Measurement of the single top tW associated production cross section at 7 TeV with 4.9 fb^{-1} of data in the dilepton decay channel at the LHC using the CMS detector

William P Martin

A thesis submitted for the degree of
Doctor of Philosophy

Centre for Sensors and Instrumentation
School of Engineering and Design
Brunel University, Uxbridge

Thursday 30th May, 2013

Abstract

The first measurement of the production cross section for a single top quark in association with a $W^{\pm}$ boson with the CMS experiment at the Large Hadron Collider is presented. The analysed data is from proton-proton collisions at a centre-of-mass energy of 7 TeV, corresponding to an integrated luminosity of 4.9 fb^{-1}. The analysis focuses on the leptonic final states of the process and implements a cut-based method. The final state consists of two leptons, two neutrinos and a jet from the fragmentation of a b-quark. The primary source of background events arises from $t\bar{t}$ production, followed by Z/γ^{*} processes. The observed signal has a significance of 3.5σ and corresponds to a cross section measurement of $14.9^{+5.0}_{-5.1}$ (stat $\oplus$ syst) pb, in agreement with the Standard Model expectations.

Declaration

I hereby declare that no part of this thesis has been previously submitted to this or any other university as part of the requirement for a higher degree. The work described herein was conducted solely by the undersigned except for those colleagues and other workers acknowledged in the text.

William P Martin
Thursday 30th May, 2013

Acknowledgements

The general perception of physics is that it is male-dominated subject. From my personal experience I have seen that to be true, from secondary school education up to and including my PhD. However, during the 4 years of my postgraduate research I have worked almost exclusively with female physicists, without whom, I would not have a thesis to write.

In particular, I want to thank *Rebeca Gonzalez Suarez* who started the CMS tW-channel cross section measurement and with whom I worked with to complete the work presented here. Her expertise, support and patience were invaluable during the last 17 months. If I were to continue on to a post-doc position, I would take Rebeca as my role-model. I would also like to thank my acting supervisor *Joanne Cole* for always being available for advice when needed. She deserves special appreciation for going through my thesis with a fine-toothed comb more than once. The quality of this document is thanks in no small part to her.

In terms of male physicists, I would like to thank my supervisor at Brunel *Peter Hobson*. I would be remiss if I did not thank *Ivan Reid* for his invaluable technical support throughout the last 4 years, and *Sam Harper* for technical advice (and recommending ski fitness classes) whilst working at CERN. I would also like to mention *Andrea Giammanco*, whose presentation on $\mathrm{t\bar{t}}$/tW interference was instrumental in my move into single top physics.

Additionally, I would like to acknowledge the community of PhD students working at CERN, a wide range of interesting people from all over the globe who made my LTA such a memorable experience. In particular, I should mention *Flavia Dias* for reminding me of my passion for physics, and that complaining about something does not mean you hate it.

I would also like to thank the Science and Technologies Funding Council for supporting me financially during my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Standard Model (SM) of particle physics describes how elementary particles interact with each other. All the predictions of this theory have been verified with an astonishing precision. As of this year, the long sought for Higgs boson can be included in this statement, with compelling analyses from both of the LHC's general purpose detectors observing a new boson compatible with SM prediction [1, 2]. Although the SM is a remarkably self-consistent theory, and has undergone much scrutiny, there are indications that it is not the fundamental theory at the Planck scale, which is a much higher energy scale than is currently accessible in the laboratory. The need to fine tune the theory to account for the difference between the grand unification scale and the electroweak scale, where it remains an effective theory, leads to the so-called hierarchy problem [3]. In order to solve this problem, along with other shortcomings, new physics beyond the SM is required.

As the top quark is the heaviest known fermion and fermions interact with the Higgs boson through the Yukawa coupling that is proportional to the fermion mass, the large mass of the top quark makes it a unique probe into physics at the electroweak scale [3]. Additionally, the top quark decays before it can hadronise, meaning that it is the only known quark that can be studied in an unbound state through its decay products. With about one top quark pair produced per second at the design luminosity and unprecedented high centre-of-mass energy of 14 TeV, the Large Hadron Collider (LHC)

provides unparalleled quantities of top quarks, allowing detailed studies of the fermion's properties. This has enabled searches for new physics in top decays, and may eventually lead to a deeper understanding of electroweak symmetry breaking and the origin of mass.

The subject of this thesis is the measurement of single top quark production with an associated $W^{\pm}$ boson, using 4.9 fb^{-1} of data collected with the CMS detector. In this analysis, the leptonic decays of the $W^{\pm}$ boson, $W \to l\nu$, and top quark, $t \to bW \to jl\nu$, are studied in three final states. (Only states containing electrons and muons are considered: ee, $e\mu$ and $\mu\mu$; decays involving τ are simulated but not directly studied). Therefore, signal events are characterised by the presence of two leptons, two neutrinos and a jet originating from the decay of a b-quark.

This particular channel proves challenging from an analysis perspective as it has a very small cross section when compared with the different backgrounds. This is especially so in the case of top-antitop ($t\bar{t}$) production, where in addition to having a cross section approximately 10 times that of the signal process, it also has an almost identical signature. Disentangling this prominent background source represents the main challenge of the analysis.

This thesis is structured in the following way. Chapter 2 gives a brief introduction to the Standard Model of particle physics. The production and the decay of top quarks are explained, as is the relevance of top quark physics for the discovery of the Higgs boson. Chapter 3 describes the experimental setup of the Compact Muon Solenoid (CMS) detector. This includes a brief summary of the specifications of the LHC and a detailed description of the CMS sub-detectors. Having explained how signals are measured in the CMS detector systems, Chapter 4 goes on to explain how these signals are used to identify elementary particles. Descriptions are given of how the reconstruction of the leptons, jets, and transverse momentum imbalance works, as well as the b-jet identification procedure. The top quark events studied in this analysis need to be extracted from the huge amount of recorded

events. Chapter 5 describes the necessary steps to select the events of interest. Specifically, it explains how the kinematic properties of the signal process topology are exploited to reject background processes and retain the desired tW production events. Details of the systematic uncertainties related to the analysis are presented in Chapter 6. Whilst the final result is presented in Chapter 7. The statistical methods used to evaluate the production cross section and signal significance are explained here, along with the calculated results. In addition, an estimate of the corresponding CKM matrix element, $|V_{tb}|$, is given. Finally a summary about all the work achieved in this thesis is given in Chapter 8.

Chapter 2

Theory

2.1 The Standard Model

The Standard Model (SM) of particle physics is a theory that encompasses all known fundamental particles and their interactions with the electromagnetic, weak and strong nuclear forces. The SM incorporates three generations of quarks and leptons, four gauge bosons, which mediate the interactions with the forces, and the (SM) Higgs boson. The Higgs boson is the smallest excitation of the Higgs field, the means by which elementary massive particles acquire their mass.

Each generation in the SM contains two quarks, a lepton and it's associated lepton neutrino. The six known quarks are the up, down, charm, strange, bottom and top. The three known leptons are the electron, muon and tau. The four gauge bosons are the photon, gluon, Z^0 and $W^{\pm}$.

The *Electromagnetic Force* occurs between all electrically charged particles and, of the fundamental forces, is the one that is best understood. It is mediated by the photon, which has no mass and hence an infinite range. The strength of the electromagnetic force is characterised by the *fine structure constant*: $\alpha = e^2/\hbar c = \frac{1}{137}$. When an electric charge moves, it has an associated electric field and magnetic field. A charged particle will be attracted (repelled) along the direction of the electrical field, and the magnetic field

will subject the charged particle to a force perpendicular to the motion of the charge and the direction of the magnetic field. These phenomena are utilised to identify particles, and are a crucial component of particle detector design (see Chapter 3). Within the SM the electromagnetic force is described by Quantum Electrodynamics (QED), a relativistic field theory that describes the interactions of charged particles via the electromagnetic field experienced through the exchange of photons.

The *Weak Interaction* affects particles with weak isospin and is mediated by the exchange of the massive charged $W^{\pm}$ boson or the neutral Z^0 boson. Weak isospin is analogous to electric charge for electromagnetism, yet is a property of all quarks and leptons, making the weak interaction the only fundamental force to couple to all fermions. Due to the large mass of the weak gauge bosons, (80.385 ± 0.015 GeV/c^2 for the $W^{\pm}$ boson, 91.1876 ± 0.0021 GeV/c^2 for the Z^0 boson [3]), they are short-lived. This causes the weak interaction to have a typical field strength several orders of magnitude less than electromagnetism or the strong nuclear force, and thus acts at a shorter range; hence the name.

The helicity of a particle is the projection of it's spin onto the direction of momentum. A particle has right-handed helicity if the direction of it's spin is the same as the direction of it's motion, and is left-handed if these directions oppose. The weak interaction distinguishes itself by being the only interaction to violate parity; the $W^{\pm}$ boson only couples to left-handed particles and right-handed antiparticles. The weak interaction theory is sometimes referred to as Quantum Flavourdynamics, as it is the only interaction capable of changing quark flavour within the SM.

At higher energies, of the order of the Z^0 boson mass, the electromagnetic and weak interactions are described by the combined *Electroweak Interaction*. Electromagnetism and the weak interaction, which manifest at low energies, are, in fact, different aspects of one fundamental electroweak force. The disparity in strength between these two observed forces is accounted for by the massive bosons that mediate the weak interaction. However, the electroweak

interaction specifies that all fundamental particles are massless, which is not the case for the $W^{\pm}$ and Z^0 bosons. The theory is that at the high temperatures of the early universe there existed electroweak symmetry in that the mediators of both the forces were massless. However, as the universe cooled this symmetry is broken via the Higgs mechanism, which causes the $W^{\pm}$ and Z^0 bosons to acquire mass and the photon to remain massless [4].

The *Strong Nuclear Force* is described by Quantum Chromodynamics (QCD) and is mediated by the gluon, of which there are 8 independent types. Analogous to the electromagnetic force, it acts between particles which have colour charge, a property of only quarks and gluons; however, whilst there is only one kind of electric charge (be it positive or negative), there are three kinds of colour charge, red, green, and blue. Since the mediating gluons themselves carry colour, unlike the electrically neutral photon, they can couple directly to other gluons. So, in addition to the fundamental quark-gluon vertex, this leads to gluon-gluon vertices, involving three or four gluons. This gluon self-coupling makes QCD more complex than QED, but introduces the possibility of novel particle states; such as bound states of interacting gluons, without the presence of quarks, know as '*glueballs*' [5].

The strength of the QCD coupling constant, α_s, varies with the separation of the particles and, unlike with the weak interaction and QED, the strong force decreases in strength with decreasing distance, and vice versa. This behaviour is known as *asymptotic freedom* and is a result of the unique self-coupling nature of the mediating gluons. In the case of QED, a cloud of virtual electron-positron pairs can be thought to exist between two charges, effectively shielding the 'bare' charges so that, from a distance, the effective charge appears reduced. The reduction occurs due to opposite charges attracting, and like charges repelling. As the two charges come closer they begin to penetrate each others virtual charge clouds, exposing themselves to each others bare charges, so that the effective charge increases. A similar process occurs for QCD, except that the virtual cloud is composed of not only quark-antiquark pairs but also gluons, giving rise to an 'anti-shielding' effect, where the effective colour charge at larger distances becomes greater

than the bare charge. The virtual quark-antiquark components drive α_s up at short distances, equivalent to the electron-positron pairs in QED, whilst the virtual gluon pairs, that dominate, drive α_s down. This has the effect that at short distances the force between quarks is weak and they behave almost like free particles.

As the distance between two quarks increases, the intermediate gluon fields form narrow tubes of colour charge, which act on the quarks to hold them together. After a certain distance this force becomes constant, after which it becomes more energetically favourable for a new quark-antiquark pair to spontaneously appear, from the energy applied to pull the quarks apart, rather than to allow this tube to extend further. This phenomenon is known as *colour confinement*; it implies that no individual free quarks can be observed as, instead, the new quarks pair up with the originals to form colour-neutral hadrons on very short time scales. This idea is reinforced by there being no experimental observations of free quarks to date.

The bottom (b) quark, whose existence was originally proposed by Kobayashi and Maskawa in 1973 [6] and was subsequently discovered in 1977 by the Fermilab E288 experiment [7], is the lightest third generation quark at $4.19^{+0.18}_{-0.06}$ GeV/c^2 (in the $\bar{\text{M}}$S scheme) [3]. It can decay into either an up or a charm quark via the weak interaction. However, these decays are suppressed by the Cabibbo-Kobayashi-Maskawa (CKM) matrix (equation 2.2) [3]. The CKM matrix describes the couplings between quark flavours in charged interactions with W$^{\pm}$ bosons. The 'weak interaction generations':

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix} \tag{2.1}$$

are related to the physical quark states by the CKM matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.2}$$

On the left of equation 2.2 is the weak interaction doublet partners of up-type quarks (up, charm and top), and on the right is the (3×3) CKM matrix along with a vector of mass eigenstates of down-type quarks (down, strange and bottom). The relative proportion of decays that proceed by the process $t \to WX$ are given by the square of the amplitude of the relevant CKM matrix element, $|V_{tX}|$. As the CKM matrix is unitary, this implies that the majority of top quark decays produce a $W^{\pm}$ boson and a b-quark. The value of $|V_{tb}|$ has been directly measured experimentally by the D0 [8] and CDF collaborations [9, 10] from s- and t-channel single top production, and recently by CMS [11] from t-channel production. The average of these measurements [3] gives a value of:

$$|V_{tb}| = 0.89 \pm 0.07 \tag{2.3}$$

The best estimate of the current parameters of the CKM matrix is provided by a global fit to the individual measurements of the parameters [3]. An overview of the numerous measurements and experiments involved can be found elsewhere [3]. The global fit yields the results shown in equation 2.4.

$$V_{CKM} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.00347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{pmatrix} \tag{2.4}$$

An important aspect of top quark physics is the short lifetime of the

particle compared to the other known quarks; a result of its large mass. The lighter quarks can all hadronise with other flavour quarks to form a variety of baryons and mesons, which themselves each have various decay topologies. The top quark, however, has a lifetime so short that it decays before hadronisation can occur and so can only decay via electroweak processes. Therefore, whilst there are several channels of top quark production, there is effectively only one decay process. In the SM, the top quark is expected to decay via the weak process to a $W^{\pm}$ boson and a b-quark approximately 100% of the time.

Top quark events are characterised according to the decay of the $W^{\pm}$ bosons, which can occur either leptonically or hadronically. The leptonic decays are of the form $W^- \rightarrow l^- + \bar{\nu_l}$ and $W^+ \rightarrow l^+ + \nu_l$, where l denotes a charged lepton, be it an electron, muon or tau, which collectively have a branching fraction of $BR(W \rightarrow l\nu) = (10.80 \pm 0.09)\%$ [3]. An hadronic decay occurs when the $W^{\pm}$ boson decays to a quark-antiquark pair of differing flavour, with a branching fraction of $BR(W \rightarrow hadrons) = (67.60 \pm 0.27)\%$ [3].

This analysis deals exclusively with leptonic decaying $W^{\pm}$ bosons, but ignoring those involving taus. The decay of a top quark to a leptonically (and hadronically) decaying W^+ boson is illustrated in Fig. 2.1. The hadronic decays of the tau lepton, due to its high mass, are problematic to reconstruct, so are omitted from many leptonic analyses. The final, dileptonic topology being studied is therefore categorised by the presence of two oppositely charged leptons (any combination of electrons and muons), a b-quark and two neutrinos.

