and down variations of each considered systematic are shown for the tW signal sample, along with the dominant $t\bar{t}$ background and 'Other' backgrounds, which contains mostly Z+jets events.

Systematic Uncertainty (ee/$e\mu$/$\mu\mu$)	tW (%)	$t\bar{t}$ (%)	Other (%)
Luminosity	±4.4	±4.4	±4.4
Lepton identification	±2.33/±1.86/±2.28	±2.33/±1.86/±2.28	±2.33/±1.86/±2.28
JER	+0.22 /−2.00 /−1.67 +0.01/ −1.71/ −0.96	+1.78 /+0.19 /+0.48 −0.16/ −0.84/ −0.07	−2.01 /−14.3 /+1.38 −0.46/ −0.28/ −12.7
JES	+0.59 /−0.20 /−1.60 +1.97/ +0.50/ −2.21	−0.01 /+0.20 /+0.32 +0.02/ −1.03/ +0.57	−15.1 /+1.85 /−6.91 −15.3/ −5.69/ −7.05
MET modeling	+2.60 /−0.10 /−3.95 −1.17/ −0.54/ −0.41	+1.03 /−0.25 /+0.26 +1.14/ −0.01/ +1.04	-
Event pile up	+1.53 /+1.40 /+1.77 +3.42/ −0.89/ −7.53	+1.76 /+1.83 /+2.77 −0.11/ −2.29/ −1.93	−4.32 /−23.60 /+3.67 −9.09/ −4.80 /−9.36
B-tagging data/MC scale factor	−0.77 /−0.60 /−1.89 +7.00/ −2.00/ +3.61	+1.81 /−0.11 /+1.07 +0.98/ −0.18/ +0.02	−1.35 /+2.37 /−1.21 −0.70/ −33.9/ −12.3
Q^2 scale	−6.42 /+5.33 /−2.85 −9.70/ −1.66/ −11.0	+11.9 /+7.44 /+2.51 −8.41/ −8.67/ −10.4	-
ME/PS matching thresholds	-	+5.59 /−0.62 /−4.86 +3.97/ +1.35/ +2.23	-
tW DR/DS scheme	+0.00 /+0.00 /+0.00 −18.1/ −10.1/ −17.4	-	-
Top quark mass	−2.47 /+4.86 /−0.60 −11.5/ −4.14/ −11.2	+9.88 /+6.70 /+4.80 −5.59/ −5.52/ −8.11	-
Lepton energy scale	−1.70 /−4.29 /−1.85 −3.00/ −1.39/ +1.08	+0.27 /−0.88 /+0.47 +0.72/ −0.79/ −0.96	−1.14 /+3.94 /−27.5 −18.3/ +0.31/ +0.80
$t\bar{t}$ cross section	-	±6.705	-
PDF	+5.95 /+5.05 /+6.45 −5.95/ −5.05/ −6.45	+4.42 /+4.73 /+4.30 −4.42/ −4.73/ −4.30	-
Z+jet MET scale factor	-	-	+17.11 /+8.37 /+19.42 −17.11/ −8.37/ −19.42

TABLE 7.5: Impact on the event yields (in percentage) for the up and down variations of each considered systematic in the *eee* channel of the tZq search. All yields are after all event selection requirements have been applied.

Systematic	tZq (%)	$t\bar{t}Z$ (%)	WZ (%)	Other (%)
JES	0.0% 0.1%	−0.3% −1.9%	0.7% 0.0%	0.1% −0.0%
bTag	2.0% −2.0%	1.4% −3.1%	3.5% −4.3%	−6.2% −11%
Trigger	3.6% −3.6%	3.6% −3.6%	3.6% −3.6%	3.6% −3.6%
PDF	6.5% −4.8%	8.4% −6.4%	7.6% −5.4%	10% −6.8%
Pile-up	−0.1% 0.1%	−4.6% 3.9%	0.6% −0.4%	−3.1% 3.1%
JER	0.7% −0.8%	−0.4% 2.7%	1.1% −0.6%	0.6% −0.1%
Statistical	±0.9	±8.7	±5.9	±3.7
Lumi	±2.6	±2.6	±2.6	±2.6

TABLE 7.6: Impact on the event yields (in percentage) for the up and down variations of each considered systematic in the *ee*μ channel of the tZq search. All yields are after all event selection requirements have been applied.

Systematic	tZq (%)	$t\bar{t}Z$ (%)	WZ (%)	Other (%)
JES	0.1% 0.1%	1.0% 0.6%	−0.4% −0.0%	0.1% −0.1%
bTag	2.0% −2.0%	−0.6% −6.3%	3.6% −3.6%	−12% −16%
Trigger	3.5% −3.5%	3.5% −3.5%	3.5% −3.5%	3.5% −3.5%
PDF	6.4% −4.8%	8.3% −6.5%	7.3% −5.3%	8.4% −6.6%
Pile-up	−0.1% 0.1%	−4.5% 4.0%	1.1% −1.2%	1.8% −3.1%
JER	0.7% −0.7%	−1.6% 1.2%	2.0% −2.9%	0.3% −0.3%
Statistical	±0.9	±8.0	±5.5	±4.3
Lumi	±2.6	±2.6	±2.6	±2.6

TABLE 7.7: Impact on the event yields (in percentage) for the up and down variations of each considered systematic in the $e\mu\mu$ channel of the tZq search. All yields are after all event selection requirements have been applied.

Systematic	tZq (%)	$t\bar{t}Z$ (%)	WZ (%)	Other (%)
JES	−0.1% 0.1%	0.6% 0.0%	−0.6% −0.0%	−0.0% −0.1%
bTag	2.0% −2.0%	1.1% −4.1%	3.8% −3.7%	−14% −25%
Trigger	4.7% −4.7%	4.7% −4.7%	4.7% −4.7%	4.7% −4.7%
PDF	6.3% −4.8%	15.8% −6.9%	7.2% −5.1%	6.7% −5.6%
Pile-up	0.1% −0.0%	−4.0% 3.0%	0.3% −0.1%	−1.8% 1.9%
JER	0.8% −0.9%	−1.3% 3.0%	3.4% −3.8%	0.2% −0.4%
Statistical	±0.8	±8.4	±5.3	±3.6
Lumi	±2.6	±2.6	±2.6	±2.6

TABLE 7.8: Impact on the event yields (in percentage) for the up and down variations of each considered systematic in the $\mu\mu\mu$ channel of the tZq search. All yields are after all event selection requirements have been applied.

Systematic	tZq (%)	$t\bar{t}Z$ (%)	WZ (%)	Other (%)
JES	−0.1% 0.1%	0.0% 1.2%	1.4% 0.0%	0.1% 0.0%
bTag	2.0% −2.0%	2.5% −2.5%	3.9% −3.9%	−27% −38%
Trigger	2.5% −2.1%	2.5% −2.1%	2.5% −2.1%	2.5% −2.1%
PDF	6.4% −4.9%	9.6% −7.2%	7.2% −5.2%	5.4% −5.4%
Pile-up	−0.3% 0.3%	−4.6% 3.9%	0.7% −0.4%	−1.8% 0.1%
JER	0.7% −0.5%	1.1% 3.2%	3.4% −3.1%	0.4% −0.2%
Statistical	±0.8	±7.8	±5.2	±4.0
Lumi	±2.6	±2.6	±2.6	±2.6

Chapter 8

Results

Once the event selection has been applied and the systematic uncertainties have been evaluated, the results from both searches (tW and tZq) are determined using the same basic statistical evaluation method. Their respective event yields are used as inputs to a likelihood function that evaluates the cross section of the signal process and the significance of the result. The statistical model, detailed in the first part of this chapter, comprises one parameter of interest, the cross section of the signal process in question, and a number of nuisance parameters including the background event yields and systematic uncertainties. The signal cross sections, along with confidence levels at one standard deviation, are evaluated using a profile likelihood method [135]. Although this method is ideal for the treatment of small samples, the tZq search has such small yields that for the individual channels an accurate cross section measurement cannot be made. In this instance an upper limit of production is calculated based on the observed data.

A maximised likelihood ratio [136] is used to quantify an observed excess of events over the background only hypothesis. By comparing a purely background prediction with the combined signal and background model a significance of the measured cross section can be calculated.

The measured tW cross section is then compared to the theoretical prediction to directly estimate the CKM matrix element $|V_{tb}|$.

The final section of this chapter will compare the results of the two presented analyses with competing results from the LHC.

8.1 Statistical Model

The cross section and statistical significance are extracted using a binned likelihood function. For the tW measurement this function takes the form of binned event counts across the three dilepton final states - ee, $e\mu$ and $\mu\mu$ - and the three defined sample regions: the signal region (1j1t) and two $t\bar{t}$ -enriched control regions (2j1t and 2j2t). The three regions must be included simultaneously to allow the $t\bar{t}$ background estimation (see Chapter 6). The function used for the tZq search is the binned output of the BDT for the four different trilepton final states, eee, $ee\mu$, $e\mu\mu$ and $\mu\mu\mu$.

Regardless of how the bins are defined, any bin i can be independently specified by the number of observed events, n_i. The probability of observing n_i events is described by a Poisson distribution:

$$P(n_i) = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \tag{8.1}$$

where λ_i is the Poisson mean prediction of the model for the ith bin. In a Poisson distribution, λ is a function of the model parameters, which in this case is only the signal cross section, σ_s. However, instead of including σ_s directly, a signal strength modifier, μ, is used as a parameter that scales the number of expected signal events, S. The expected number of signal events is calculated for an arbitrary signal cross section, in this case the SM cross section σ_s^{SM}, meaning that the actual calculated cross section will be of the form $\mu \cdot \sigma_s^{SM}$. The expected event count in the bin may then be written as

$$\lambda_i = \mu S_i + \sum_k B_{k,i} \tag{8.2}$$

where k runs over all background processes and $B_{k,i}$ is the background contribution of process k in bin i.

When considering an entire distribution rather than an individual bin, the probability of obtaining the dataset, n, is the product of the Poisson probability of obtaining the yield in each of the bins:

$$P(n|\lambda) = \prod_i \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \tag{8.3}$$

where i runs over all the bins in the distribution. In fact, this may be generalised further to include multiple channels and sample regions, such as those used in the analyses in this thesis, by allowing i to run over not only all of bins in one distribution, but across all bins in the distribution obtained from each of the intended regions.

The event yields are also affected by the systematic uncertainties discussed in Chapter 7. For each independent source of uncertainty, u, a nuisance parameter, δ_u, is introduced. The expected event yield for bin i, λ_i, is then a function of these nuisance parameters. Each nuisance parameter takes the form of a Gaussian prior with a mean of 0 and a standard deviation of 1, which allows the corresponding event yield to vary up or down by one standard deviation. A Gaussian probability density for any value x is given by

$$Gauss(x|x_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\left(\frac{x_0 - x}{\sigma}\right)^2\right) \tag{8.4}$$

where x_0 is the mean and σ the standard deviation of the Gaussian.

The likelihood function, which will be used in the extraction of both cross section and signal strength significance, is given by the product of the Poisson probability of the dataset, given in equation 8.3, and the Gaussian priors for the parameters δ_u:

$$L(\mu, \delta|n) = \prod_i \frac{\lambda_i(\mu, \delta)^{n_i} e^{-\lambda_i(\mu,\delta)}}{n_i!} \cdot \prod_u \frac{1}{\sqrt{2\pi}} exp\left(-\frac{\delta_u}{2}\right)^2 \tag{8.5}$$

It will be necessary to assess the agreement of the observation, n, with one prediction using parameters $\{\mu, \delta\}$ relative to another prediction with parameters $\{\mu', \delta'\}$. This is done through the 'likelihood ratio' (LR):

$$LR = \frac{L(\mu, \delta|n)}{L(\mu', \delta'|n)} \tag{8.6}$$

In practice the most commonly used likelihood ratio at the LHC is the profiled log-likelihood ratio [137]:

$$q_\mu(n) = \begin{cases} -2\ln\frac{L(\mu,\delta)}{L(\mu',\delta')} & , \quad \mu \geq \mu' \geq 0 \\ 0 & , \quad \text{else} \end{cases} \tag{8.7}$$

This ratio will be used for the extraction of limits and the significance of cross section measurements.

8.1.1 Cross section extraction

The cross section of the signal process is calculated using the profile likelihood method, which maximises the function in equation 8.5 allowing μ and δ_u to float. The value of μ obtained in this way is then translated into a cross section.

Along with this central value, a 68% confidence limit is produced for the cross section. To calculate this, the likelihood ratio introduced in equation 8.7 is used. For this calculation, the denominator is set to the central value of the cross section determined above. The nuisance parameters in the numerator are allowed to float to maximise the likelihood, and the value of μ that represents a 68% agreement between the two models is determined.

8.1.2 Signal strength significance

To calculate the significance of any signal excess over the background-only hypothesis, the likelihood ratio from equation 8.7 is used. In this case the likelihood ratio is used to compare the combined signal and background prediction in the numerator, and the background-only prediction in the denominator:

$$q_\mu(n) = -2\ln\left(\frac{L_{s+b}}{L_b}\right) = -2\ln\frac{max_{\mu,\delta}L(\mu,\delta|n)}{max_\delta L(\mu=0,\delta|n)} \tag{8.8}$$

Both an expected significance, taken purely from simulation, and an observed significance, taking into account the observed data, are calculated. In practice this ratio is difficult to calculate analytically, so toy simulations based on the expected parameters are used. For both the expected and observed significance, the denominator is calculated by creating many toy simulations with no signal (i.e. $\mu = 0$) and creating a distribution for the likelihood function. The observed significance is then the fraction of these toys whose

value of L_b lies above the observed value. This is usually quoted as a number of standard deviations away from the mean value of L_b for the background-only hypothesis.

The expected significance is calculated by creating a set of toy simulations assuming the SM signal cross section, i.e. $\mu = 1$. The expected significance is then the fraction of the background-only toys whose values of L_b are above the median value of L_{s+b}. As with the cross section measurement, a central value for the expected significance is calculated, and also the 68% confidence levels either side of the median L_{s+b}.

If the measured cross section is in good agreement with the signal prediction, the observed and expected significances should be compatible within errors.

8.1.3 Results of the statistical analysis

8.1.3.1 tW channel search

Figure 8.1 shows the event yields for the combination of all considered channels scaled to the result of the statistical analysis. The tW cross section was measured to be 33.9 ± 8.6 pb, corresponding to an excess over the background-only hypothesis of 3.6σ. The expected significance of the result was $2.8^{+0.9}_{-0.8}\sigma$.

8.1.3.2 tZq search

Figure 8.2 shows the distribution of the BDT discriminant used as the binned distribution for the statistical analysis. The combination of all considered final states found a cross section for tZq production of 783^{+1000}_{-543} fb. This corresponded to an excess of signal over the background-only hypothesis of 3.432σ. The expected significance for this channel was found to be $1.389 \pm 0.005\sigma$. This cross section measurement is within uncertainties of the SM prediction. At the time of writing, the tZq search was close to beginning the internal approval procedure within the CMS collaboration. It should be noted that the uncertainties from the reweighting of the data-driven background and from matching and Q^2 variations in the WZ generator are not yet propagated into this result. In addition the error quoted on the expected significance is from statistics only, meaning that the final quoted uncertainty will be larger. This means that although the result presented here is not final, it should be very close to what is published by CMS in the future.