2.2 Top quark topology

At 173.5 $\pm$ 0.6(stat.) $\pm$ 0.8(syst.) GeV/c^2 [3] the top quark is the most massive of all observed elementary particles, approximately 40 times heavier than the next largest quark, the b-quark. Thus, large amounts of energy are needed to create a top quark, which is experimentally done in particle

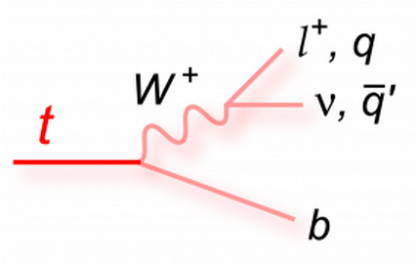


Figure 2.1: Feynman diagram of a top quark decaying to a W^+ boson and a b-quark, with the subsequent leptonic or hadronic decay of the W^+ boson [12].

colliders, such as the Tevatron and Large Hadron Collider (LHC).

At both colliders the primary source of top quark production is via the strong interaction producing a top-antitop quark pair. This process was the discovery mode for the Tevatron [13], and represented the first top quark measurement for the general purpose LHC experiments ATLAS [14] and CMS [15]. In $\sqrt{s} = 7$ TeV proton-proton collisions at the LHC, $t\bar{t}$ production proceeds predominantly via gluon-fusion (∼85%), with the remaining $t\bar{t}$ pairs coming from quark-antiquark annihilation [16], shown at LO in Fig. 2.2. At the LHC design energy of $\sqrt{s} = 14$ TeV, the gluon-gluon fusion rate will reach as high as 90% [16].

Conversely, at the Tevatron, quark-antiquark annihilation dominates $t\bar{t}$ production (∼90%) [16]. This rate difference in production mechanism is, in part, due to the change in Parton Distribution Function (PDF) caused by the increase in collision energy between the two experiments; another factor is the choice of collision particles. The Tevatron collides protons with anti-protons so that the annihilating quarks are both valence quarks/antiquarks, whilst at the LHC a valence quark from one proton must annihilate with a sea antiquark from the other proton; or similarly, the annihilation of a sea quark and a sea antiquark.

Top quarks are produced singly via the weak interaction, which occurs in

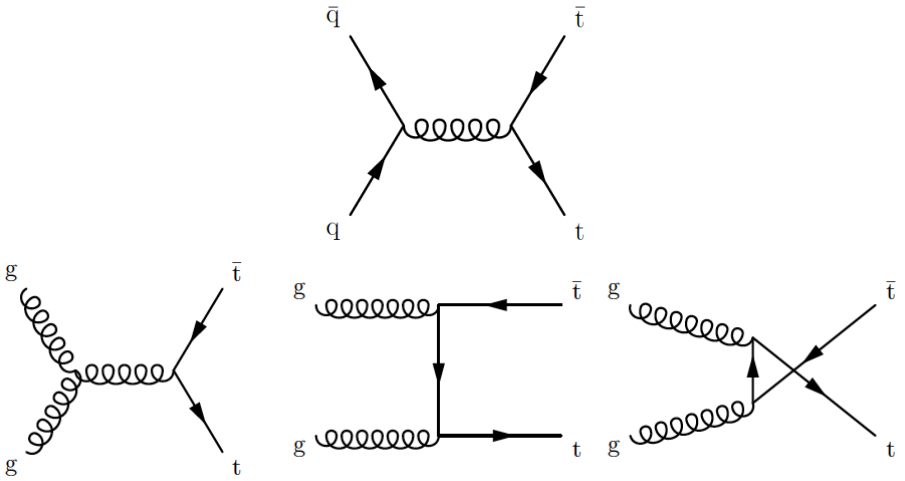


Figure 2.2: Leading order Feynman diagrams for $t\bar{t}$ production at the LHC [16].

Table 2.1: Standard Model theory for single top quark production cross section per channel and $t\bar{t}$ at the Tevatron [17] and LHC, shown at next-to-next-to-leading-order [18, 19, 20, 21].

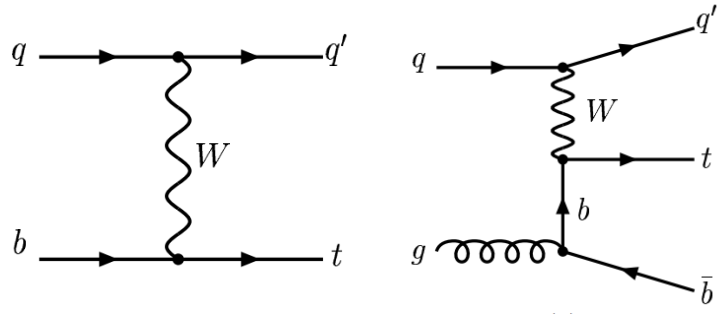
	$\sqrt{s}$ [TeV]	s-channel [pb]	t-channel [pb]	tW-channel [pb]	$t\bar{t}$ [pb]
Tevatron	1.96	1.046	2.08	0.266	7.31
LHC	7	4.59	64.2	15.6	163

three different ways, referred to as channels. As each process has different initial and final states, they are deemed separate processes and thus are, in principle, separately measurable. The cross sections for each single top process, and for $t\bar{t}$ production, are presented in Table 2.1 for both the LHC and Tevatron, assuming a top quark mass (m_t) of 173 GeV/c^2 for the LHC rates, and 172 GeV/c^2 for the Tevatron estimates. At the LHC, the production rate for $t\bar{t}$ production is estimated to be approximately twice that of the combined single top modes.

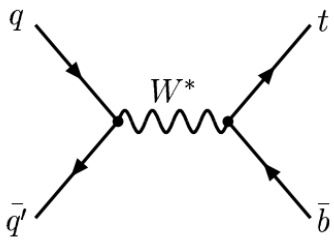
The t-channel process proceeds via the exchange of a virtual, space-like $W^{\pm}$ boson, see Fig. 2.3(a). The t-channel partonic processes are of the form

$qb \to q't$ and $\bar{q}b \to \bar{q}'t$. This channel is the dominant single top quark process at both the Tevatron and LHC experiments. At the LHC, the t-channel production cross section is approximately a factor 32 larger than at the Tevatron, at $\sqrt{s} = 7$ TeV, due to the increased possible momentum fraction carried by gluons produced in the larger centre-of-mass collisions.

The s-channel process, also known as quark-antiquark annihilation, proceeds via the exchange of an off-shell (time-like) $W^{\pm}$ boson which decays to a top and antibottom quark; see Fig. 2.3(b). Its lowest order processes are of the form $q\bar{q}' \to \bar{b}t$, of which the dominant process is $u\bar{d} \to \bar{b}t$ as well as processes involving the charm quark and CKM-supressed contributions [18]. It has a relatively large rate at the Tevatron, but is comparatively small at the LHC because it is driven by initial state antiquark parton densities.



(a) t-channel



(b) s-channel

Figure 2.3: Leading order Feynman diagram for single top quark production in the t-channel [21] and s-channel [18].

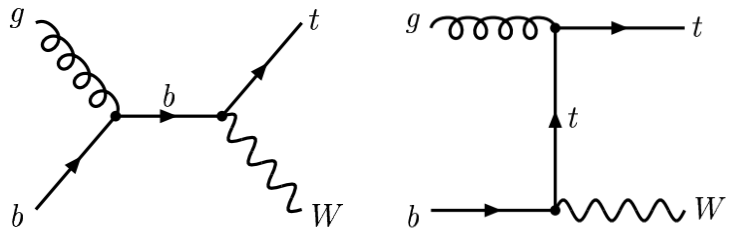


Figure 2.4: Leading order Feynman diagrams for single top quark production in the tW-channel [22].

This analysis is concerned with the third method of single top production, where an initial state b-quark emits a real, on-shell $W^{\pm}$ boson resulting in a final state of a top quark with an associated $W^{\pm}$ boson, $bg \to tW^-$ and $\bar{b}g \to \bar{t}W^+$ (shown in Fig. 2.4). This is referred to as the tW-channel. What differentiates this process from the other production channels is that both the top quark and the $W^{\pm}$ boson are observable as final state particles, and as such the production channel is sensitive only to physics (new or otherwise) which directly affects the Wtb interaction. In addition to the main decays listed, all three single top production channels also have CKM-suppressed contributions and which are therefore much rarer.

2.3 Interference with $t\bar{t}$ production

At leading order (LO), the tW-channel single top and $t\bar{t}$ processes are well-defined, with $\sigma_{tW} < \sigma_{t\bar{t}}$ (as shown in Table 2.1). When progressing to next-to-leading-order (NLO) in QCD, the tW-channel includes higher-order corrections of the type shown in Fig. 2.5. However, these additional diagrams can also be thought of as the production of a top quark pair at LO, with a $\bar{t}$-quark decaying to produce the W^- and $\bar{b}$ quark (or similarly for $t \to W^+b$).

Fig. 2.5 shows a subset of the diagrams that contribute to $gg \to tW\bar{b}$ or $q\bar{q} \to tW\bar{b}$, which represent the interference between the tW-channel single

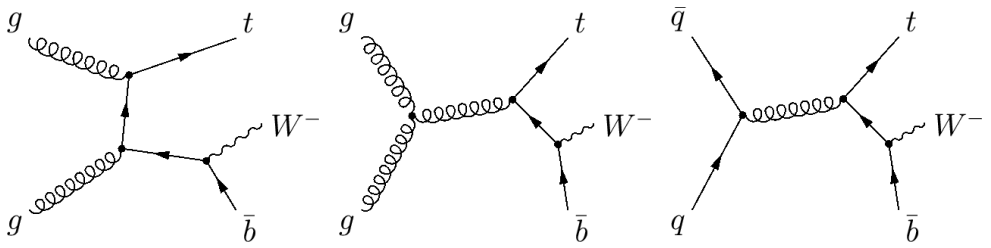


Figure 2.5: Feynman diagrams for NLO tW-channel single top quark production that are removed from the signal definition in the Diagram Removal simulation scheme [22].

top production and $t\bar{t}$ production. They are denoted as *doubly resonant*, in that the intermediate antitop quark can be on- or off-shell. So-called *singularly resonant* diagrams are those which are identified with the tW-channel only. The interference becomes a problem if the invariant mass of the final state $W^{\pm}b$ system is close to m_t, as the propagator for the intermediate top particle becomes large and the cross section for these problematic diagrams increases. This is because in this mass region the diagrams can then represent the production of a $t\bar{t}$ pair, with the subsequent decay into the W^- and $\bar{b}$ quark, which exhibits a higher production cross section than the equivalent tW single top processes [23]. Subsequently, the interference is suppressed outside this mass region.

In practise, although very similar at LO, the two processes do have distinct features that can be used to separate them; principally a second b-quark in $t\bar{t}$ events, from the decay of the second top quark, leads to a final state with more jets and more jets originating from b-quarks (see Section 5.4 for more details). Although additional b-quarks can also be seen in NLO tW events, as in Fig. 2.5, they occur via gluon-splitting, so are typically softer and thus distinguishable from the harder additional b-quarks in $t\bar{t}$ events.

The problem of interference actually lies within any computation that includes contributions beyond the LO, i.e. the simulation of the two processes. In previous tW-channel analyses (including that of the ATLAS experiment

[24]), LO simulation descriptions were used for the tW process, with a large associated modelling uncertainty, avoiding $t\bar{t}$ interference in the signal simulation sample. The issue is defining the tW-channel beyond LO in a way that is applicable in an event generator context, where both initial- and final-state *parton showers* (jets and hadrons produced in the decay of the underlying hard process particles) are present. Ignoring the interference in the NLO simulation production and then removing the problematic events in the sample by applying restrictions on the kinematic features is impractical. That approach would require long production times to generate simulation samples, that, when the large amount of 'useless' interference events are removed, are of sufficient size and purity for use in analyses. For efficient simulation it is desired to have a large 'pure' sample, hence the requirement of specialised tW definitions.

There are two working definitions of tW-channel production for event generation purposes, designed such that, by comparing them, one can directly assess the impact of the interference with $t\bar{t}$ production [23]. Initially, these approaches were part of the MC@NLO event generator formalism [23], and have since been incorporated into the POWHEG generator [25], which produces the single top simulations used in this analysis (see Section 4.1). The definitions are:

- The Diagram Subtraction (DS) scheme modifies the NLO tW cross section by introducing a subtraction term that locally cancels the $t\bar{t}$ contribution.
- In the Diagram Removal (DR) scheme all doubly resonant NLO tW diagrams are removed (i.e. Fig. 2.5), dealing with the interference at the amplitude level. Although effective, it should be noted that this method breaks gauge invariance [26].

For the DS scheme, in the computation of the NLO cross section - for the process $\alpha\beta \to tW\bar{b}$, where $\alpha\beta = gg$ or $q\bar{q}$ - the contribution from the doubly resonant diagrams, $\mathcal{D}_{\alpha\beta}$, is given by the magnitude squared of the

amplitude for $t\bar{t}$ production, $|\mathcal{A}_{\alpha\beta}^{(t\bar{t})}|^2$ [23]. The DS scheme locally defines a gauge-invariant counteracting term, $\tilde{\mathcal{D}}_{\alpha\beta}$, such that it can be subtracted and therefore totally cancel the doubly resonant contributions when $M_{\bar{b}W} = m_t$. This subtraction term is thus constructed as:

$$\tilde{\mathcal{D}}_{\alpha\beta} = \frac{BW(M_{\bar{b}W})}{BW(m_t)} |\hat{\mathcal{A}}_{\alpha\beta}^{(t\bar{t})}|^2 \tag{2.5}$$

where $BW()$ denotes a Breit-Wigner (or Lorentz) function, widely used for describing the non-interfering cross-section of particle resonant states [27]; and the momenta of $|\mathcal{A}_{\alpha\beta}^{(t\bar{t})}|^2$ have been 'reshuffled' so that the kinematics of $\tilde{\mathcal{D}}_{\alpha\beta}$ represent the full NLO $\alpha\beta \to tW\bar{b}$ process whilst simultaneously having the $\bar{t}$ being on-shell.

While the DS scheme has the advantage that all diagrams are to a certain extent considered, the subtraction term can lead to negative weights for some simulated tW-channel events. For simplicity, as DR does not lead to negative weights, it is chosen as the default scheme for top physics analyses, including that presented in this thesis. The effect of using the DS sample is taken as a systematic uncertainty; see Section 6.2.5 for a full list of systematic uncertainties. Comparing the event yields using the two schemes, by construction, provides an estimate of the size of the interference with $t\bar{t}$.

2.4 Single top quark at the LHC

The electroweak production of single top quarks was first reported by the D0 and CDF experiments at the Tevatron Collider, at the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, USA [28, 29, 30], where they were able to measure both s- and t-channel cross sections. Associated W production was not measured as the production cross section at the Tevatron is too small to be measured. This is because the lower energy Tevatron collisions produce fewer gluons carrying large momentum fractions compared

to the LHC, so the quark-induced s- and t-channel production modes are favoured. The increase of the centre-of-mass energy from the Tevatron to the LHC has non-uniformly increased the rates for all three production channels, shifting the PDF such that the gluon or b-quark induced processes see a greater cross section increase than the valence quark induced processes. As shown in Table 2.1, the tW-channel production now exceeds that in the s-channel, so as to become a viable study. Meanwhile, the s-channel has become the challenging measurement due to its small cross section. In contrast to the Tevatron, however, s-channel measurements will be possible at the LHC as the amount of data accumulates. An s-channel measurement has already been attempted by the ATLAS experiment, but it exhibits 60-100% systematic uncertainties due to data and simulation statistics [31].