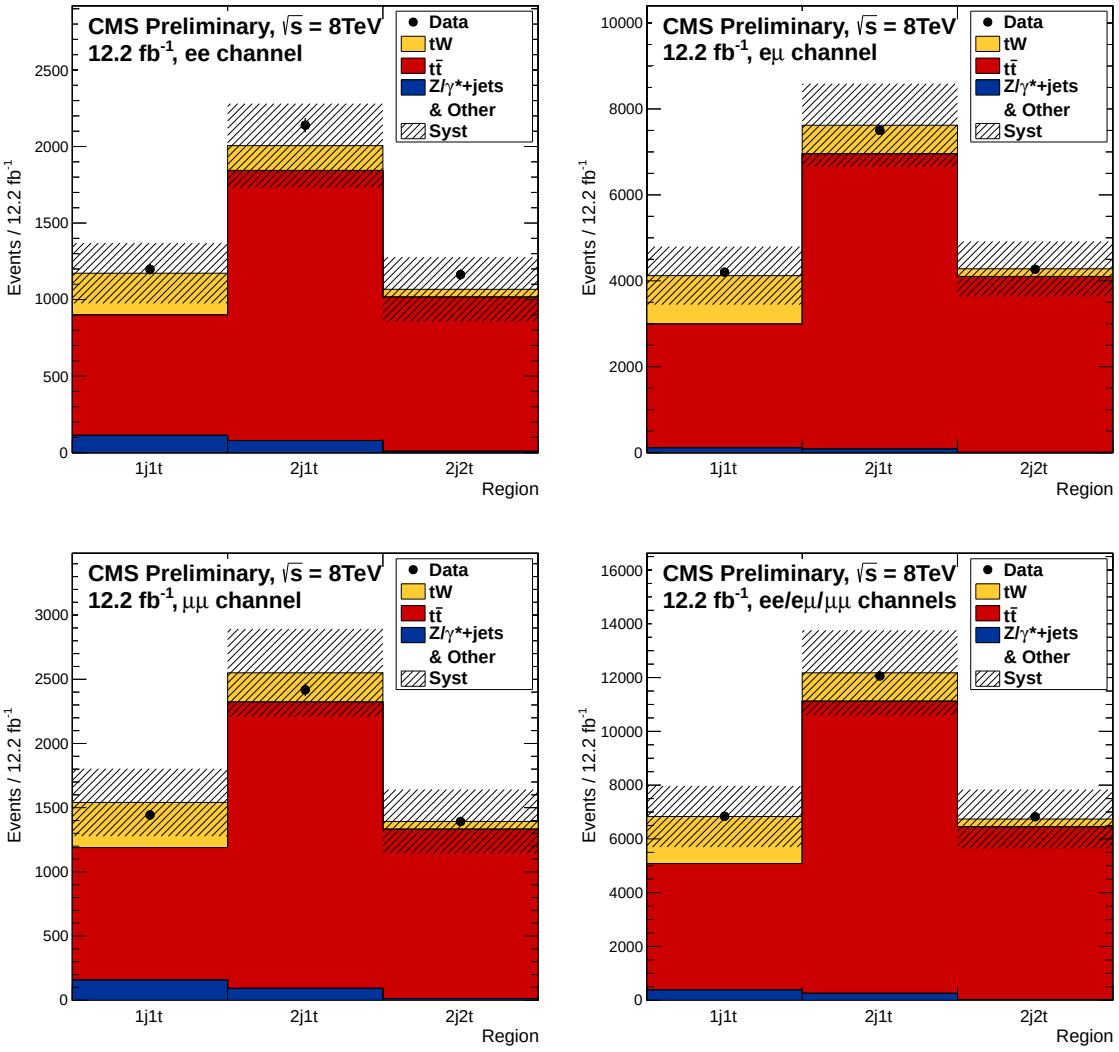


FIGURE 8.1: Comparison between event yields in data and simulation for the tW cross section measurement for the combination of all considered final states. The three bins represent the three sample regions used in the analysis, from left to right the signal (1j1t) region and two $t\bar{t}$ enriched control regions (2j1t and 2j2t). From top-left to bottom-right, the plots show the ee, $e\mu$, $\mu\mu$ and combined final states, respectively. The event yields have been scaled to the outcome of the statistical analysis.

8.2 $|V_{tb}|$ Calculation

The measurement of the tW cross section can be used to test the CKM matrix parameter $|V_{tb}|$, under the assumption that $|V_{td}|$ and $|V_{ts}|$ are much smaller than $|V_{tb}|$, as shown in Section 2.1.2. The strength of the Wtb coupling is directly proportional to the CKM coupling constant, $|V_{tb}|$. In the tW process there are two such couplings: one at the vertex where the top quark is produced and one at the vertex where the top quark decays. This means that the cross section measurement is proportional to the vertex function squared,

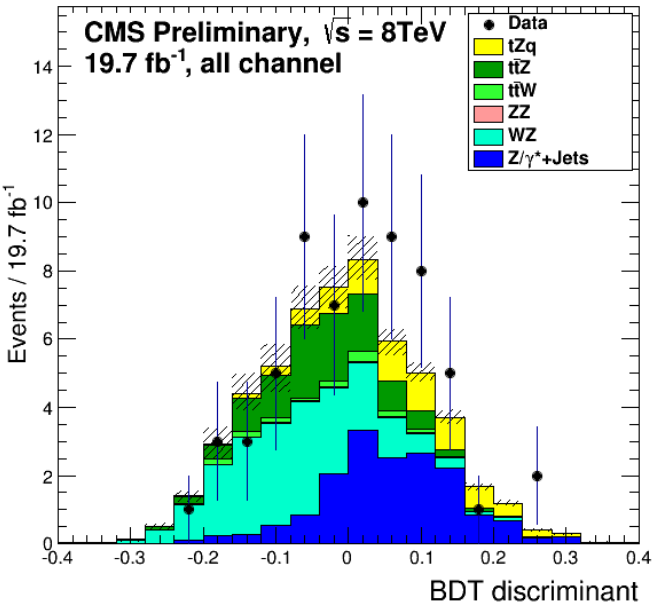


FIGURE 8.2: The distribution of the BDT discriminant for events passing all event selection requirements. The plot shows the combination of all four final states considered.

and so $|V_{tb}|^2$. $|V_{tb}|$ can therefore be calculated as follows:

$$|V_{tb}| = \sqrt{\frac{\sigma_{tW}}{\sigma_{tW}^{th}}} = 1.23^{+0.15}_{-0.17} \tag{8.9}$$

where σ_{tW} is the observed tW cross section and σ_{tW}^{th} is the SM cross section assuming $|V_{tb}| = 1$.

The single top t-channel cross section measured by the CDF and D0 experiments [138] give a value of $|V_{tb}| = 1.03 \pm 0.06$. At the LHC, $|V_{tb}|$ has been measured using the t-channel cross section at $\sqrt{s} = 7$ TeV by CMS [139] and ATLAS [140] to give an average value of $|V_{tb}| = 1.03 \pm 0.05$. At $\sqrt{s} = 8$ TeV the combined ATLAS and CMS result for the t-channel cross section implies a value $|V_{tb}| = 0.99 \pm 0.07$ [141]. The average of these gives $|V_{tb}| = 1.021 \pm 0.032$, a value compatible with the measurement made in the tW measurement.

In addition to a direct measurement of $|V_{tb}|$ a lower bound can be calculated under the Standard Model assumption that $0 \leq |V_{tb}|^2 \leq 1$. The measured tW cross section corresponds to a lower bound at 95% confidence level of $|V_{tb}| > 0.75$. The CDF and D0 have reported 95% confidence level limits of $|V_{tb}| > 0.78$ [142] and $0.99 > |V_{tb}| > 0.90$ [143],

respectively. CMS has presented a similar limit, also at 95% confidence level, based on data collected at $\sqrt{s} = 7$ TeV of $|V_{tb}| > 0.92$ [144].

8.3 Other results from the LHC

The search for the single top t-channel produced in association with a Z boson presented here is the first of its kind, at an LHC experiment or otherwise. As such, there are no competing experimental results with which to compare those obtained here, but they may be compared with the theoretical results originally presented in [51]. The paper predicted that the LHC should have a similar sensitivity to tZq production as for $t\bar{t}Z$. CMS has presented evidence for the $t\bar{t}Z$ channel from data collected during $\sqrt{s} = 8$TeV [49], so for the same running period we should expect to have a similar sensitivity to the tZq. In reality, the different final state topology opens the tZq search up to a wider variety of backgrounds, making it more difficult to isolate the signal sample. Nevertheless, the first evidence for tZq production of a comparable level to that for $t\bar{t}Z$ has been found.

The tW channel, on the other hand, has several other published analyses from the LHC. Both CMS and ATLAS have produced a measurement of the tW cross section using a boosted decision tree (BDT) to separate the tW signal from the $t\bar{t}$ background, and CMS also presents a further analysis that uses the transverse momentum of the tW system as the input to the statistical analysis.

8.3.1 Other CMS analyses

The tW observation paper published by CMS [145] contains a total of three analyses. The cut based search presented in this thesis constituted one of the two cross-check analyses that were included alongside the main BDT analysis.

The idea of using a multivariate technique was initially conceived when it was predicted that a cut based analysis alone would not be able to reach the 5σ significance required to claim an observation of tW production before the start of LS1 in 2013. Being a well understood tool, a BDT was chosen to separate the tW signal from the dominant $t\bar{t}$ background. The first iteration of the BDT [43] was very simple, containing only 4 variables, and relied heavily on the p_T^{sys} variable, the vector sum of the transverse momentum of all objects

within the tW system; the two leptons, $\not{E}_T$ and the single jet. Despite the simplicity there was already a marked gain in significance and an improvement in agreement with the SM compared to the cut based analysis presented alongside it.

For the $\sqrt{s} = 8$ TeV data many additional variables were considered, and a total of 13 variables were eventually chosen based on their separation power and the level of agreement between data and simulation when checked in the control regions. It was discovered that the most powerful discriminators originated in additional, softer jets that didn't pass the signal event selection criteria. For this reason the requirement that there be no additional loose jets in a signal event was removed from the BDT analysis. There was also no additional requirement placed on the H_T of the system in the $e\mu$ channel, due to its limiting effect on the statistics in the BDT.