The ATLAS and CMS experiments have both measured the t-channel production cross section, the details of which can be found elsewhere [32, 11]. The ATLAS collaboration was also the first experiment to present evidence for tW associated production [24]. A cross section measurement of $16.8 \pm 2.9(\text{stat.}) \pm 4.9(\text{syst.})$ pb, with an observed (expected) significance of 3.3σ (3.4σ) is achieved using 2.05 fb^{-1} of data from $\sqrt{s} = 7$ TeV pp collisions. The CMS tW associated production analysis, presented here, was published in January 2013 [22].

Outside of the importance of the actual measurement in and of itself, the production of a single top quark with an associated on-shell $\mathrm{W}^{\pm}$ boson has several useful applications. It provides information about the Wtb interaction vertex which, by probing a different kinematic region, is complimentary to that provided by s- and t-channel measurements. Like the other single top quark processes it allows for a direct measurement of the $|V_{tb}|$ CKM matrix element, so is a sensitive probe for a potential fourth quark family. It is also an important background for future $\mathrm{t\bar{t}}$ production and $H \to WW$ measurements.

In 2012 CDF used their latest precision measurements of the $\mathrm{W}^{\pm}$ boson

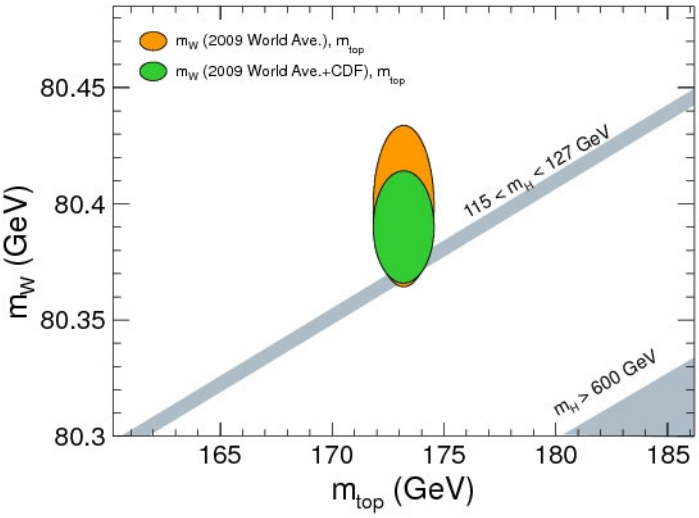


Figure 2.6: Higgs mass constraint plot using leading precision mass estimates for the top quark and W boson [33].

and top quark mass (calculated from single top and $t\bar{t}$ production) to restrict the, then unobserved, Higgs boson mass to be to less than 145 GeV/c^2 [33], as shown in Fig. 2.6. The grey bands show the remaining allowed regions for the Higgs boson mass, after exclusions obtained at LEP, the Tevatron and the LHC. The achieved result is compatible with the Higgs-like-particle results published by the LHC experiments in 2012 [34].

Chapter 3

The Compact Muon Solenoid detector

The Compact Muon Solenoid (CMS) detector is located at the European Centre for Nuclear Research (CERN) on the Large Hadron Collider (LHC). The LHC was designed to deliver proton-proton collisions at a centre-of-mass energy of up to $\sqrt{s} = 14$ TeV, with a peak luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} [35]. This luminosity corresponds to 2808 bunches of protons per beam with a maximum number of 1.15×10^{11} protons per bunch. The spacing between the proton bunches at design luminosity is 25 ns, resulting in a collision rate of 40 MHz. For a safe operation and to avoid magnet quenches (like the one that occurred in September 2008 [36]), the LHC operated with $\sqrt{s} = 7$ TeV until the end of 2011; the delivered and recorded integrated luminosity during this data taking period is shown in Fig. 3.1. In 2012, the energy was raised to 8 TeV, with an extended maintenance shutdown planned for 2013 before the energy is then raised up to the full design limit of 14 TeV. The analysis reported in this thesis was conducted using data recorded in 2011, where the maximum number of bunches was 1380 at a bunch spacing of 50 ns. This corresponds to a peak luminosity of 3.65×10^{33} cm^{-2}s^{-1} [37].

CMS is a general-purpose detector, distinguished by its large, high field solenoid magnet, a silicon inner tracking system and a homogenous scintillat-

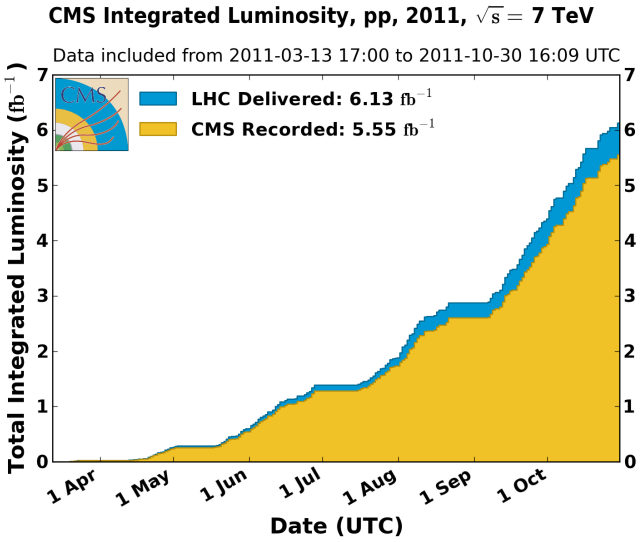


Figure 3.1: The delivered, and recorded integrated luminosity at the CMS experiment for the 2011 data taking period. The delivered luminosity corresponds to that delivered by the LHC. The recorded luminosity to when CMS was recording data, corresponding to a recording efficiency of 90.5% [38].

ing crystal electromagnetic calorimeter system. The design of CMS is motivated by the physics programme intended for the LHC, primarily the goal of understanding the mechanism for electroweak symmetry breaking, and the practical implications of dealing with the high luminosities of the collider. In particular, the electronics and triggering system are designed to be fast enough to take advantage of the high collision frequency. The sub-detectors near to the LHC beam pipe, in particular the forward hadronic calorimeter (HCAL) and inner edge of the electromagnetic calorimeter (ECAL) endcaps, which receive the highest fluence, also have to be sufficiently radiation-hard so as to handle the hostile environment there, inherent in high luminosity running.

The CMS detector reference frame is centred at the Interaction Point of CMS (IP5) of the LHC, employing a right-handed Cartesian coordinate system. The origin is defined as the nominal interaction point, with the x-axis pointing into the centre of the LHC ring, the y-axis pointing upwards

perpendicular to the ground and the z-axis pointing in the anti-clockwise direction, along proton beam 2 of the LHC. The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, ϕ, is defined as the angle in the $x - y$ plane. It is often useful to express sections or locations in the θ plane using pseudorapidity (η), which is defined in terms of θ in equation 3.1.

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

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

The instantaneous luminosity L describes the rate dN/dt of a certain process per its cross section σ:

$$L = \frac{dN}{dt} \cdot \sigma^{-1} \tag{3.2}$$

It needs to be distinguished from the integrated luminosity, $\mathcal{L}$, which denotes the total amount of recorded events of a certain process per cross section:

$$\mathcal{L} = \int L dt = N \cdot \sigma^{-1} \tag{3.3}$$

Whilst the cross section is specific to each individual process, the luminosity is a parameter of the LHC. The protons in the LHC are organised into n_b bunches, each containing N_b particles per bunch. Given that both proton beams have the same energy, the luminosity can be expressed as the following function of the accelerator parameters:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta^*} F \tag{3.4}$$

In this equation, $f_{rev} = c/r_{LHC} = 11.25$ kHz is the revolution frequency, $\gamma_r = 1/\sqrt{1 - v^2/c^2}$ the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, β^* the focus of the beam and F the geometric luminosity reduction factor due to the crossing angle at the interaction point of the beams.

3.1 Overview

The CMS detector is 21.6 metres long and 14.6 metres in diameter with a weight of 14,000 tons; the overall layout of CMS and its sub-detectors is illustrated in Figure 3.2. The detector has a superconducting solenoid magnet 12.5 m long and with an inner diameter of 6 m, designed to generate an axial magnetic field of 4 T (in actuality, 3.8 T), twice that of the field used in the ATLAS detectors inner solenoid [39]. The bore of the magnet coil is large enough to accommodate the tracking system, the ECAL system and the HCAL system. Located closest to the beam-line is the inner tracking system consisting of layers of pixel detectors and highly granular silicon strip detectors. Together they provide precise measurements of interaction vertices and charged particle tracks. Surrounding the tracking system is the homogeneous ECAL, which provides an excellent energy resolution for reconstructed photons and electrons. It is made of lead tungstate ($PbWO_4$) scintillating crystals, with the scintillation light being detected by silicon avalanche photo-diodes in the barrel region and vacuum photo-triodes in the endcap regions. The HCAL system fills the remaining space between the ECAL and superconducting magnet, consisting of alternating layers of brass and active plastic scintillating material to collect energy from hadronic jets. The magnetic field of CMS is closed by an iron return yoke surrounding the superconducting magnet, which supports the muon detectors and additional forward hadronic calorimeters on the outer layers of the detector.

This chapter will focus on the sub-detectors essential to the performance of this analysis; a more complete description of the detector can be found elsewhere [40].

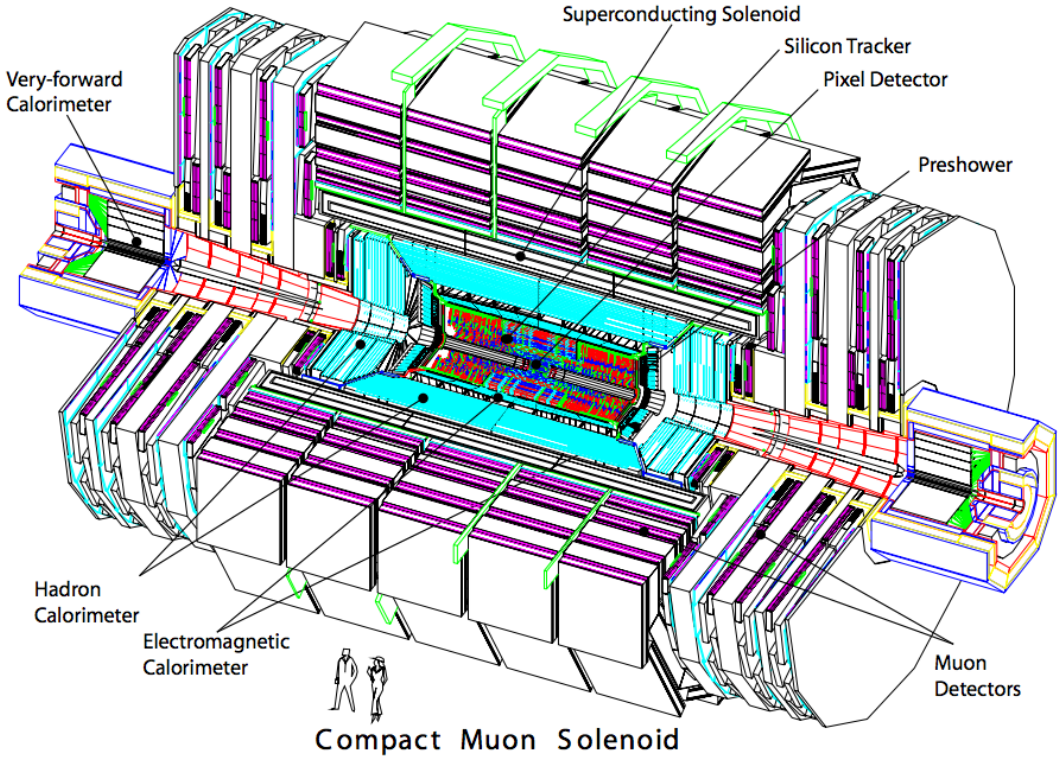


Figure 3.2: Cut away diagram of the CMS detector illustrating the location and relative size of the sub-detectors [40].

3.2 Inner tracking system

3.2.1 Overview

The CMS inner tracking system is designed to resolve the trajectories of charged particles as they traverse the detector and to provide high resolution vertex reconstruction. By measuring the paths of these particles in the magnetic field that encompasses the entire inner tracker, the momentum and charge of the particles can be measured. As such it is an important component for the reconstruction of these charged particles. Specifically for this analysis, the tracker is used for electron and muon identification and the reconstruction of secondary decay vertices observed in heavy flavour decays. The tracker system consists of pixel layers at the centre surrounded by layers of silicon strip detectors, which are illustrated in the schematic diagram of Fig. 3.3. The cluster information from the pixel layer is, as of 2011, used to measure the offline instantaneous luminosity on CMS, and to provide corrections for the online measurement made using the Forward Hadronic Calorimeter [41]; refer to Section 6.1 for the explanation of this measurement. In total, the tracking system has a length of 5.8 m and diameter of 2.5 m, with the geometrical acceptance extending up to pseudorapidities of $|\eta| < 2.5$.

Closest to the beam-line lies the pixel detector, consisting of three layers at radii of 4.4, 7.3 and 10.2 cm, arranged in cylindrical layers. Two disks of silicon pixels are placed at each end, at $\pm$ 34.5 and $\pm$ 46.5 cm, to complete the forward coverage. Each individual pixel covers an area of 100×150 μm^2 in the $r - \phi$ and z coordinates, respectively. In total, the pixel detector contains 66 million pixels, corresponding to an active area of 1 m^2.

Around the pixel detector is the silicon strip tracker, covering the radial region from 20 to 116 cm. In the barrel, the silicon strip tracker is composed of three sub-sections; the Tracker Inner Barrel (TIB) and Disks (TID) and the Tracker Outer Barrel (TOB). The inner section, comprised of the TIB and TID, covers the region up to 55 cm in radius and is composed of 4 strip

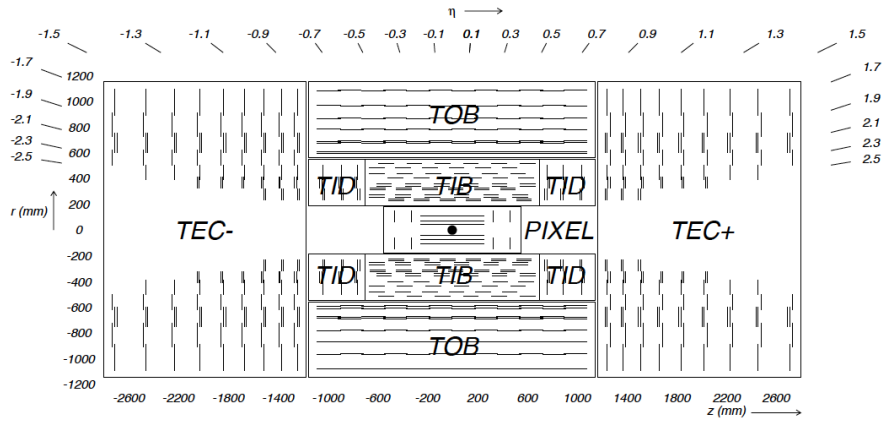


Figure 3.3: Schematic cross section through the CMS tracker. Each line represents a detector module. Double line indicate back-to-back modules which deliver stereo hits [40].

layers in the barrel and 3 disk layers on each end. Additionally, the inner two layers of the TIB and the first two disks of the TID have extra silicon strips mounted on their backs at small stereo angles (denoted by double lines in Fig. 3.3). This enables measurements of the z-coordinate in the barrel and r in the disks. The typical cell size of the silicon strips in this inner region of the tracker is 10 cm $\times$ 80 μm, with a thickness of 320 μm.