Besides these two differences in the event selection requirements, the signal and control regions are defined identically to the cut based analysis presented in this thesis. The statistical analysis was carried out in the same way as described in this chapter, but with the BDT discriminant, which can be seen in Figure 8.3, used as the input instead of the raw event yields. The cross section was found to be 23.4 ± 5.4 pb, corresponding to an observed (expected) significance of 6.1σ ($5.4 \pm 1.4\sigma$). This constituted the first observation of the associated production of a single top quark with a $W^{\pm}$ boson. The cross section measurement gave a value of $|V_{tb}| = 1.03 \pm 0.12(\text{exp}) \pm 0.04(\text{th.})$. Under the SM assumption of $0 < |V_{tb}|^2 < 1$, a lower bound at 95% confidence level of $|V_{tb}| > 0.78$ is found.

Two cross check analyses were included in the publication to validate the use of the BDT, one of which was the event yield based search presented in this thesis. By carrying out the same analysis process but extracting the cross section from different distributions the strength of the observed signal in the BDT analysis can be justified. Both cross checks used the same event selection, including the additional soft jet veto and H_T requirement in the $e\mu$ channel. The first chosen variable to be checked, and that presented here, was the event yields of the simulation and data. This 'cut and count' method was expected to be the least sensitive of the analyses, but is a very well established technique that gives reliable and trusted, if somewhat conservative, results. The second cross check analysis was to use the p_T^{sys} variable to extract the tW cross section.

Top quark pair events should contain one jet from each of the top quarks; however, to pass the tW signal requirements one of these jets must not have been properly identified.

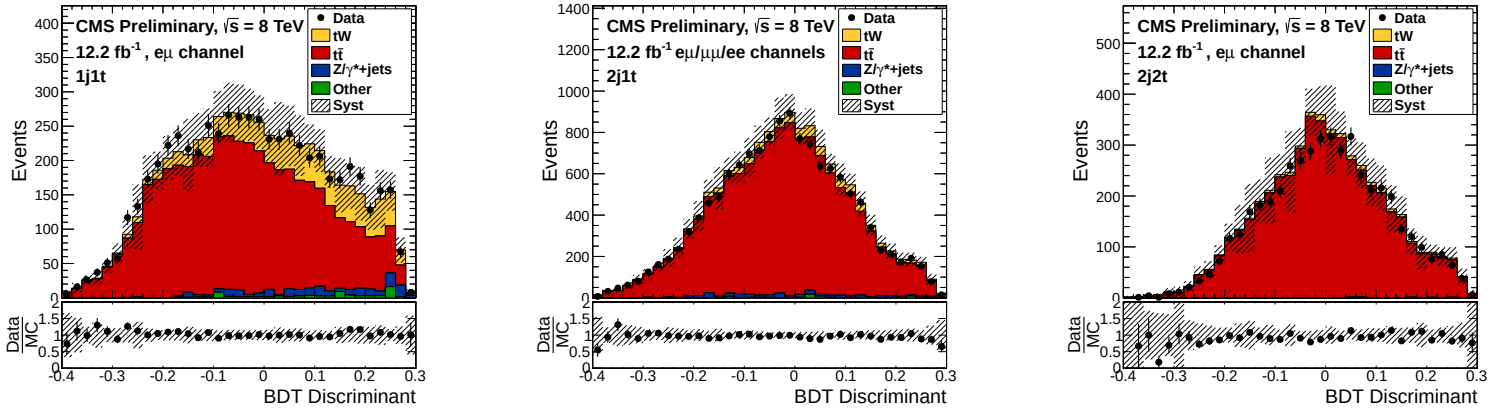


FIGURE 8.3: The BDT discriminant for all considered final states in the three sample regions; the signal (1j1t) region, and the two $t\bar{t}$ control regions (2j1t and 2j2t) [146].

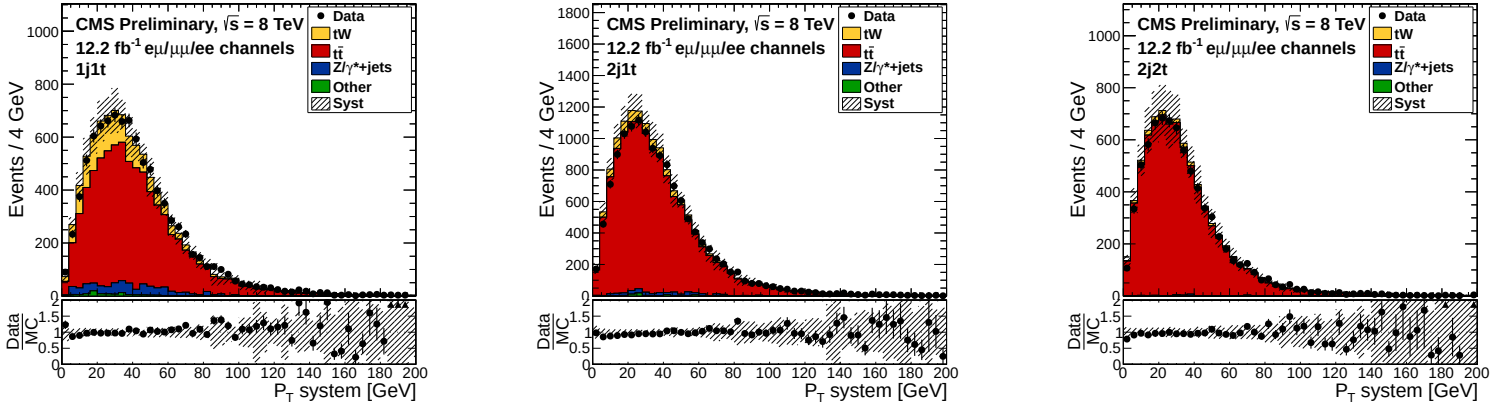


FIGURE 8.4: The transverse momentum of the tW system for all considered final states in the three sample regions; the signal (1j1t) region, and the two $t\bar{t}$ control regions (2j1t and 2j2t) [146].

This jet should have been reconstructed within the detector, so is not included in the $\not{E}_T$ of the event, but by not including it the vector sum of the transverse momentum will be shifted away from zero. In signal tW events all the objects expected to be found have been identified and included in the event topology, so the p_T^{sys} should be close to zero. This gives the p_T^{sys} variable, the distribution of which can be seen in Figure 8.4, excellent discrimination power between tW and $t\bar{t}$ events, and for this reason a cross section extracted from this variable should have a greater significance than from the event yields alone. The measured cross section of 24.3 ± 8.6 pb corresponds to a 4.0σ excess above the background-only hypothesis, compared to an expected significance of $3.2^{+0.4}_{-0.9}\sigma$. This is, as expected, more significant than the analysis based on event yields, whilst having less sensitivity than the BDT analysis.

All three of the cross sections presented in the tW publication are consistent with each other and the SM prediction of 22.2 ± 0.6 (scale) ± 1.4 (PDF) pb. The results obtained for $|V_{tb}|$ from the three analyses are consistent with 1, and therefore the SM.

8.3.2 ATLAS tW results

The ATLAS experiment published the first evidence for the associated production of a single top quark with a $W^{\pm}$ boson in 2012 using 2.05 fb^{-1} of proton-proton collision data at $\sqrt{s} = 7$ TeV [147]. A cross section of 16.8 ± 2.9 (stat.) ± 4.9 (syst.) pb was measured with an observed (expected) significance of 3.3σ (3.4σ).

ATLAS has also presented preliminary results for the cross section, conducted using 20.3 fb^{-1} at $\sqrt{s} = 8$ TeV [148]. A cross section of $\sigma_{Wt} = 27.2 \pm 5.8$ pb was obtained, at a 4.2 σ significance over the background-only hypothesis. The expected significance of the result based on SM predictions was 4.0 σ. From this a value of $|V_{tb}| = 1.10 \pm 0.12$ (exp.) $\pm$ 0.03 (th.) was calculated. Under the SM assumption of $0 < |V_{tb}|^2 < 1$, a lower bound at 95% confidence level of $|V_{tb}| > 0.72$ is found.

The ATLAS BDT results differs from the CMS BDT result in several ways. Firstly, although the ATLAS result takes into account the full dataset available for $\sqrt{s} = 8$ TeV running, it concentrates solely on the $e\mu$ channel, rather than considering all possible combinations of electrons and muons. This channel was chosen because of the large signal yield and low contamination of Z+jets background, reducing dependence on reweighting

techniques for the fake lepton background. Although the $e\mu$ channel proved the most sensitive in the CMS analysis, the overall increase in sensitivity gained from the addition of the ee and $\mu\mu$ channels was found to outweigh the additional systematics and backgrounds introduced.