The TOB surrounds the TIB and TID, forming the outer-most radial layer of the tracker system, covering $\pm$118 cm in z and extending the radial coverage up to 116 cm. The TOB consists of six layers of silicon micro-strips which, as the TOB is located further from the interaction point, are constructed with larger sensors of dimensions up to 25 cm $\times$ 180 μm, with an increased thickness of 500 μm. The two inner-most strip layers of the TOB are 'double-sided', with additional rear mounted strips like the TIB/TID. Finally, the inner tracker is completed by the Tracker EndCaps, denoted as TEC+ and TEC-, respectively, depending on their location in the z-direction, which each contain nine disks of silicon sensors, all of which are double-sided. In total the CMS silicon strip tracker comprises 9.3 million strips with an

active area of 198 m^2.

3.2.2 Performance

The primary vertex resolution strongly depends on the number of tracks used in fitting the vertex and on their transverse momentum, p_T. To measure the resolution, the tracks in an event with only one vertex are randomly split into two different sets and used to independently fit the primary vertex [42]. The distribution of the difference in the fitted vertex positions can then be used to extract the resolution by fitting a Gaussian to it and dividing σ by $\sqrt{2}$. To examine the effect of the p_T of the tracks in the vertex, the resolution versus the number of tracks in the vertex is studied for different average track p_T in the vertex. Fig. 3.4 shows the x, y, and z resolutions for different average p_T ranges. While the resolution differs considerably depending on p_T and multiplicity, the simulation accurately reproduces the data results.

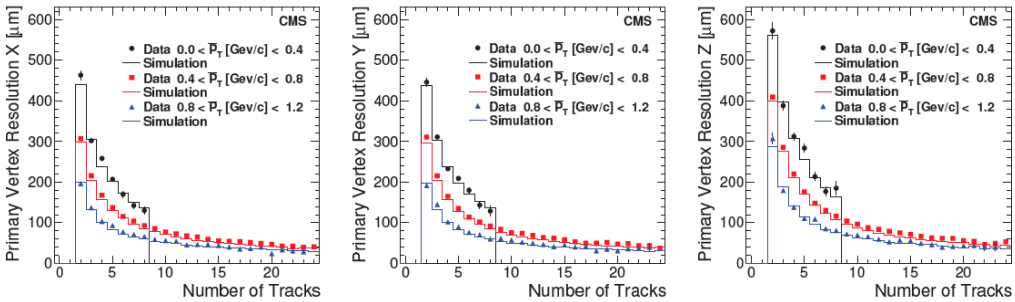


Figure 3.4: Primary vertex resolution distributions in x (left), y (middle), and z (right) versus number of tracks. The three sets of results in each plot show different average p_T ranges and within each p_T range, data and simulation are compared [42].

Fig. 3.5 shows the dependence on pseudo-rapidity of the resolution on the p_T and transverse impact parameter for samples of isolated muons with p_T of 1, 10 and 100 GeV/c [43]. The resolutions on both track parameters generally worsens for larger values of $|\eta|$ because the extrapolation length from the

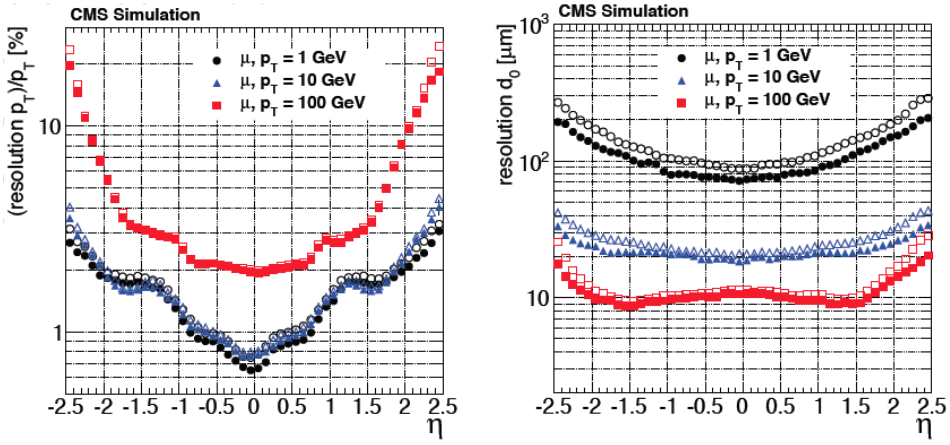


Figure 3.5: Resolution, as a function of pseudorapidity, for single, isolated muons with p_T of 1, 10, and 100 GeV/c: transverse momentum (left), transverse impact parameter (right). For each bin in η, the solid symbol corresponds to the width of a gaussian fitted to the residuals distribution and the open symbol represents the RMS of the same distribution [43].

innermost hit to the beam axis, where the parameters are calculated, becomes larger. At high momentum (100 GeV/c), the resolution of the transverse momentum is around 1-2% up to a pseudorapidity of $|\eta| = 1.6$, but gets worse at higher pseudorapidity values. The degradation around $|\eta| = 1.0$ and beyond is due to the gap between the barrel and the end-cap disks. At a transverse momentum of 100 GeV/c, the material in the tracker accounts for between 20 and 30% of the transverse momentum resolution; at lower momenta, the resolution is dominated by multiple scattering and its value reflects the amount of material traversed by the track. The relative precision with which the p_T is measured is best for tracks with p_T of approximately 3 GeV/c.

At high momentum, the impact parameter resolution is dominated by the resolution of the innermost hit in the pixel detector, whilst at lower momenta, the resolution is progressively degraded by multiple scattering, until the latter becomes dominant.

3.3 Electromagnetic calorimeter

3.3.1 Overview

After passing through the tracking system particles enter the electromagnetic calorimeter where the energy of electrons and photons is measured [40]. The ECAL consists of a barrel section (EB) covering a range up to $|\eta| < 1.479$ and two endcaps (EE) in the range $1.479 < |\eta| < 3.0$, as shown in Fig. 3.6. In addition, there is a preshower detector in front of the endcaps ranging from 1.653 to 2.6 in $|\eta|$. The EB and EE regions of the ECAL are made of scintillating lead tungstate crystals ($PbWO_4$) [44]. With a high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radius (2.2 cm) they produce a compact calorimeter with a fine granularity. To measure the intensity of the scintillating light produced by the crystals two different choices of photodetector are used, based on the configuration of the magnetic field and the expected levels of radiation; avalanche photo-diodes (APDs) are used in the barrel region and vacuum photo-triodes (VPTs) in the endcaps [40].

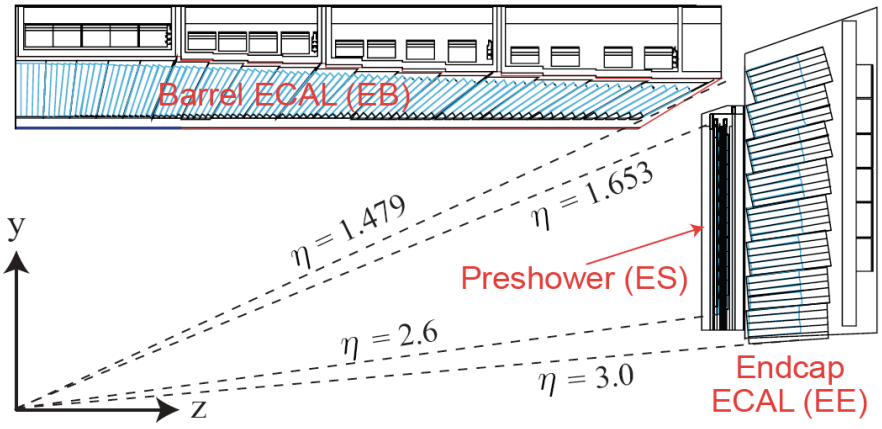


Figure 3.6: A schematic view of the $+y$ and $+z$ section of the ECAL. The barrel (top) and an endcap half with the preshower detector (right) are shown [45].

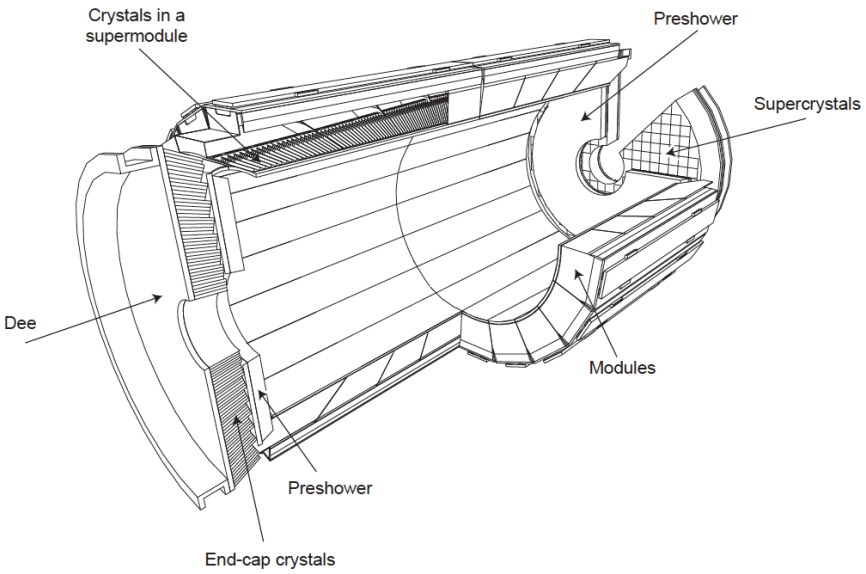


Figure 3.7: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front [40].

The ECAL barrel is comprised of 61,200 crystals arranged in a radial configuration covering the absolute pseudorapidity range up to 1.479. Clusters of between 400 to 500 crystals are grouped into a total of 36 supermodules, each covering a 20° region in ϕ and one half of the full length of the barrel, as illustrated in Fig. 3.7. Each crystal corresponds to a solid angle of approximately 0.0174 $\times$ 0.0174 in $\eta-\phi$, which is equivalent to a cross section of 22×22 mm^2 at the front face of the crystal and 26×26 mm^2 at the rear face. The length of each crystal is 230 mm, which corresponds to 25.8 radiation lengths.

The ECAL endcaps cover the fiducial region $1.479 < |\eta| < 3.0$, and are placed at $z = \pm3.154$ m from the interaction point. Each endcap is split into two halves, known as "Dees", each containing 3662 crystals arranged into 5×5 arrays (supercrystals), illustrated in Fig. 3.7. In total, the endcaps are comprised of 14,846 identical crystals, with dimensions of 28.62×28.62 mm^2 and 30×30 mm^2 for the front and rear faces respectively. The length of each endcap crystal is 220 mm, which corresponds to 24.7 radiation lengths.

Finally, the ECAL preshower detector is a two-layered sampling calorimeter located in front of the endcaps, covering the pseudorapidity range of $1.653 < |\eta| < 2.6$ [40]. The aim of this sub-detector is to identify the decays of neutral pions and to improve the position resolution of electrons and photons detected in the endcap regions. The preshower is constructed from layers of lead radiators, to initiate electromagnetic showers, backed by silicon strip sensors, to measure the deposited energy. There are two such layers of lead and strip sensors to provide a two coordinate measurement of incident particles. The lead plate in the first layer is 2 radiation lengths thick, whilst the second corresponds to only 1 radiation length. The silicon strip sensors each have a width of 2 mm, with a total preshower width of 20 cm.

3.3.2 Performance

A complete supermodule from the ECAL barrel was tested using a test beam of electrons with momenta between 20 and 250 GeV/c [46]. The supermodule was fully equipped with the front end electronics and cooling and temperature systems; however it was tested without an applied magnetic field and without any inert material in front of the crystals. Measurements were conducted for incident electrons with seven different energies, shown in Fig. 3.8, aligned on the centres of the crystals. The energy resolution of the ECAL can be parameterised as:

$$(\frac{\sigma}{E})^2 = (\frac{S}{\sqrt{E}})^2 + (\frac{N}{E})^2 + C^2 \tag{3.5}$$

where E is in GeV, S is the stochastic term, N the noise term, and C the constant term. The Stochastic term arises from the combination of event-to-event fluctuations in the lateral shower containment ($\sim$1.5%), a photostatistics contribution ($\sim$2.3%), and fluctuations in the energy deposited in the preshower absorber (where present) with respect to what is measured in the preshower silicon detector ($\sim$5%) [47]. The noise term is comprised of detector noise from digitisation, the preamplifier and from pile-up effects; ran-

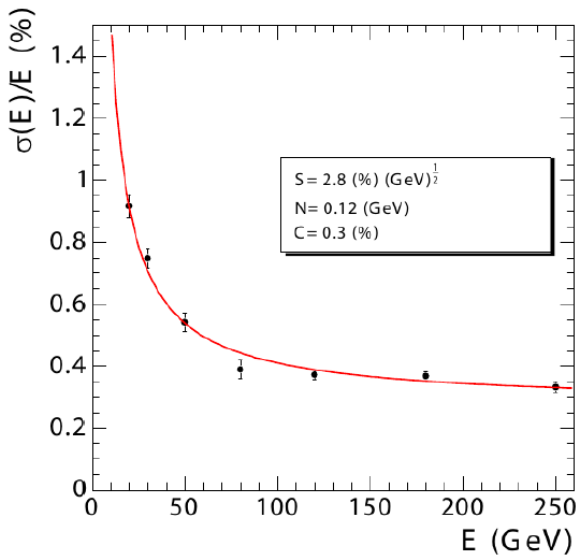


Figure 3.8: The energy resolution for electrons measured on a test beam of electrons with momenta between 20 and 250 GeV/c [40].

ging from 155 (770) MeV, at low luminosities, up to 210 (915) MeV at high luminosities, in the barrel (endcap) region [47]. The constant term has several contributions, the most important being crystal inter-calibration errors, energy leakage from the back of crystals and non-uniformity of the longitudinal light collection; the total constant term is taken as 0.55% [47]. These terms were measured in 2004, with electron beams of momenta between 20 and 250 GeV/c, to be 2.8%, 0.3% and 0.12 GeV for the stochastic, noise and constant terms respectively [46]; consistent with expectations.

Data from collisions in 2010, corresponding to 250 nb^{-1}, were used to measure the energy scale calibration of the ECAL sub-detectors [48]. The measured ratio of the reconstructed invariant mass peak between the data and simulation, using π^0 and η candidates that decay to $\gamma\gamma$, provides the ECAL scale correction. The EB energy scale was found to agree with the simulation to within 1%, whilst the EE scale agreed to within 3%.

3.4 Hadronic calorimeter

3.4.1 Overview

The hadronic calorimeter (HCAL) plays an important role in measuring the energies of hadrons, including protons, neutrons, pions and kaons. It is also used in determining the missing transverse energy in events containing neutrinos or exotic particles. The HCAL is a sampling calorimeter system comprised of four distinct sub-detectors; the HCAL Barrel (HB), HCAL Endcaps (HE), the Outer Hadronic calorimeter (HO) and the Forward Hadronic calorimeter (HF). The HF is used to measure the online instantaneous luminosity on CMS, shown in Fig. 3.1. The locations of all the four sub-detectors are illustrated in Fig. 3.9.

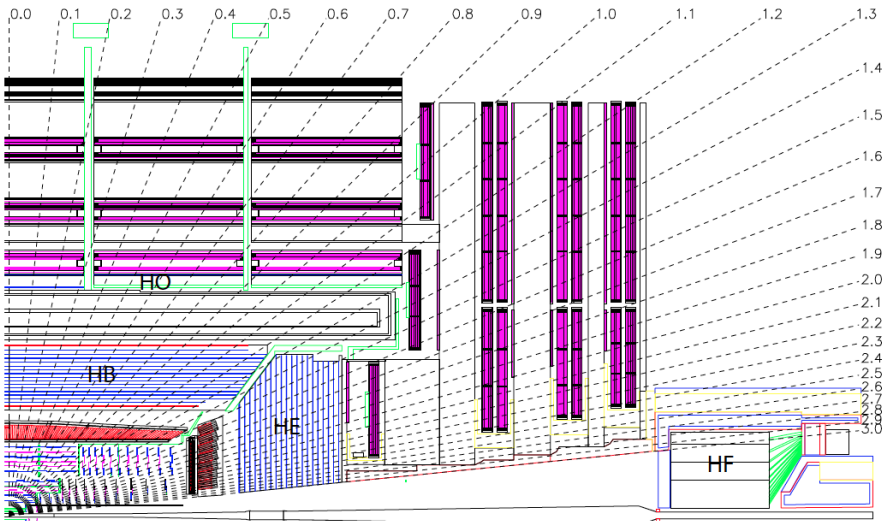


Figure 3.9: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [40].

The HB calorimeter is a sampling calorimeter that covers the pseudorapidity range $|\eta| < 1.3$. It is formed of 36 identical wedges, containing eight 50.5 mm-thick and six 56.55 mm-thick brass absorber plates, interspaced with

3.7 mm thick plastic scintillator tiles. Additionally, two stainless steel absorbers are used as the innermost and outermost plates (40 mm and 75 mm respectively) for structural stability. The plates are bolted together in a staggered geometry so as to give full radial coverage for the entire wedge. The plastic scintillator is divided into 16 η sectors, resulting in a segmentation of 0.087×0.087 in η and ϕ. The light from each scintillator tile is extracted by wavelength shifting fibres and is then measured using hybrid photo-diodes (HPDs). At $\eta = 0$ the effective thickness of the HB is 5.82 interaction lengths, increasing at higher pseudorapidities to a maximum of 10.6 interaction lengths. In terms of hadronic shower development, the ECAL adds an additional 1.1 interaction lengths in front of the HB.

The two HE calorimeters, covering the range $1.3 < |\eta| < 3.0$, close off the HB calorimeter at each end. This substantial portion of the rapidity range is expected to contain 34% of the final state particles produced. The HE calorimeters are constructed using the same absorber and scintillation materials as the HB. The brass plates are 79 mm-thick with 9 mm gaps to accommodate the scintillators. In terms of η and ϕ the granularity of the HE sub-detectors matches that of the HB for $|\eta| < 1.6$, but reduces to 0.17×0.17 for $|\eta| \geq 1.6$. The total length of the HE calorimeter, including electromagnetic crystals, is equivalent to approximately 10 interaction lengths.

The restriction in space inside the solenoid has led to the combined stopping power of the EB plus HB not providing sufficient containment for hadronic showers. To this end the HB is complimented by the additional HO calorimeter sub-detector lying outside the solenoid, so as to measure energies of late or extended hadronic showers. The HO calorimeter utilises the solenoid coil as an additional absorber equal to $1.4/\sin\theta$ interaction lengths, where θ is the polar angle; in this manner the HO calorimeter extends the minimum effective absorber thickness to 11.8 interaction lengths. The HO is divided into five rings in η, to match the structure of the muon system (see following section for description), being placed as the first sensitive layer in each of these rings. In the central ring only, where the total depth of the

HB is smallest, the HO has two layers of scintillator, either side of a steel absorber block, to increase the number of interaction lengths. The sizes and positions of the tiles in the HO are also intended to roughly map onto the layers of the HB, with a granularity of 0.087×0.087 in η and ϕ, such that consistent towers of hadronic calorimeter sub-detectors can be constructed. As with the HB and HE, the scintillation light is carried by wavelength shifting fibres to hybrid photo-diodes for measurement.

The HF calorimeters are cylindrical structures, with a radial extension between 12.5 cm to 130.0 cm, covering the range $3.0 < |\eta| < 5.2$. The faces of the HF systems are placed at $z = \pm 11.2$ m, covering the forward pseudorapidity region of $3 < |\eta| < 5.2$. To withstand this extreme radiation environment quartz fibres were chosen as the scintillation medium, exploiting the Cherenkov effect for signal generation. When a charged particle passes through a dielectric medium at a speed greater than the phase velocity of light in that medium, it produces *Cherenkov Radiation*. The charged particle polarises the molecules in the medium, which rapidly return to their ground state emitting a cone of visible electromagnetic radiation, centred on the incident charged particle [49]. The calorimeter is constructed from a steel absorber, comprised of 5 mm plates, with grooves for the fibres. The absorber is 165 cm long, equivalent to 10 interaction lengths, and is instrumented with two sets of fibres. One set runs the full length, and the second set starts at a depth of 22 cm, making it possible to distinguish showers generated by electrons & photons, which deposit a large fraction of their energy in the first 22 cm, from those generated by hadrons. The fibres run parallel to the beam line, and are bundled to give a resolution of 0.175×0.175 in η and ϕ. The front-end electronics for the HF are contained in Readout BoXes (RBXs) attached to the sub-detector. The Cherenkov light emitted in the quartz fibres is channeled to photomulitplier tubes (PMTs) housed in the RBXs. An RBX contains three front-end electronics boards, each reading out 8 PMTs, in total servicing half a wedge (10^o in ϕ) of the sub-detector.

3.4.2 Performance

Parts of the various HCAL subsystems were exposed to test beams, of electrons, pions, protons and muons, to measure the characteristics of the detectors and to obtain a reference calibration for them [50]. An ECAL module was also included in the test beam setup. The hadronic energy resolution of the combined barrel HCAL and ECAL is parameterised as:

$$(\frac{\sigma}{E})^2 = (\frac{S}{\sqrt{E}})^2 + C^2 \tag{3.6}$$

where E is in GeV, S is the stochastic term, and C the constant term. The values were measured as $S = 0.847 \pm 0.016$ GeV$^{1/2}$ and $C = 0.074 \pm 0.008$ [51]. The energy resolution was found to be similar in the endcaps, with the corresponding values for the HF of $S = 1.98$ GeV$^{1/2}$ and $C = 0.09$ [52]. As the forward jets typically have very high energies, this higher S-term value is expected for the HF, while still being able to provide the required energy resolution.

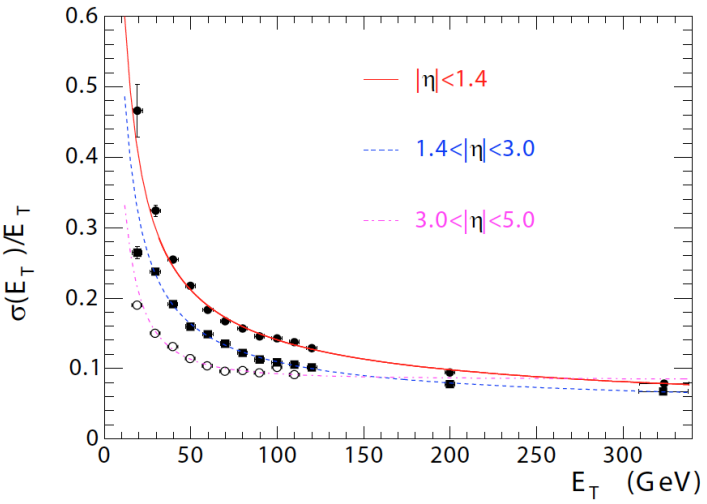


Figure 3.10: Jet transverse energy resolution, separated into the barrel (red), endcap (blue) and forward (pink) regions [40].

The transverse energy resolution of each sub-detector is shown in Fig. 3.10, and is typically of order 10% for high E_T jets.

3.5 Muon system

3.5.1 Overview

The muon system has three functions: muon identification, muon momentum measurement and triggering. The muon systems of CMS are designed to be sensitive to the whole kinematic region accessible in LHC collisions and as such have a large geometric coverage [53]. Optimised to the varying magnetic fields and radiation environments present at different areas outside the solenoid, the muon system is made up of three separate sub-detectors utilising different technologies, illustrated in Fig. 3.11.

In the barrel region Drift Tube (DT) chambers are used due to the low muon rate, backgrounds, and relatively low magnetic field strength. There are a total of 250 DTs in the barrel, covering the pseudorapidity range $|\eta| < 1.2$. The DT wire length is approximately 2.4 m in the $r - \eta$ direction, constrained by the longitudinal segmentation of the barrel iron yoke. The cell width, corresponding to the maximum drift path, was selected as 21 mm. This results in a maximum drift time of 380 ns in the Argon (15%) and CO_2 (85%) gas mixture.

The DT chambers are grouped into 4 stations, forming concentric cylinders around the beam-line, divided into 5 wheels running along the axis of the beam-line. The stations are interspersed between the layers of the iron yoke. There are a total of sixty DT chambers in the first three stations, and seventy in the outermost. The DT chambers typically consist of three superlayers (SLs), each made of 4 layers of rectangular drift cells staggered by half a cell distance. The wires in the two outer SLs run parallel to the beam-line providing a track measurement in the magnetic bending plane ($r - \phi$). In the central SL, the wires run orthogonal to the beam-line measuring the z-

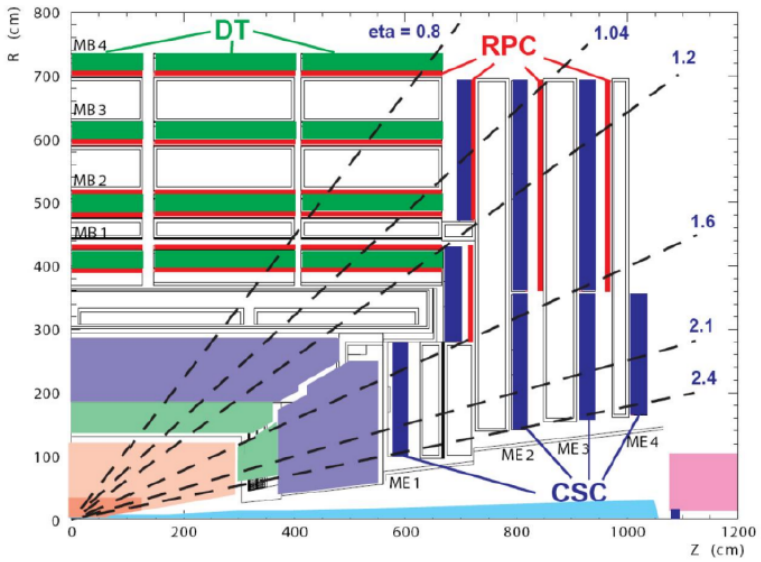


Figure 3.11: A schematic diagram of one quadrant of the muon system showing the drift tubes (DTs) in the barrel, the cathode strip chambers (CSCs) in the endcaps and the resistive plate chambers (RPCs) [54].

position along the beam. The spatial separation of the $r - \phi$ measuring SLs within each DT chamber offers an improved angular resolution. The outermost station is constructed without the third, z-measuring SL and therefore measures only the ϕ coordinate.

Cathode Strip Chambers (CSCs) are used in the endcap regions, covering the pseudorapidity range $0.9 < |\eta| < 2.4$, overlapping with the barrel region. The CSCs provide the functions of precision muon measurement and muon trigger in one device. This technology is used here due to their fast response time ($\sim$4.5 ns) and ability to operate in a high non-uniform magnetic field. The CSCs are multi-wire proportional chambers comprised of 6 anode wire planes interleaved among 7 cathode panels. The anode wires run azimuthally to provide a measurement of the polar angle of a track. The CSCs are placed in a radial configuration, interspersed between the iron yoke plates, giving measurements with a spatial resolution, in $r - \phi$, of 75 μm for the inner

chambers and 150 μm for the outer chambers.

The third detector employed by the muon system is Resistive Plate Chambers (RPCs), gaseous parallel-plate detectors that are used in both the barrel and the endcaps. An RPC consists of two parallel electrodes constructed from graphite coated plastic. The coverage of the RPCs extends up to pseudorapidities of $|\eta| < 1.6$ currently, and will be extended to $|\eta| < 2.1$ for high-luminosity running. They have a fast response time, measured at 1.26 ns for the design used on CMS. As this is far shorter than the 25 ns between consecutive LHC bunch crossings they can unambiguously identify the bunch crossing any given muon track is associated to. To this end, the RPCs are primarily used for triggering, but also provide adequate spatial resolution for use in track reconstruction. In total there are six layers of RPCs in the barrel, one on either side of the DT chambers in the first and second muon station and one on the inner side of the DT chambers in the outer two muon stations. In addition, there are another three RPCs layers in the endcaps.

3.5.2 Performance

Cosmic ray muon data, taken in 2008 during the Cosmic Run At Four Tesla (CRAFT), was used to study the performance of the sub-detectors of the muon system; the details of which can be found elsewhere [55, 56, 57].

The muon reconstruction performance was measured more recently in 2010 using 40 pb^{-1} of data at $\sqrt{s} = 7$ TeV [58]. The identification efficiency for muons with p_T greater than a few GeV/c is found to be above 95% in all detector regions, with a maximum misidentification rate of 1%. For muons with p_T in the range $20 - 100$ GeV/c, the relative resolution is found to be between 1.3% and 2% in the barrel, and better than 6% in the endcaps. For highly-energetic muons, with p_T of approximately 1 TeV, the resolution is found to be better than 10%.

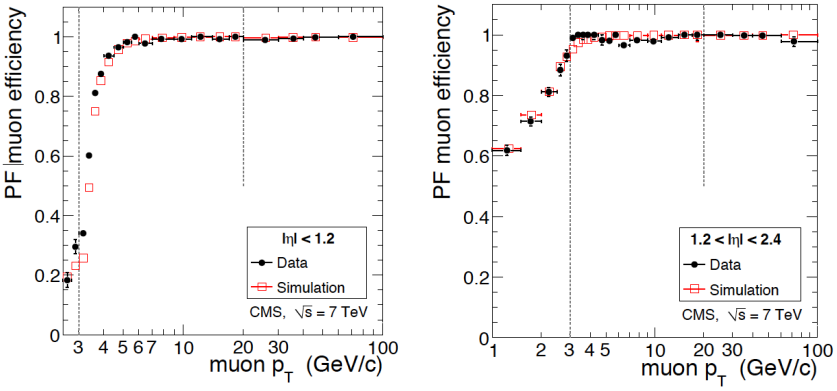


Figure 3.12: Tag-and-probe results for the muon efficiency in data compared to simulation. Given that a tracker track exists, the plots show the efficiency as a function of muon p_T for Particle Flow muons in the barrel and overlap regions (left), and in the endcaps (right) [58].