The second difference is in the definition of the $t\bar{t}$ control regions. Whilst all the CMS analyses define two separate control regions, based on whether one or two of the jets were b-tagged, the ATLAS analysis combines these two regions into one definition: exactly two jets present, either one or both of which are b-tagged.

Finally, the ATLAS analysis uses two separate BDTs, one trained on single jet events and the other trained exclusively on two jet events. This allows the use of different variables that might be more appropriate to each region, for example variables that involve both jets in the control region. There are 19 variables used in the signal region and 20 for the control region, the most powerful of which are found to be those that use the p_T^{sys} variable, defined similarly to that used in the CMS analyses: the vector sum of the transverse momentum of the selected leptons, jets and missing transverse energy of the system.

The cross section and $|V_{tb}|$ measurements presented by the ATLAS experiment are consistent with those published by CMS. The ATLAS result has a similar level of sensitivity to the p_T^{sys} fit analysis conducted within CMS, but is considerably less powerful than the CMS BDT result. The uncertainty on the cross section measurement is similar in both; the CMS BDT cross section measurement has a smaller absolute uncertainty, but the higher central measurement at ATLAS means that the relative uncertainty is less.

Chapter 9

Conclusion

Two analyses were presented in this thesis: the observation and cross section measurement of single top production in association with a W boson, and the search for t-channel single top production in the presence of a radiated Z boson. Both analyses made use of proton proton collisions at $\sqrt{s} = 8$ TeV collected by the CMS detector at the LHC.

Using a subset of the available data from 2012, corresponding to an integrated luminosity of 12.2 fb^{-1}, a cut-based analysis searching for the dilepton tW final state was defined. The cross section was measured at 33.9 ± 8.6 pb, corresponding to an observed excess over the background-only hypothesis with a significance of 3.6σ. The expected significance, taken from simulation only, was calculated to be $2.8^{+0.9}_{-0.8}\sigma$.

This result was published in Physical Review Letters alongside a multivariate analysis that achieved the first observation of tW production [145], which currently represents the leading tW measurement. The cut-based analysis forms a robust cross-check to the main BDT result, validating the use of the multivariate techniques to achieve observation. All the published results were consistent with each other and the SM prediction for the process.

The complete 2012 dataset, corresponding to a total luminosity of 19.7 fb^{-1}, was used for the search for the trilepton tZq final state. A boosted decision tree was used to further separate the tZq signal from the dominant $t\bar{t}Z$ and WZ backgrounds. A cross section of 783^{+1000}_{-543} fb was measured for the signal process, corresponding to an observed excess of data events over the background-only hypothesis of 3.432σ. The simulation-only expected significance of this result was calculated as $1.389 \pm 0.005\sigma$. This constitutes the first

evidence for tZq production at the LHC, with a measured cross section that agrees within errors with the SM prediction.

9.1 Future Measurements

As well as the dilepton tW analysis presented in this thesis, there is currently an investigation under way within CMS to search for the lepton plus jets tW final state: the case in which one W boson decays leptonically and the other hadronically. The additional jets and fewer leptons in the final state make this a more challenging channel to study because of the high level of QCD background, which further complicates estimation of the irreducible $t\bar{t}$ background.

An obvious extension to the dilepton tW analysis would be to include the full available luminosity of the $\sqrt{s} = 8$ TeV dataset. This would increase the dataset by $\sim 50\%$, a substantial increase. This was not done for the original tW observation paper because it was found that the considered subset of data could achieve observation, the primary motive of the paper. In fact, as seen in Chapter 7, the uncertainty in the measurement is dominated by the theoretical uncertainties in the simulation modelling rather than limited statistics. This means that the additional statistics afforded by the larger dataset would not go far to reducing the uncertainty on the cross section measurement. They would, however, allow the measurement of other interesting properties of the channel, such as the ratio of tW^- to $\bar{t}W^+$ events.

The greatest reduction in uncertainty on the cross section measurement would come from improving the modelling uncertainty on the signal and $t\bar{t}$ background samples used in the analysis. One potential improvement being considered is the inclusion of tW and $t\bar{t}$ events in the same event simulation. Whilst this resolves the problem of the interference between the two processes by treating them as one, it would require many more $t\bar{t}$ events than are currently generated to maintain the same number of tW signal events. Additionally, the top quark mass uncertainty, which was problematic throughout the tW analysis, is no longer considered for current analyses. This consideration would help achieve greater accuracy.

Looking ahead to the restart of the LHC, there is much potential for tW analyses. Of the single top channels, the cross section of tW scales most favourably with the increase in

centre-of-mass energy that the restart will bring (see Table 2.2 and Figure 2.4). Where the cross section of top quark pair production increases by a roughly a factor of four, the tW channel cross section is approximately 3.7 times larger and the t-channel cross section increases by less than three times. The s-channel cross section increases the least of all, with only slightly over twice its $\sqrt{s} = 8$ TeV cross section at 14 TeV.

The initial aim will be to re-discover the tW channel, which should be possible within the first few months of the high energy running. Once this has been achieved, finer measurements of the cross section, along with possible differential and fiducial cross section measurements, will be conducted. Eventually the channel will then be used to measure and confirm other properties of the top quark, such as the mass.

As seen in Figure 2.7, the tZq cross section scales similarly to the t-channel single top production with the increasing centre-of-mass energy at the LHC. The leading background processes, $t\bar{t}Z$ and WZ diboson events scale in a similar manner, so the extraction of the signal is not expected to become more difficult at higher energies.

The first aim when studying the tZq channel after LS1 will be to improve the significance of the observed result to a 5σ level, in order to claim observation of the process. This should be achievable solely through the addition of more data, as the low cross section of the process combined with the limited available data make the result largely statistically limited. The accuracy of the cross section measurement will also be improved by larger simulated samples; presently the number of simulated background events containing three leptons is very small, hindering the training of the BDT. Once it is possible to accurately measure a cross section for the tZq process, the couplings of interest that it probes - the WZ and Zt couplings - can be evaluated and compared to the SM theoretical predictions.

Bibliography

[1] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. doi: 10.1016/j.physletb.2012.08.021. URL http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[2] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. doi: 10.1016/j.physletb.2012.08.020. URL http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[3] Katherine Garrett and Gintaras Duda. Dark Matter: A Primer. *Adv.Astron.*, 2011: 968283, 2011. doi: 10.1155/2011/968283.

[4] W.M. Alberico and S.M. Bilenky. Neutrino oscillations, masses and mixing. *Phys.Part.Nucl.*, 35:297–323, 2004.

[5] J. Beringer et al. Review of Particle Physics. *Physical Review D*, 86(1), 2012. doi: 10.1103/physrevd.86.010001. URL http://dx.doi.org/10.1103/PhysRevD.86.010001.

[6] URL http://press.web.cern.ch/backgrounders/top-quark. CERN top quark press relese, accessed May 2015.

[7] The CDF Collaboration. Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab . *Phys. Rev. Lett.*, 74(14):2626–2631, 1995. doi: 10.1103/physrevlett.74.2626. URL http://dx.doi.org/10.1103/PhysRevLett.74.2626.

[8] The D0 Experiment. Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV . *Phys. Rev. Lett.*, 74(13):2422–2426, 1995. doi: 10.1103/physrevlett.74.2422. URL http://dx.doi.org/10.1103/PhysRevLett.74.2422.

[9] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13(9):321–323, 1964. doi: 10.1103/physrevlett.13.321. URL http://dx.doi.org/10.1103/PhysRevLett.13.321.

[10] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132–133, 1964. doi: 10.1016/0031-9163(64)91136-9. URL http://dx.doi.org/10.1016/0031-9163(64)91136-9.

[11] The ATLAS, CDF, CMS, D0 Collaborations. First combination of Tevatron and LHC measurements of the top-quark mass. *arXiv:1403.4427*, 2014. URL http://arxiv.org/abs/1403.4427.

[12] Michael Peskin. *An introduction to quantum field theory*. Addison-Wesley, Reading, Mass, 1995. ISBN 978-0201503975.

[13] Herbert Goldstein. *Classical mechanics*. Addison Wesley, San Francisco, 2002. ISBN 9780201657029.

[14] E. Noether. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918. URL http://eudml.org/doc/59024.

[15] Rym Bouchendira, Pierre Cladé, Saïda Guellati-Khélifa, François Nez, and François Biraben. New Determination of the Fine Structure Constant and Test of the Quantum Electrodynamics. *Phys. Rev. Lett.*, 106(8), 2011. doi: 10.1103/physrevlett.106.080801. URL http://dx.doi.org/10.1103/PhysRevLett.106.080801.