Figure 3.12 shows the muon reconstruction and identification efficiency, ϵ_{rec+id}, for the different detector regions in data and simulation. The measurements were made using $J/\psi \rightarrow \mu^+\mu^-$ events for $p_T < 20$ GeV/c and $Z \rightarrow \mu^+\mu^-$ events for $p_T > 20$ GeV/c; the Tag-and-Probe method was used [58], described in Chapter 6.1. The muons are reconstructed using the Particle Flow algorithm, which is described in Chapter 4.

3.6 Trigger system

3.6.1 Overview

Due to the high collision rate of up to 40 MHz at the LHC and the large number of read out channels, it is not possible to store all events that are produced. Rather, only a drastically reduced rate of events, that are potentially interesting, can be recorded and processed. The task of the trigger is to decide very quickly whether or not an event is worth recording. The CMS experiment uses a two-level online trigger system with the aim to keep events of interest whilst reducing the data rate to about 100 Hz due to limited

storage and computing capacity.

The first level is hardware-based and is referred to as the **Level-1 Trigger** (L1). It consists of custom-designed, highly programmable electronics with a maximum output rate of 100 kHz; sitting partly on the detector and partly in the underground control room approximately 90 m from the detector cavern. The second level is software-based, referred to as the **High-Level Trigger** (HLT).

3.6.2 L1 Trigger

As the L1 trigger has to process a very high rate of collision data it is not possible to use the full set of sub-detectors in CMS; instead it uses information from only the calorimeters and the muon system. The L1 trigger consists of local, regional and global components, roughly split into muon and calorimeter pipelines, under a final Global Trigger stage which makes the final decision on each event. An overview of the system is shown in Fig. 3.13.

For the calorimeter trigger pipeline, the calorimeters are divided into regions known as *trigger towers*. These consist of a readout from both the ECAL and the HCAL in a region of the detector, where the ECAL crystals are segmented to match the coarser granularity of the HCAL towers. Energy deposits in these trigger towers form the first stage of the calorimeter trigger pipeline, referred to as Trigger Primitive Generators. These are then passed to the Regional Calorimeter Trigger, which identifies electron/photon candidates. These electron/photon candidates are split into isolated and non-isolated collections, with four of each being passed to the Global Calorimeter Trigger. At this stage jets are reconstructed and the missing transverse energy is calculated along with other global quantities. This information is then passed to the Global Trigger.

The muon trigger system uses all three sub-detectors (DT chambers, CSCs and RPCs), which identify the bunch crossing to which a signal is related. The DT chambers provide information of track segments in the ϕ

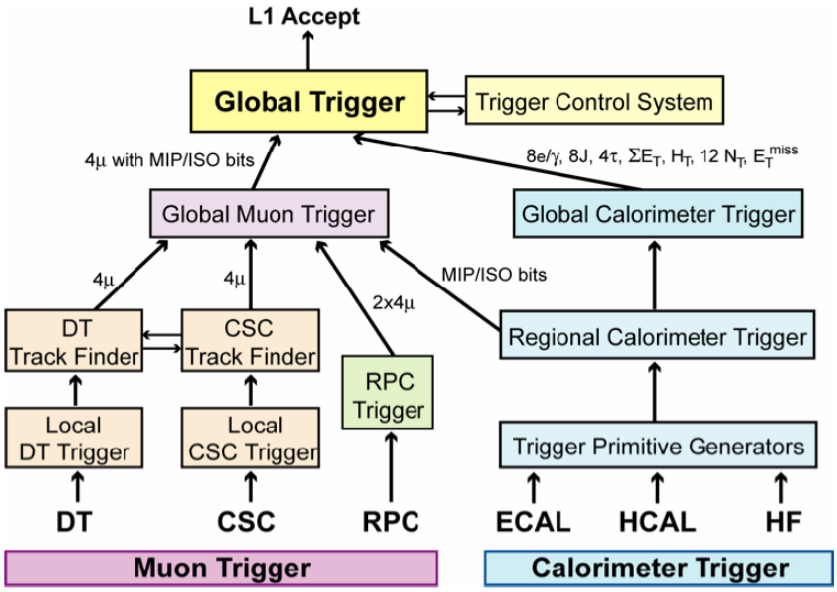


Figure 3.13: Architecture of the Level 1 (L1) trigger system [40].

direction and hit patterns in the η direction, the CSCs produce 3-dimensional track segments and the RPCs produce their own track candidates. Up to four muon candidates from the DT chambers and CSCs are passed to the Regional Muon Trigger (RMT), plus four each from the barrel and forward RPCs. At the Global Muon Trigger, the RPC muon candidates are matched to those from the other two sub-detectors and the resulting combined candidates are then sorted by transverse momentum and quality. Four muon candidates are then passed to the Global Trigger.

The final step for the L1 trigger process is the Global Trigger. This uses all of the information passed from the individual trigger steps to make a final decision whether an event should be accepted for further examination by the HLT. Completely programmable logic is used to apply trigger algorithms to make the decision based on the reconstructed physics objects and global quantities. A total of 128 algorithms can be run in parallel by the Global Trigger.

3.6.3 HLT

The HLT makes the final decision as to whether an event is permanently read out and stored for offline analysis. The software-based HLT runs on a large farm of commercial multi-core computers with access to all the information in the event, including information from the L1 trigger. Sophisticated algorithms are run, similar to those in offline physics analyses, capable of the full reconstruction of event topologies and filtering according to desired physics criteria. This enables the data rate to be reduced from 100 kHz, at the L1 trigger level, to the order of 100 Hz. Several thousand 3 GHz computer cores are used, based on an estimated mean processing time for the HLT of 50 ms. To accommodate the necessary longer processing time of the HLT, sub-detector front end systems are used to temporarily store data in 40 MHz pipelined buffers. The pure software implementation of the HLT allows for easy changes to the algorithms, so they can be optimised according to the running condition of the LHC.

Chapter 4

Simulation samples and reconstruction

Monte Carlo (MC) simulation is an essential part of particle physics analyses, necessary for the development of methods as well as the validation of results. The production of MC samples for use in CMS data analyses is done centrally by the *Generator Group*. The first section of this chapter describes the programs used to generate these samples, and details the process of correcting them to higher-order cross sections.

CMS has developed an algorithm that combines the information from all of its sub-detectors to identify and reconstruct all particle types produced in a collision event, i.e. electrons, muons, photons, charged hadrons and neutral hadrons. This is not a partial identification of a single particle type, but a full recreation of the event, with every deposit and track accounted for. By making full use of detector information the particle identification performance is significantly improved. The resulting list of particles, produced by the algorithm for each event, is then used to construct various 'higher-level' objects, such as jets (including b-tag information) and missing transverse energy. This algorithm is known as *Particle-flow* (PF) within CMS [59]. Variations of this technique have been used at the SLAC experiments, and is central to the SiD (Silicon Detector) concept for a future Linear Collider

[60]. The rest of this chapter will detail the various object reconstruction methods used on CMS, focusing on the PF algorithm.

4.1 Simulated samples

Single top quark events have been simulated with the POWHEG event generator v301 [25], designed to describe the full next-to-leading order (NLO) properties of these processes. POWHEG is used to generate separate MC samples with top or antitop quarks, for all three single top production channels. The events generated using POWHEG are then passed to PYTHIA v6.4.24 [61] to simulate the showering and hadronisation of the patrons.

The leading-order (LO) MadGraph matrix-element based generator v5.1.1 [62] is used to produce $t\bar{t}$ samples and inclusive single boson production samples: $W^{\pm}$ or Z^0 to light or heavy partons (separate samples referred to as W+jets and Z+jets, respectively). PYTHIA6 is again used to simulate the showering and hadronisation of the partons, with the matrix element to parton-shower matching performed using the Kt-MLM algorithm [63]. This matching process introduces a systematic uncertainty for the MadGraph generated samples, detailed in section 6.2.5.

The remaining background samples are entirely simulated using PYTHIA6, including diboson production and QCD processes enriched in events with muons produced in the decay of b-quarks, c-quarks and long-lived hadrons. The decay of τ leptons is handled by the separate TAUOLA package [64], containing sub-programs for several different τ decay modes. It is used in conjunction with all the MC generators listed to add tau decay information to every simulation sample, apart from the QCD background samples. The CTEQ6.6M parton distribution functions [65] are used in all samples.

The simulated events are then passed through a simulation of the CMS detector. A full reconstruction of the CMS detector has been created in GEANT4 [66, 67] for this purpose.

In the generation of all simulation samples, the top quark mass was taken to be $m_t = 172.5$ GeV/c^2. The Z+jets sample was generated with the constraint that the invariant mass of the lepton pair, m_{ll}, is greater than 50 GeV/c^2. The predicted inclusive QCD production cross section is so large that generating a single sample for such processes with simulation is neither practical nor effective. Therefore, the QCD multijet events that are of interest for this analysis are produced with generator-level filters, to ensure that only the heavy flavour enriched events are kept. This increases the number of simulated events that are likely to pass the analysis event selection. For this analysis, the QCD samples are generated with a total event $p_T > 20$ GeV/c and with at least one muon of $p_T^{\mu} > 15$ GeV/c - and as such is called 'muon-enriched' QCD.

Several 'electron enriched' QCD samples were used, with an isolated electron (*EMenriched*) or with an electron originating from the decay of b- or c-quarks (*BCtoE*), covering the same p_T region as the muon-enriched sample. However, it was found that all events from all the samples were eliminated by the first kinematic requirement of the analysis event selection (see Chapter 5). As such, these samples are not listed in Table 4.1 as they did not contribute to the analysis result.

The simulation samples are corrected after generation to a consistent higher order so as to improve the description of the data. Approximate NNLO results are derived for the top quark p_T distribution and $t\bar{t}$ production cross section ($\sigma_{t\bar{t}} = 163^{+11}_{-10}$ pb) by adding NNLO soft gluon corrections (derived from Next-to-Next-to-Leading-Log resummation) to the exact generated NLO (LO) quantities for the single top ($t\bar{t}$) samples [20]. The tW-channel top and antitop quark simulation samples are then combined, for use in this analysis, to give a total NNLO cross section of 15.74 pb. All single top simulations are inclusive, meaning that they contain both the leptonic and hadronic $W^{\pm}$ boson decay channels. The inclusive single boson production samples are both normalised to full NNLO calculations of the inclusive Z+jets and W+jets cross sections using the FEWZ program [68, 69], corresponding to 3048 pb and 31314 pb, respectively. The diboson production

samples of $W^{\pm}W^{\mp}$, $W^{\pm}Z^0$ and Z^0Z^0, with the corresponding inclusive cross sections of 42.9 pb, 18.3 pb, and 7.67 pb, respectively, are calculated to NLO with the MCFM program [70]. The QCD sample remains uncorrected at LO. A complete analytic result for the NLO QCD cross section [71] was published too late to be used to generate the simulation. However, as QCD is not a significant background for this analysis the restriction to LO is not considered an issue. The LO cross section quoted for the QCD sample is the effective cross section including the filter efficiencies. The simulated samples, with their corresponding production generators and cross sections, are summarised in Table 4.1.

Table 4.1: Simulated processes used in the analysis, including reference name, cross section, total number of generated events and generator used. 'DR' denotes the diagram removal scheme, as explained in section 2.3.

Reference	Process	$\sigma\ [pb] \cdot BR$	# of events	Generator
tW	single top, tW-channel DR (t, inclusive)	7.87 (NNLO)	813743	powheg-pythia6-tauola
	single top, tW-channel DR ($\bar{t}$, inclusive)	7.87 (NNLO)	689462	powheg-pythia6-tauola
single top	single top, t-channel (t, inclusive)	41.92 (NNLO)	3337875	powheg-pythia6-tauola
	single top, t-channel ($\bar{t}$, inclusive)	22.65 (NNLO)	1943627	powheg-pythia6-tauola
	single top, s-channel (t, inclusive)	3.19 (NNLO)	259777	powheg-pythia6-tauola
	single top, s-channel ($\bar{t}$, inclusive)	1.44 (NNLO)	137889	powheg-pythia6-tauola
$t\bar{t}$	$t\bar{t}$	163 (NNLO)	3160707	madgraph-pythia6-tauola
Z+jets	$Z/\gamma^*(\rightarrow l^+l^-) + jets$	3048 (NNLO)	26523984	madgraph-pythia6-tauola
W+jets	$W(\rightarrow l\nu) + jets$	31314 (NNLO)	49708092	madgraph-pythia6-tauola
diboson	$W^{\pm}W^{\mp}$	42.9 (NLO)	4223785	pythia6-tauola
	$W^{\pm}Z^0$	18.3 (NLO)	3863081	pythia6-tauola
	Z^0Z^0	7.67 (NLO)	4188624	pythia6-tauola
QCD	μ-enriched QCD	84679.3 (LO)	25079892	pythia6

4.2 Preliminary object reconstruction

The PF event reconstruction aims to identify and reconstruct all stable particles in the event using a combination of all CMS sub-detectors to determine the particle direction, energy and type. The recorded data and simulation samples are both reconstructed using the same PF algorithm to get directly comparable results.

However, PF is not a standalone algorithm, rather an additional 'layer' of reconstruction that utilises pre-existing identification and reconstruction algorithms. The fundamental elements of the PF reconstruction are the charged-particle tracks, calorimeter clusters and muon tracks provided by preliminary reconstruction algorithms.

4.2.1 Iterative tracking

Charged hadrons are measured in the CMS tracker system with a greater resolution than in the calorimeters for p_T of up to several hundreds of GeV/c. In addition, the tracker provides a precise measurement of the charged particle direction at the production vertex, before any deviation by the magnetic field, which is needed for extrapolating tracks to the calorimeters. As the charged particles of a jet carry approximately two thirds of the jets energy, the tracker is the most important contributor of information to the PF algorithm. This makes it important to reduce fake tracks as, if the jet momenta were randomly distributed, it would lead to potentially large energy excesses when reconstructing the event. It is thus highly beneficial to the effectiveness of the PF algorithm to have a high tracking efficiency and low fake rate. To achieve this an iterative tracking method is used to identify track candidates [72].

The process of track reconstruction can be broadly summarised into 5 stages:

- Local reconstruction - signals in the strip and pixel silicon detectors that make up the tracker are clustered into *hits* produced by charged particles;
- Track seeding - initial track candidates are identified for use in the full track reconstruction. The seed defines initial trajectory parameters and errors for the track;
- Pattern recognition - using a combinatorial variation of a global Kalman filter (the Combined Kalman Filter) [43], tracker hits are grouped into possible particle trajectories. These are then propagated in parallel working outwards layer by layer, to prevent bias and double-counting;
- Fitting - a refit using the Kalman filter, starting at the innermost hit with the seeding estimate, to reduce any bias from the seeding stage, and then a second smoothing fit working from the outer layers inwards, to reduce bias from building the tracks;
- Quality - tracks are filtered in order to reject fake tracks and are labeled with a quality, either *loose*, *tight* or *highPurity*.

In the iterative tracking approach, these stages are performed 6 times, starting from the *zeroth* iteration. At the end of each iteration, the hits associated with high quality tracks are removed and the next iteration performed with different, typically looser, requirements on a progressively smaller subset of reconstructed hits [72].

Tight reconstruction criteria are implemented in the zeroth and first iterations, requiring pixel-triplet then pixel-pair seeds, to reconstruct the majority of high p_T primary vertex tracks. The second and third iterations require looser p_T seeding constraints so as to identify low p_T tracks. This relaxed criteria improves the tracking efficiency, whilst the fake rate is kept low by having a reduced pool of usable hits for reconstruction. The fourth and fifth iterations switch to using seeds from the Strip Tracker stereo layers and use relaxed constraints on the primary vertex. This reconstructs particles produced outside the Pixel Tracker volume, such as from photon conversions or

the decay of long-lived particles such as K_S^0 and K_L^0.

With the iterative technique, charged particles with as little as three hits, with p_T as little as 150 MeV/c and a production vertex greater than 50 cm away from the beam axis, are reconstructed with a fake rate of the order of 1% [72].

4.2.2 Primary vertex reconstruction

The primary vertex reconstruction makes use of the reconstructed charged particle tracks in the event. Track selection is made with requirements on the transverse impact parameter significance with respect to the LHC beam line [73]. They are also required to have a minimum number of strip and pixel hits in the tracker and to pass a cut on the normalised χ^2. From the tracks that meet these requirements, vertex candidates are constructed by clustering tracks according to their z-coordinates at the point of closest approach to the beam line. The maximum separation between a track and its nearest neighbour is taken to be $z_{sep} = 1$ cm for grouping into a common vertex. The vertex candidates, containing two or more tracks, are then passed to the adaptive vertex fitter which performs a three dimensional fit to reconstruct all possible vertices from combinations of the candidates [74]. The reconstructed vertices are then ranked according to the scalar sum of the transverse momenta squared of all the tracks associated to the vertex. The fit also returns other vertex parameters, such as position and the number of degrees of freedom, n_{dof}, the latter being an indicator of the fit quality. The performance of the vertex reconstruction is intrinsically linked to that of the tracking, particularly the number of tracks used in the fit and their respective p_T.

4.2.3 Calorimeter clustering algorithm

The pre-existing CMS calorimeter clustering algorithm has been used for PF event reconstruction to measure the energy and direction of particles with a high detection efficiency [75]. The main aim of this algorithm is to improve the energy measurement of charged hadrons with poorly resolved or high-momentum tracks; this helps in separating them from the energy deposits of neutral hadrons or photons. The clustering is performed separately in each sub-detector system, i.e. the ECAL barrel, ECAL endcaps, HCAL barrel, HCAL endcaps and the first and second layers of the preshower. In the Hadronic Forward Calorimeter, no clustering is performed due to its larger cell sizes, so in this system each cell corresponds to a single cluster.

First, local calorimeter cells with energy deposits over a certain energy are identified as cluster seeds. These seeds are summed together with nearby cells that share at least one common side, forming topological clusters. These clusters are required to have energy greater than a threshold value corresponding to two standard deviations above the ECAL and HCAL electronics noise. The thresholds are: 80 MeV for the ECAL barrel, 300 MeV in the ECAL endcaps and 800 MeV in the HCAL [75]. However, due to the effects of Bremsstrahlung and photon conversions, the energy deposited in the ECAL can be spread out in the ϕ direction. This spread energy is grouped together by the clustering algorithm, forming clusters of topological clusters, which are extended in ϕ. It is these 'superclusters' that are linked with particle tracks in the main PF algorithm.

4.2.4 Muon reconstruction

With respect to identifying muons, the PF algorithm is more of an expansion of the existing algorithm than a new approach, applying a single additional constraint on top of pre-PF reconstruction methods. In the standard CMS muon reconstruction, tracks are reconstructed independently in the silicon tracker (referred to as *tracker tracks*) and in the muon system (*muon tracks*).

These two track types form the basis for two muon reconstruction approaches [54]:

- **Tracker Muon reconstruction** – any tracker track with $p_T > 0.5$ GeV/c and $|p| > 2.5$ GeV/c is considered as a muon candidate and is extrapolated out to the muon system, taking into account expected energy loss and the uncertainty due to multiple scattering. If the muon system contains even one muon segment matching the extrapolated track position, the tracker track qualifies as a *tracker muon track*. Often referred to as the 'inside-out' approach;
- **Global Muon reconstruction** – starting from a muon track, a corresponding track in the silicon tracker is identified. The hits from these two tracks are combined into a *global muon track* by performing a global fit. Often referred to as the 'outside-in' approach;

In the PF scheme, muon candidates are required to pass both the global and tracker muon reconstructions, which are then passed to the main PF reconstruction algorithm for further validation.

4.2.5 Electron identification

The majority of charged particles are massive enough for multiple Coulomb scattering to affect them when crossing between materials. It is for that reason that a Kalman filter is the default track reconstruction algorithm for CMS, as it can incorporate these effects as Gaussian fluctuations. However, with electrons the dominant effect is Bremsstrahlung emission, which is a highly non-Gaussian process. In addition, the changes in electron trajectory caused by Bremsstrahlung emission can even result in the standard Kalman filter pattern recognition either failing to follow the complete electron path or producing a fitted track with all the hits but an unacceptably large χ^2.

To this end a dedicated electron track reconstruction algorithm is used to model the asymmetric Bremsstrahlung energy loss [76]. A relaxed Kalman

filter is used to reconstruct the entire electron trajectory, on which a refit is performed using a Gaussian-Sum Filter (GSF), a non-linear generalisation of the Combined Kalman Filter (CKF) [77]. In the GSF, the state vectors describing each track are treated as Gaussian mixtures, rather than a single Gaussian in the case of the CKF filter. By using a larger number of Gaussian components, the GSF can better handle the changes in electron trajectory, albeit with a large increase in computation time.

The PF scheme utilises two complimentary seeding strategies for reconstructing electron tracks, collectively known as *electron identification* [78]. The *ECAL-driven approach* uses an ECAL supercluster as the seed, made up of both electron and Bremsstrahlung energy deposits, identified by the calorimeter clustering algorithm. From the centre of the supercluster, the position of hits in the pixel detector is inferred and the general track seeds that match these predicted positions are selected. It is optimised for isolated, high-p_T electrons, where the number of potential tracker seeds is limited and the ECAL signal will not be affected by overlapping jet energy deposits that could bias the supercluster energy or position.

The *tracker-driven approach* was developed for PF to increase the seeding efficiency for non-isolated and low p_T electrons by using high-purity Kalman Filter tracks for the seeds [59]. Iterative tracking was introduced for this purpose, before becoming the default track reconstruction method. When the Bremsstrahlung emission is negligible, the electron momentum can be determined with reasonable precision using the CKF, and the track extrapolated to the ECAL surface to allow matching with the closest energy deposit. If the ratio of the cluster energy to the track momentum is close to one, the track is selected. However, if the Bremsstrahlung energy loss is more substantial, the track characteristics are exploited for electron reconstruction.

Firstly, a selection based on the number of tracker hits and χ^2_{KF} is carried out, followed by a GSF refit of each electron track. A boosted decision tree multivariate estimator [79] is then determined using these initial selection quantities, along with the χ^2 of the GSF refit, as well as the energy loss as

measured from the track.

The seeds obtained from the tracker-driven and ECAL-driven approaches are then merged into a single collection, whilst retaining the method by which each seed was obtained. A GSF is run on the seed collection to determine the final track properties. By using a GSF fit, more tracker hits are used in the reconstruction, leading to better modelling of the electron momentum, vital for energy loss estimation. These GSF electron tracks are used by both the standard and PF reconstruction algorithms.

4.3 Particle Flow algorithm

Any given particle is generally expected to register in several CMS sub-detectors, such as energy clusters in the ECAL and HCAL and/or charged particle tracks in the tracker. The PF scheme must be able to efficiently reconcile these separate elements to identify and reconstruct all particles in a given event. It is important that all the elements resulting from a given particle interaction are properly linked, as missing connections could result in the mis-identification of additional particles or double-counting of energy in the event.

The PF scheme uses a link algorithm to pair event elements into manageable 'blocks' of likely connected sub-detector readings. These simplified blocks form the input to the PF reconstruction algorithm. By combining, typically, only one to three elements into each block ensures that the event complexity does not greatly affect the performance of the algorithm.

4.3.1 Link algorithm

The link algorithm considers pairs of elements in the event, with the quality of the link between them quantified by their relative distance from each

other. Generally, starting with a track, links are made to calorimeter clusters by extrapolating from the last measured hit in the tracker to the calorimeter system concerned, be it ECAL, HCAL or preshower. A link is made if the extrapolated position of the track falls within the boundaries of the calorimeter cluster. The link distance is defined as ΔR, in the (η, ϕ) plane, between the extrapolated track position and the cluster position. In all cases linking with calorimeter clusters, the cluster boundaries are widened by up to a cell in each direction to account for potential non-uniformity in the calorimeters, such as gaps between the cells or modules and the uncertainty on the position of the shower maximum.

If an ECAL cluster lies within the envelope of an HCAL cluster they are linked. The link distance is defined as ΔR between the two cluster positions. To identify Bremsstrahlung photons, tangents are extrapolated from the intersection points between a given track and each of the tracker layers out to the ECAL. An ECAL cluster is linked to the track as a potential Bremsstrahlung photon if the extrapolated tangent falls within the cluster boundaries. Reconstructed global muons have already linked charged particle tracks in the tracker with muon tracks in the muon system. These are passed straight to the PF reconstruction algorithm.

4.3.2 Particle flow reconstruction

The reconstruction and identification of particles from each block of elements, returned by the link algorithm, is then performed by the PF algorithm. Similar to the iterative tracking, when tracks and/or clusters are matched to form a particle, they are removed from the block and any further reconstruction efforts.

First, the combined momentum of each global muon candidate is compared to that returned from the associated track. If the momenta match within three standard deviations it qualifies as a "particle-flow muon" and the corresponding track is removed from the block.

Electron identification is carried out next, and, as with muons, the majority of the reconstruction has already been done. To discriminate between electrons and charged hadrons, each GSF electron track requires an associated ECAL cluster. The lateral shape of this ECAL cluster must also be narrow to be considered as being from a genuine electromagnetic shower. The outermost measurement of the GSF track is associated to the nearest ECAL cluster. If one is not found the GSF track is removed from consideration for electron reconstruction.

Electron tracks seeded from the tracker-driven method retain information of their initial KF tracks, such as the number of hits and χ^2_{KF}, which are used to avoid pion misidentification. The supercluster from the ECAL-driven approach contains, by construction, the information of the ECAL cluster associated to the GSF track and the ECAL cluster(s) associated to the Bremsstrahlung track tangent(s). This various information is used to construct a multivariate estimator, onto which a kinematic requirement is placed. Electron candidates that pass qualify as "particle-flow electrons" [59]. The corresponding track and ECAL clusters (including those identified as Bremsstrahlung photons) are then removed from the block.

After the lepton identification has been performed, the remaining elements in a block can be from charged hadrons, photons or neutral hadrons. If a track has links to several HCAL (ECAL) clusters, the closest cluster is favoured and the other links discarded. It then becomes a matter of comparing the sum of the energy in the calorimeter clusters, E, to the track momenta, p, in these blocks. If the quantities are comparable, "particle-flow charged hadrons" are identified, one per track in the system. However, if the clustered energy is larger than the track momentum, E > p $+\sigma_E$ where σ_E is the uncertainty in the energy measurement, then it is assumed to be either a photon or a neutral hadron, depending on how the energy is distributed. If the observed excess is larger than the total ECAL cluster energy, the ECAL deposit is identified as being from a "particle-flow photon" and the remaining excess (E - p) attributed to a "particle-flow neutral hadron". Otherwise the

excess is considered as only a PF photon. Any remaining ECAL and HCAL clusters, either previously not linked to any track or passed over for a different cluster, are taken to be PF photons and PF neutral hadrons, respectively.

4.4 Higher-level object reconstruction

This section describes the techniques available within CMS to identify and reconstruct higher-level physics objects such as, jets, missing transverse energy and b-jets.

4.4.1 Jets

Jets are reconstructed in CMS data using the *anti-k_t* algorithm [80]. By using successive combination, the *anti-k_t* algorithm produces jets from a list of object positions and transverse momenta, never assigning a particle to more than one jet [81]. The algorithm proceeds by calculating the quantities d_{ij} for combinations of pairs of objects and d_{iB} for individual objects, defined by equations 4.1 and 4.2, respectively:

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

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

where k_t is the particle(s) transverse momentum, $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ (y denoting the rapidity) and R is the size parameter which is set to 0.5 as standard for CMS jets.

The calculated values of d_{ij} and d_{iB} are then compared to identify the smallest, d_{min}. If d_{min} is from an individual object the particle cannot be merged with another, so is labelled as a completed jet and removed from further consideration. If d_{min} is found from equation 4.1 the two particles

are merged into a 'protojet', with its position and transverse momentum calculated as:

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

This process repeats until all objects are clustered into jets. Although this is the only jet reconstruction algorithm used on CMS, it is made to produce four distinct types of jet objects [82]:

- **Calo-jets** are reconstructed from energy deposits in the ECAL and HCAL calorimeter towers [82]. If the combined deposited energy exceeds 1 GeV a *calo-jet* is produced;
- **JPT-jets** (Jet-Plus-Tracks) utilise the CMS tracking systems to improve the p_T resolution and response of calorimeter jets [83];
- **PF-jets** are reconstructed using the list of particles produced by the PF algorithm;
- **Gen-jets** are produced from the cluster energies of MC generator-level particles, in simulation-based studies. Used to validate jet methods.

The tracker and calorimeter are prominent in jet reconstruction as the typical jet energy fractions carried by charged particles and photons are 65% and 25% respectively [75]. The remaining 10% is carried by neutral hadrons, which are hampered by poor hadron calorimeter resolution and associated calibration corrections. This means that the PF algorithm, by fully reconstructing charged and neutral particles, not just energy deposits and/or tracks, can reconstruct with better precision the energy and direction of a given jet, compared to the other jet types. For this reason, and the more robust and intrinsic redundancy of the PF algorithm reconstruction, PFjets are used in this analysis.

A detector-level jet, that is reconstructed and measured from the detector inputs, typically has a different energy than the corresponding generator-level jet from the simulation, even when using the same jet algorithm. The main cause of the discrepancy is the expected non-uniform performance by the CMS calorimeters and other differences between the CMS simulation and actual performance, such as electronics noise and event pile-up. To this end, the energy of any reconstructed jet is calibrated to correct for these effects. On CMS, the jet energy corrections are factorised into separate levels, with each level correcting for a different effect by scaling the jet four-momentum [84]. The first three levels, and those used in this analysis, are:

- **L1 Pile-up** - corrects for additional measured energy which does not belong to the hard process, e.g. from electronics noise or pile-up;
- **L2 Relative Jet Correction** - an η dependent scale factor to correct for variations in jet response, derived from events with at least two hard jets, one of which is reconstructed in the uniform barrel region, $|\eta| < 1.3$;
- **L3 Absolute Jet Correction** - addresses the p_T dependence of the reconstructed jet by correcting back to the generator-level jet using corrections determined from $\gamma^\star$/Z+jets events;

Detailed descriptions of these jet corrections can be found in [84]. The final correction is a product of these individual factors, with their related uncertainties used to derive the associated systematic uncertainty (see Section 6.2.2). There also exist flavour and electromagnetic energy fraction specific corrections, but these are not yet supported for all types of reconstructed jets nor are they implemented in this analysis.