[16] Luis Álvarez-Gaumé and John Ellis. Eyes on a prize particle. *Nature Physics*, 7(1):2–3, 2011. doi: 10.1038/nphys1874. URL http://dx.doi.org/10.1038/nphys1874.

[17] Richard Feynman. *High energy collisions : Third International Conference held at Stony Brook, N.Y.* Gordon and Breach, New York London, 1969. ISBN 978-0-677-13950-0.

[18] V.N. Gribov and L.N. Lipatov. Deep inelastic ep scattering in perturbation theory. *Sov.J.Nucl.Phys.*, 15:438–450, 1972.

[19] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl.Phys.*, B126:298, 1977. doi: 10.1016/0550-3213(77)90384-4.

[20] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov.Phys.JETP*, 46:641–653, 1977.

[21] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C.-P. Yuan. Implications of CTEQ global analysis for collider observables. *Physical Review D*, 78(1), 2008. doi: 10.1103/physrevd.78.013004. URL http://dx.doi.org/10.1103/PhysRevD.78.013004.

[22] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys.Rept.*, 97:31–145, 1983. doi: 10.1016/0370-1573(83)90080-7.

[23] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. A Modified cluster hadronization model. *Eur.Phys.J.*, C36:381–395, 2004. doi: 10.1140/epjc/s2004-01960-8.

[24] J. Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *The European Physical Journal C*, 53(3):473–500, 2007. doi: 10.1140/epjc/s10052-007-0490-5. URL http://dx.doi.org/10.1140/epjc/s10052-007-0490-5.

[25] Makoto Kobayashi and Toshihide Maskawa. C P -Violation in the Renormalizable Theory of Weak Interaction . *Prog. Theor. Phys.*, 49(2):652–657, 1973. doi: 10.1143/ptp.49.652. URL http://dx.doi.org/10.1143/PTP.49.652.

[26] M. L. Perl et al. Evidence for Anomalous Lepton Production in e+ - e- Annihilation . *Phys. Rev. Lett.*, 35(22):1489–1492, 1975. doi: 10.1103/physrevlett.35.1489. URL http://dx.doi.org/10.1103/PhysRevLett.35.1489.

[27] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39(5):252–255, 1977. doi: 10.1103/physrevlett.39.252. URL http://dx.doi.org/10.1103/PhysRevLett.39.252.

[28] The UA1 Collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at UA1. *Physics Letters B*, 122(1):103–116, 1983. doi: 10.1016/0370-2693(83)91177-2. URL http://dx.doi.org/10.1016/0370-2693(83)91177-2.

[29] The UA2 Collaboration. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider. *Physics Letters B*, 122(5–6):476 – 485, 1983. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/0370-2693(83)91605-2. URL http://www.sciencedirect.com/science/article/pii/0370269383916052.

[30] LEP Electroweak Working Group. Precision Electroweak Measurements and Constraints on the Standard Model. 2010.

[31] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, R. Kogler, K. Mönig, M. Schott, and J. Stelzer. The electroweak fit of the standard model after the discovery of a new boson at the LHC. *The European Physical Journal C*, 72(11), 2012. doi: 10.1140/epjc/s10052-012-2205-9. URL http://dx.doi.org/10.1140/epjc/s10052-012-2205-9.

[32] Giuseppe Degrassi, Stefano Di Vita, Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, et al. Higgs mass and vacuum stability in the Standard Model at NNLO. *JHEP*, 1208:098, 2012. doi: 10.1007/JHEP08(2012)098.

[33] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Hoche, et al. General-purpose event generators for LHC physics. *Phys.Rept.*, 504:145–233, 2011. doi: 10.1016/j.physrep.2011.03.005.

[34] Kirill Melnikov and Timo van Ritbergen. The Three loop relation between the MS-bar and the pole quark masses. *Phys.Lett.*, B482:99–108, 2000. doi: 10.1016/S0370-2693(00)00507-4.

[35] The CMS Collaboration. Measurement of top quark polarization in t-channel single-top production. Technical Report CMS-PAS-TOP-13-001, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1601800.

[36] The CMS Collaboration. Measurement of the W boson helicity in events with a single reconstructed top quark in pp collisions at $\sqrt{s}$=8 TeV. *JHEP*, 1501:053, 2015. doi: 10.1007/JHEP01(2015)053.

[37] Nikolaos Kidonakis. Differential and total cross sections for top pair and single top production. pages 831–834, 2012. doi: 10.3204/DESY-PROC-2012-02/251.

[38] Luca Fiorini. Top-Quark Physics Results From LHC. *arXiv:1201.5844*, 2012. URL http://arxiv.org/abs/1201.5844.

[39] . URL http://cms.web.cern.ch/news/measurement-t-channel-single-top/quark-production-rates-pp-collisions-7-tev. CMS measurement of t-channel press release, accessed May 2015.

[40] The CDF and D0 Collaborations. Observation of s -Channel Production of Single Top Quarks at the Tevatron . *Phys. Rev. Lett.*, 112(23), 2014. doi: 10.1103/physrevlett.112.231803. URL http://dx.doi.org/10.1103/PhysRevLett.112.231803.

[41] The D0 Collaboration. Observation of Single Top-Quark Production. *Phys. Rev. Lett.*, 103(9), 2009. doi: 10.1103/physrevlett.103.092001. URL http://dx.doi.org/10.1103/PhysRevLett.103.092001.

[42] . URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures. Public Top Summary Figures, accessed May 2015.

[43] The CMS Collaboration. Evidence for Associated Production of a Single Top Quark and W Boson in $p\bar{p}$ Collisions at $\sqrt{s} = 7$ TeV . *Phys. Rev. Lett.*, 110(2), 2013. doi: 10.1103/physrevlett.110.022003. URL http://dx.doi.org/10.1103/PhysRevLett.110.022003.

[44] The ATLAS Collaboration. Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV. *Physics Letters B*, 716(1):142–159, 2012. doi: 10.1016/j.physletb.2012.08.011. URL http://dx.doi.org/10.1016/j.physletb.2012.08.011.

[45] Stefano Frixione, Eric Laenen, Patrick Motylinski, Chris White, and Bryan R Webber. Single-top hadroproduction in association with a W boson . *J. High Energy Phys.*, 2008(07):029–029, 2008. doi: 10.1088/1126-6708/2008/07/029. URL http://dx.doi.org/10.1088/1126-6708/2008/07/029.

[46] Paolo Nason and Bryan Webber. Next-to-Leading-Order Event Generators. *Ann.Rev.Nucl.Part.Sci.*, 62:187–213, 2012. doi: 10.1146/annurev-nucl-102711-094928.

[47] M. Nowakowski and A. Pilaftsis. On gauge invariance of Breit-Wigner propagators. *Zeitschrift für Physik C Particles and Fields*, 60(1):121–125, 1993. doi: 10.1007/bf01650437. URL http://dx.doi.org/10.1007/BF01650437.

[48] Achilleas Lazopoulos, Thomas McElmurry, Kirill Melnikov, and Frank Petriello. Next-to-leading order QCD corrections to $t\bar{t}Z$ production at the LHC. *Phys.Lett.*, B666:62–65, 2008. doi: 10.1016/j.physletb.2008.06.073.

[49] The CMS Collaboration. Measurement of top quark-antiquark pair production in association with a W or Z boson in pp collisions at $\sqrt{s} = 8$ TeV. *The European Physical Journal C*, 74(9), 2014. doi: 10.1140/epjc/s10052-014-3060-7. URL http://dx.doi.org/10.1140/epjc/s10052-014-3060-7.

[50] The CMS Collaboration. Measurement of the inclusive top-quark pair + photon production cross section in the muon + jets channel in pp collisions at 8 TeV. Technical Report CMS-PAS-TOP-13-011, CERN, Geneva, 2014. URL http://cds.cern.ch/record/1644573.

[51] John Campbell, R. Keith Ellis, and Raoul Röntsch. Single top production in association with a Z boson at the LHC. *Phys.Rev.*, D87(11):114006, 2013. doi: 10.1103/PhysRevD.87.114006.

[52] The CMS Collaboration. Measurement of tqZ and search for FCNC in single top events. . Paper in preparation.

[53] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN, Geneva, 2004. URL https://cds.cern.ch/record/782076.

[54] URL https://www.stfc.ac.uk/3387.aspx. CERN accelerator complex, accessed May 2015.

[55] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *J. Inst.*, 3(08):S08003–S08003, 2008. doi: 10.1088/1748-0221/3/08/s08003. URL http://dx.doi.org/10.1088/1748-0221/3/08/S08003.

[56] The CMS Collaboration. The CMS experiment at the CERN LHC. *J. Inst.*, 3(08): S08004–S08004, 2008. doi: 10.1088/1748-0221/3/08/s08004. URL http://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[57] Jr. Alves, A. Augusto et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. doi: 10.1088/1748-0221/3/08/S08005.

[58] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *J. Inst.*, 3 (08):S08002–S08002, 2008. doi: 10.1088/1748-0221/3/08/s08002. URL http://dx.doi.org/10.1088/1748-0221/3/08/S08002.

[59] Karlheinz Schindl. The injector chain for the LHC. (CERN-OPEN-99-052):6 p, 1999. URL http://cds.cern.ch/record/397574.

[60] Geoff Brumfiel. Eight-month delay for LHC. *Nature*, 455(7216):1015–1015, 2008. doi: 10.1038/4551015a. URL http://dx.doi.org/10.1038/4551015a.

[61] R et al. Alemany-Fernandez. Operation and Configuration of the LHC in Run 1. 2013. URL https://cds.cern.ch/record/1631030.

[62] URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Public CMS luminosity results, accessed May 2015.

[63] URL http://cms.web.cern.ch/news/cms-detector-design. CMS detector diagram, accessed May 2015.

[64] The CMS Collaboration. CMS, tracker technical design report. 1998.

[65] D. Sprenger, M. Weber, R. Adolphi, R. Brauer, L. Feld, et al. Validation of Kalman Filter alignment algorithm with cosmic-ray data using a CMS silicon strip tracker endcap. *J. Inst.*, 5:P06007, 2010. doi: 10.1088/1748-0221/5/06/P06007.

[66] The CMS Collaboration. CMS tracking performance results from early LHC operation. *The European Physical Journal C*, 70(4):1165–1192, 2010. doi: 10.1140/epjc/s10052-010-1491-3. URL http://dx.doi.org/10.1140/epjc/s10052-010-1491-3.

[67] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *J. Inst.*, 9(10):P10009–P10009, 2014. doi: 10.1088/1748-0221/9/10/p10009. URL http://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[68] The CMS Collaboration. CMS: The electromagnetic calorimeter. Technical design report. 1997.

[69] A Benaglia. The CMS ECAL performance with examples. *J. Inst.*, 9(02):C02008–C02008, 2014. doi: 10.1088/1748-0221/9/02/c02008. URL http://dx.doi.org/10.1088/1748-0221/9/02/C02008.

[70] The CMS Collaboration. Energy resolution of the barrel of the CMS Electromagnetic Calorimeter. *J. Inst.*, 2(04):P04004, 2007. URL http://stacks.iop.org/1748-0221/2/i=04/a=P04004.

[71] The CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. URL https://cds.cern.ch/record/357153.

[72] The CMS Collaboration. The CMS-HF quartz fiber calorimeters. *J. Phys.: Conf. Ser.*, 160:012014, 2009. doi: 10.1088/1742-6596/160/1/012014. URL http://dx.doi.org/10.1088/1742-6596/160/1/012014.

[73] A Penzo, Y Onel, and the CMS Collaboration. The CMS-HF quartz fiber calorimeters. *J. Phys.: Conf. Ser.*, 160:012014, apr 2009. doi: 10.1088/1742-6596/160/1/012014. URL http://dx.doi.org/10.1088/1742-6596/160/1/012014.

[74] CMS Collaboration. Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data. *J. Inst.*, 5(03):T03012–T03012, 2010. doi: 10.1088/1748-0221/5/03/t03012. URL http://dx.doi.org/10.1088/1748-0221/5/03/T03012.

[75] The CMS Collaboration. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. *The European Physical Journal C*, 55(1):159–171, 2008. doi: 10.1140/epjc/s10052-008-0573-y. URL http://dx.doi.org/10.1140/epjc/s10052-008-0573-y.

[76] The CMS Collaboration. Design, performance, and calibration of CMS forward calorimeter wedges. *The European Physical Journal C*, 53(1):139–166, 2007. doi: 10.1140/epjc/s10052-007-0459-4. URL http://dx.doi.org/10.1140/epjc/s10052-007-0459-4.

[77] The CMS Collaboration. CMS, the Compact Muon Solenoid. Muon technical design report. 1997.

[78] Min Suk Kim et al. CMS reconstruction improvement for the muon tracking by the RPC chambers. *PoS*, RPC2012:045, 2012. doi: 10.1088/1748-0221/8/03/T03001.

[79] The CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s}$ = 7 TeV . *J. Inst.*, 7(10):P10002–P10002, 2012. doi: 10.1088/1748-0221/7/10/p10002. URL http://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[80] The TOTEM Collaboration. Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at s = 8 TeV . *Phys. Rev. Lett.*, 111(1), 2013. doi: 10.1103/physrevlett.111.012001. URL http://dx.doi.org/10.1103/PhysRevLett.111.012001.

[81] The CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS. . URL https://cds.cern.ch/record/706847.

[82] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002. URL https://cds.cern.ch/record/578006.

[83] Rene Brun and Fons Rademakers. ROOT - An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389(1-2):81–86, 1997. doi: 10.1016/s0168-9002(97)00048-x. URL http://dx.doi.org/10.1016/S0168-9002(97)00048-X.

[84] Federico De Guio. The CMS data quality monitoring software: experience and future prospects. *J.Phys.Conf.Ser.*, 513:032024, 2014. doi: 10.1088/1742-6596/513/3/032024.

[85] The CMS Collaboration. Status of the CMS Detector Control System. *J. Phys.: Conf. Ser.*, 396(1):012023, 2012. doi: 10.1088/1742-6596/396/1/012023. URL http://dx.doi.org/10.1088/1742-6596/396/1/012023.

[86] V Rapsevicius. CMS Run Registry: Data Certification Bookkeeping and Publication System. *J. Phys.: Conf. Ser.*, 331(4):042038, 2011. doi: 10.1088/1742-6596/331/4/042038. URL http://dx.doi.org/10.1088/1742-6596/331/4/042038.

[87] M Pernicka, D Kotlinski, W Johns, H Steininger, and H Schmid. The CMS Pixel FED. 2007. URL https://cds.cern.ch/record/1091743.

[88] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, 2009. URL https://cds.cern.ch/record/1194487.

[89] Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, 2010. URL https://cds.cern.ch/record/1247373.

[90] GEANT4 Collaboration. Geant4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506(3):250–303, 2003. doi: 10.1016/s0168-9002(03)01368-8. URL http://dx.doi.org/10.1016/S0168-9002(03)01368-8.

[91] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys.Rev.*, D82:074018, 2010. doi: 10.1103/PhysRevD.82.074018.

[92] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *JHEP*, 2006(05):026, 2006. URL http://stacks.iop.org/1126-6708/2006/i=05/a=026.

[93] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 1006:043, 2010. doi: 10.1007/JHEP06(2010)043.

[94] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *JHEP*, 2011(6):128, 2011. doi: 10.1007/JHEP06(2011)128.

[95] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP*, 0701:013, 2007. doi: 10.1088/1126-6708/2007/01/013.

[96] Z. Was. TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report. *Nucl.Phys.Proc.Suppl.*, 98:96–102, 2001. doi: 10.1016/S0920-5632(01)01200-2.

[97] Michał Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Computer Physics Communications*, 185 (11):2930–2938, 2014. doi: 10.1016/j.cpc.2014.06.021. URL http://dx.doi.org/10.1016/j.cpc.2014.06.021.

[98] P. Kant, O.M. Kind, T. Kintscher, T. Lohse, T. Martini, et al. HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Comput.Phys.Commun.*, 191:74–89, 2015. doi: 10.1016/j.cpc.2015.02.001.

[99] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order. *Comput.Phys.Commun.*, 182:2388–2403, 2011. doi: 10.1016/j.cpc.2011.06.008.

[100] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. *Nucl.Phys.Proc.Suppl.*, 205-206:10–15, 2010. doi: 10.1016/j.nuclphysbps.2010.08.011.

[101] Pierre Billoir. Progressive track recognition with a Kalman-like fitting procedure. *Computer Physics Communications*, 57(1-3):390–394, 1989. doi: 10.1016/0010-4655(89)90249-x. URL http://dx.doi.org/10.1016/0010-4655(89)90249-X.

[102] W Erdmann. Vertex reconstruction at the CMS experiment. *J. Phys.: Conf. Ser.*, 110(9):092009, 2008. doi: 10.1088/1742-6596/110/9/092009. URL http://dx.doi.org/10.1088/1742-6596/110/9/092009.

[103] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, 2009. URL http://cds.cern.ch/record/1194487.

[104] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, 2005. doi: 10.1088/0954-3899/31/9/n01. URL http://dx.doi.org/10.1088/0954-3899/31/9/N01.

[105] The CMS Collaboration. Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV. Technical Report CMS-PAS-PFT-10-003, CERN, 2010. URL http://cds.cern.ch/record/1279347.

[106] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl. Instrum. Meth. A*, 543(2-3):577–584, 2005. doi: 10.1016/j.nima.2004.12.018. URL http://dx.doi.org/10.1016/j.nima.2004.12.018.

[107] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti- k t jet clustering algorithm . *J. High Energy Phys.*, 2008(04):063–063, 2008. doi: 10.1088/1126-6708/2008/04/063. URL http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[108] Stephen D. Ellis. Successive combination jet algorithm for hadron collisions. *Physical Review D*, 48(7):3160–3166, 1993. doi: 10.1103/physrevd.48.3160. URL http://dx.doi.org/10.1103/PhysRevD.48.3160.

[109] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *J. Inst.*, 6(11):P11002–P11002, 2011. doi: 10.1088/1748-0221/6/11/p11002. URL http://dx.doi.org/10.1088/1748-0221/6/11/P11002.

[110] Serguei Chatrchyan et al. Missing transverse energy performance of the CMS detector. *J. Inst.*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.

[111] Saranya Samik Ghosh. Performance of MET reconstruction and pileup mitigation techniques in CMS. 2015.

[112] The CMS Collaboration. Identification of b-quark jets with the CMS experiment. *J. Inst.*, 8:P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013.

[113] The CMS Collaboration. Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, tbart and boosted topology events. *CMS PAS BTV-13-001*, 2013. URL http://cms-physics.web.cern.ch/cms-physics/public/BTV-13-001-pas.pdf.

[114] Prerit Jaiswal, Karoline Kopp, and Takemichi Okui. Higgs Production Amidst the LHC Detector. *Phys.Rev.*, D87(11):115017, 2013. doi: 10.1103/PhysRevD.87.115017.

[115] The CMS Collaboration. Results on b-tagging identification in 8 TeV pp collisions. Feb 2013. URL https://cds.cern.ch/record/1528183.

[116] Jet Performance in pp Collisions at 7 TeV. Technical Report CMS-PAS-JME-10-003, CERN, Geneva, 2010. URL https://cds.cern.ch/record/1279362.

[117] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[118] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive Logistic Regression: a Statistical View of Boosting. *Annals of Statistics*, 28:2000, 1998.

[119] Andreas Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, et al. TMVA - Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[120] CMS Collaboration. Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update. 2012.

[121] Simon White. Luminosity Scans at the LHC. Luminosity Scans at LHC. 2011. URL https://cds.cern.ch/record/1357865.

[122] CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1598864.

[123] The CMS Collaboration. Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 2011(1), 2011. doi: 10.1007/jhep01(2011)080. URL http://dx.doi.org/10.1007/JHEP01(2011)080.

[124] The CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. 2015.

[125] The CMS collaboration. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC . *J. Inst.*, 8(11):P11002–P11002, 2013. doi: 10.1088/1748-0221/8/11/p11002. URL http://dx.doi.org/10.1088/1748-0221/8/11/P11002.

[126] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C.-P. Yuan. CT10 next-to-next-to-leading order global

analysis of QCD. *Physical Review D*, 89(3), 2014. doi: 10.1103/physrevd.89.033009. URL http://dx.doi.org/10.1103/PhysRevD.89.033009.

[127] M.R. Whalley, D. Bourilkov, and R.C. Group. The Les Houches accord PDFs (LHAPDF) and LHAGLUE. 2005.

[128] The CMS Collaboration. 8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb of data in CMS. *CMS Public Note DP-2013-033*, 2013. URL https://cds.cern.ch/record/1627305/.

[129] The CMS Collaboration. Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV. *CMS PAS JME-10-014*, 2011. URL http://inspirehep.net/record/925240/.

[130] Angela Barbaro Galtieri, Fabrizio Margaroli, and Igor Volobouev. Precision measurements of the top quark mass from the Tevatron in the pre-LHC era. *Rep. Prog. Phys.*, 75(5):056201, 2012. doi: 10.1088/0034-4885/75/5/056201. URL http://dx.doi.org/10.1088/0034-4885/75/5/056201.

[131] The CMS Collaboration. Combination of the CMS top-quark mass measurements from Run 1 of the LHC. *CMS PAS TOP-14-015*, 2014. URL http://cds.cern.ch/record/1951019.

[132] The CMS Collaboration. Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV. *The European Physical Journal C*, 73(3), 2013. doi: 10.1140/epjc/s10052-013-2339-4. URL http://dx.doi.org/10.1140/epjc/s10052-013-2339-4.

[133] The CMS Collaboration. Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV. *The European Physical Journal C*, 73(3), 2013. doi: 10.1140/epjc/s10052-013-2339-4. URL http://dx.doi.org/10.1140/epjc/s10052-013-2339-4.

[134] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001. doi: 10.1088/1126-6708/2001/01/010.

[135] Wolfgang A. Rolke, Angel M. López, and Jan Conrad. Limits and confidence intervals in the presence of nuisance parameters. *Nucl. Instrum. Meth. A*, 551(2-3):493–503, 2005. doi: 10.1016/j.nima.2005.05.068. URL http://dx.doi.org/10.1016/j.nima.2005.05.068.

[136] Alexander L. Read. Presentation of search results: The CL(s) technique. *J.Phys.*, G28:2693–2704, 2002. doi: 10.1088/0954-3899/28/10/313.

[137] The CMS Collaboration. Search for Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s}$=7 TeV. Technical Report CMS-PAS-HIG-11-029, CERN, Geneva, 2011. URL http://cds.cern.ch/record/1406353.

[138] The D0 Collaboration. Evidence for s-channel single top quark production in collisions at. *Physics Letters B*, 726(4-5):656–664, 2013. doi: 10.1016/j.physletb.2013.09.048. URL http://dx.doi.org/10.1016/j.physletb.2013.09.048.

[139] The CMS Collaboration. Measurement of the single-top-quark t-channel cross section in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 2012(12), 2012. doi: 10.1007/jhep12(2012)035. URL http://dx.doi.org/10.1007/JHEP12(2012)035.

[140] The ATLAS Collaboration. Measurement of the t-channel single top-quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys.Lett.*, B717:330–350, 2012. doi: 10.1016/j.physletb.2012.09.031.

[141] The ATLAS and CMS Collaborations. Combination of single top-quark cross-sections measurements in the t-channel at $\sqrt{s} = 8$ TeV with the ATLAS and CMS experiments. Technical Report CMS-PAS-TOP-12-002. ATLAS-CONF-2013-098, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1601029.

[142] The CDF Collaboration. Measurement of b (t $\rightarrow$ w b) / b (t $\rightarrow$ w q) at the collider detector at fermilab. *Phys. Rev. Lett.*, 95(10), aug 2005. doi: 10.1103/physrevlett.95.102002. URL http://dx.doi.org/10.1103/PhysRevLett.95.102002.

[143] The D0 Collaboration. Precision measurement of the ratio b (t $\rightarrow$ w b) / b (t $\rightarrow$ w q) and extraction of v tb. *Phys. Rev. Lett.*, 107(12), sep 2011. doi: 10.1103/physrevlett.107.121802. URL http://dx.doi.org/10.1103/PhysRevLett.107.121802.

[144] CMS Collaboration. First measurement of B(t -> Wb)/B(t -> Wq) in the dilepton channel in pp collisions at sqrt(s)=7 TeV. Technical Report CMS-PAS-TOP-11-029, CERN, Geneva, 2012. URL http://cds.cern.ch/record/1429972.

[145] The CMS Collaboration. Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s}$ =8 TeV. *Phys. Rev. Lett.*, 112(23), 2014. doi: 10.1103/physrevlett.112.231802. URL http://dx.doi.org/10.1103/PhysRevLett.112.231802.

[146] URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOP12040. CMS tW observation public web page, accessed May 2015.

[147] The ATLAS Collaboration. Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s}$ = 7 TeV. *Phys.Lett.*, B716:142–159, 2012. doi: 10.1016/j.physletb.2012.08.011.

[148] The ATLAS Collaboration. Measurement of the cross-section for associated production of a top quark and a W boson at $\sqrt{s}$ = 8 TeV with the ATLAS detector. Technical Report ATLAS-CONF-2013-100, CERN, Geneva, 2013. URL http://cds.cern.ch/record/1600799.