4.4.2 Missing transverse energy

The missing transverse energy (E_T^{miss}) in an event is a useful indicator of the presence of weakly interacting particles, such as neutrinos, and as such

is a useful tool in identifying events which contain leptonically decaying $W^{\pm}$ bosons. The missing transverse momentum is the opposite sign of the transverse-momentum vector sum of all final-state particles reconstructed in the detector. The E_T^{miss} is then taken as the magnitude/modulus of this vector. CMS has developed three distinct reconstruction algorithms to determine E_T^{miss}; what distinguishes them is which sub-detectors are used to derive the energy and/or momentum measurements:

- **Calorimeter** E_T^{miss} – utilises only calorimeter energies and tower geometry [85];
- **Track Corrected** E_T^{miss} – tracks reconstructed in the inner tracker are used to correct the calorimeter based E_T^{miss} measurement, to compensate for the non-linearity of the CMS calorimeter [86];
- **Particle-flow** E_T^{miss} – calculated using the complete PF scheme, incorporating information from all CMS sub-detectors. Individual reconstructed particles take the place of the calorimeter energy deposits used in the previous approaches;

The calorimeter E_T^{miss} method requires several corrections due to the non-uniformity of the calorimeter towers. Jet Energy Scale (JES) corrections, referred to as Type 1, are made to adjust the measured calorimeter jet energy back to the energy of the final state particle-level jet. The corrections are defined by the JES group, and determined using simulation, where the generated jet energy values are known. JES is discussed in more detail in Chapter 6. Calorimeter E_T^{miss} must also be corrected for possible muons present in the event, by replacing the muon energy deposits in the calorimeter with the momentum measurement from the muon system and central tracker. The Type 1 corrections are less reliable at low energies, typically below 20 GeV, which can distort the corrected E_T^{miss} distribution. To remedy this, so-called Type 2 corrections are made to the unclustered energy and jets below this threshold.

In the case of PF E_T^{miss}, by reconstructing the charged hadrons from tracker information, and with a correction factor already applied to the neut-

ral hadron and photon energies, PF reconstructed particles are naturally closer to the correct energy scale [75]. This improvement in energy scale means that when PF E_T^{miss} was first commissioned for use in analyses, Type 1 and the secondary Type 2 corrections were not considered necessary. It should be noted, however, that these corrections for PF E_T^{miss} have recently been investigated [87], with Type 1 PF corrections being introduced for single top quark analyses using 2012 collision data.

It has been shown [87] that algorithms using tracker information show an improved E_T^{miss} resolution compared to purely calorimeter derived E_T^{miss}, with the global event reconstruction of PF performing the best. For this reason (uncorrected) PF E_T^{miss} is used in this analysis, as per the recommendations of the CMS Top Physics Analysis Group (PAG).

4.4.3 b-Jet identification

The b-quark is the second heaviest quark and can decay into either an up or charm quark via the weak interaction. However, these decays are suppressed by the CKM matrix (see Section 2.1) which results in b-mesons having a longer lifetime than expected for such a massive particle. This enables the particles they hadronise into to travel a measurable distance within the detector geometry before decaying; unlike the hadrons formed from the lighter quark flavours which can escape the detector without decaying. It is this feature that can be exploited to identify b-quarks and jets arising from them.

The identification of jets associated with the production of b-quarks is an invaluable tool for studying the top quark due to the almost 100% branching ratio of the top quark to a $W^{\pm}$ boson and a b-quark (see Section 2.2). By accurately identifying jets as originating from b-quark decay, it is possible to reduce the backgrounds from other channels, such as W+jets, Z+jets and QCD multijet events, which are dominated by jets from light-flavour quarks (u, d, s), c-quark fragmentation and gluons.

CMS uses several robust algorithms, of varying complexity, to distinguish between b- and light-flavour jets [88]. Each algorithm draws on a variety of reconstructed objects, such as tracks, vertices and leptons, either individually or in combination, to produce a discriminator value for each jet in the event; the higher this value the more likely it is that it originated from a b-quark. So-called 'working points' are defined for each algorithm by the B-Tagging Physics Object Group, corresponding to a nominal misidentification probability for light-flavoured jets of 10% (loose), 1% (medium) and 0.1% (tight), respectively, at an average jet p_T of approximately 80 GeV/c. Broadly speaking, the algorithms are grouped into two categories: those that use the impact parameters of particle tracks to identify b-jets and those that use reconstructed secondary vertices.

The impact parameter (IP) is defined as the distance of closest approach between a track in a jet and the reconstructed primary vertex. The IP is calculated in three dimensions, taking advantage of the excellent z-axis pixel detector resolution in the CMS detector (see Section 3.2). The sign of the IP is determined by the sign of the scalar product of the IP and the reconstructed jet direction; with tracks originating from the decay of particles travelling along the jet axis tending to have positive IP values.

This parameter is used to identify tracks that are inconsistent with the jet direction, such as prompt tracks (particle tracks that originate from the primary vertex) and tracks typically found in the jets from light-flavour decays. Since the jet direction from a b-quark decay approximates the flight path of the B hadron, the tracks from a b-quark decay are expected to have positive IP values. The IP distribution is shown in data compared to simulated QCD multijet samples on the left in Fig. 4.2.

The discrimination between the decay products of b- and non-b-jets is made using the IP significance, defined as IP/σ_{IP}, where σ_{IP} is the related uncertainty. The distribution of the IP significance can be seen on the right in Fig. 4.2, where the b-quark contribution can be seen mostly at positive values where it dominates. The *Track Counting* (TC) algorithm sorts the tracks

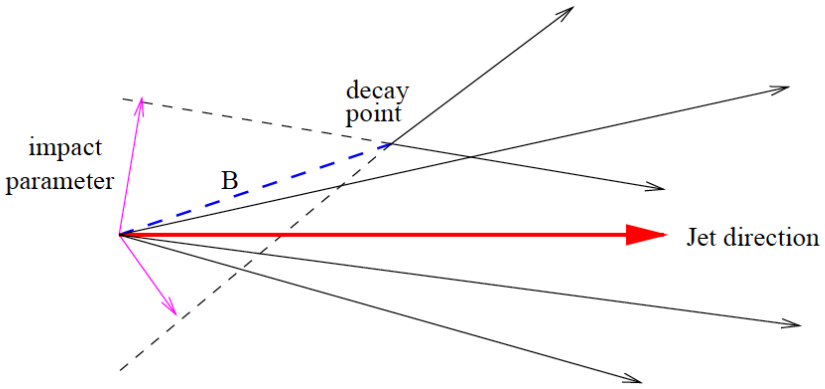


Figure 4.1: An event with a secondary vertex from a B hadron decay and it's impact parameter [5].

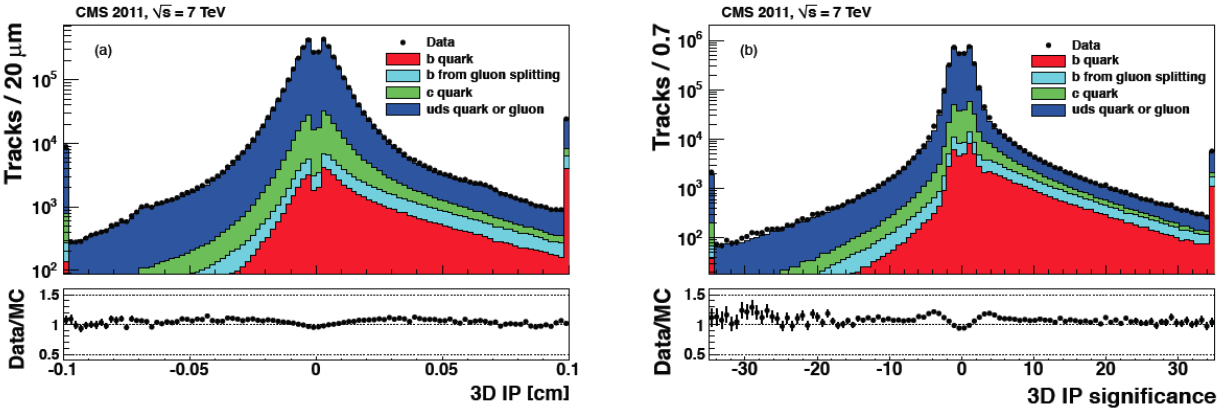


Figure 4.2: Distributions of the 3D impact parameter (left) and the significance of the 3D impact parameter (right) for all selected tracks. In each histogram, the rightmost bin includes all events from the overflow. Contributions from different components are taken from multijet QCD simulation and are renormalised to match the yields in data. The sample corresponds to a trigger selection with jet $p_T > 60$ GeV/c. [88].

found in a jet in order of decreasing IP significance. The High Efficiency version (TCHE) discriminates using the IP significance of the second ranked track, whilst the High Purity version (TCHP) uses the third ranked track. By requiring more tracks above the discriminant threshold, the TCHP algorithm rejects relatively more light-flavour decays and thus shows improved purity.

The concept of the TC algorithm has been extended to two, more complex algorithms that consider the structure of the jet, not just the component tracks. The Jet Probability (JP) algorithm implements a likelihood estimate that the jet-associated tracks originate from the primary vertex. The Jet B Probability (JBP) algorithm gives weights to the tracks with the highest IP significances, up to a maximum of four - the average number of reconstructed charged particles from b-hadron decays. The distribution of the JP discriminator is shown in Fig. 4.3, where it can be seen that with the tight working point discriminator cut of 0.790 the remaining events are almost exclusively b-flavour events.

The *Simple Secondary Vertex* (SSV) algorithm uses a function of the signed, three dimensional flight distance significance of secondary vertices as a discriminant. The discriminant is calculated using equation 4.4, where D_{3D} is the signed three-dimensional flight distance between the secondary and primary vertices, and $\sigma_{D_{3D}}$ is the uncertainty on this parameter. In the cases with more than one reconstructed secondary vertex the one with the highest significance is used. The SSV flight distance significance distribution in data compared to simulated QCD multijet samples is shown on the left in Fig. 4.4. Contributions from different components are taken from multijet QCD simulation and are renormalised to match the yields in data. All events in data and simulation have passed a jet $\mathrm{p_T} > 60$ GeV/c trigger requirement. The relative size of the b-quark sample, compared to the other components, demonstrates the discriminating power of the variable.

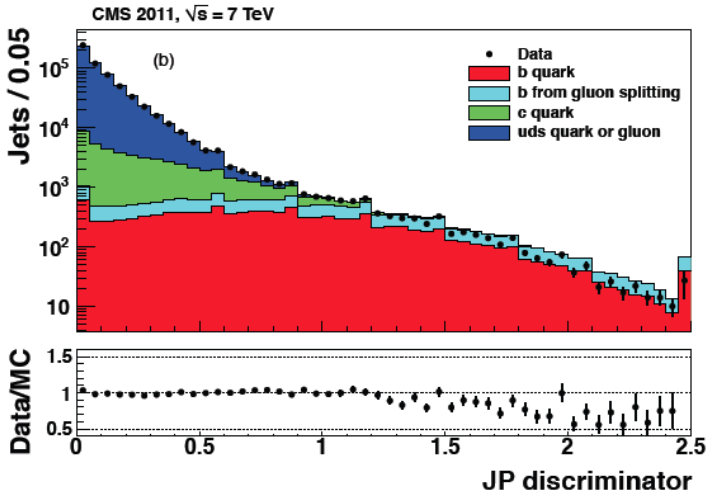


Figure 4.3: Discriminator values for the JP algorithm. The small discontinuities in the distribution are due to the $> 0.5\%$ single track probability requirement. Contributions from different components are taken from multijet QCD simulation and are renormalised to match the yields in data. [88].

$$SSV_{Disc} = log(1 + \frac{D_{3D}}{\sigma_{D_{3D}}}) \tag{4.4}$$

The SSV algorithm distinguishes between b- and non-b-jets by reconstructing a secondary decay vertex and placing requirements on the associated variables, designed to increase the b-identification purity. The SSV algorithm requires that:

- secondary vertices that share at least 65% of their tracks with the primary vertex are discarded;
- the significance of the radial distance between the primary and secondary vertices must be greater than 3σ;
- SV candidates with a radial distance of > 2.5 cm with respect to the primary vertex or with masses exceeding 6.5 GeV/c^2 or comparable to

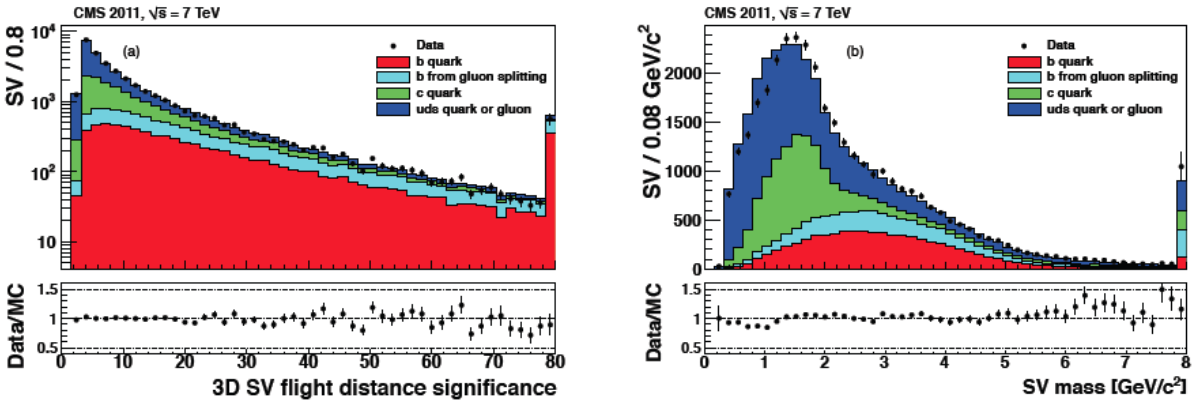


Figure 4.4: Properties of reconstructed decay vertices used in the secondary vertex (SV) algorithm: significance of the 3D flight distance (left) and associated mass (right). Contributions from different components are taken from multijet QCD simulation and are renormalised to match the yields in data [88].

that of the K^0 meson are rejected (SV mass distribution shown on the right in Fig. 4.4). This is to reduce contamination due to the decay of massive long-lived mesons and of vertices from particles interacting with the detector material;

- the flight direction of the SV candidates are required to lie within a cone of $\Delta R = 0.5$ around the jet direction.

There are two versions of the SSV algorithm optimised for different purities; the *High Efficiency* (SSVHE) version requires at least two tracks associated to the secondary vertex, whilst the *High Purity* (SSVHP) version requires at least three. The distributions of these two discriminants are shown in Fig. 4.5.

There exists a more complex approach which uses a combination of secondary vertices and track-based lifetime information of the jet particles. This *Combined Secondary Vertex* (CSV) algorithm exhibits greater efficiency with respect to the standard SSV algorithms as it becomes possible to discriminate between jets in the cases when no secondary vertices are found. A selection of

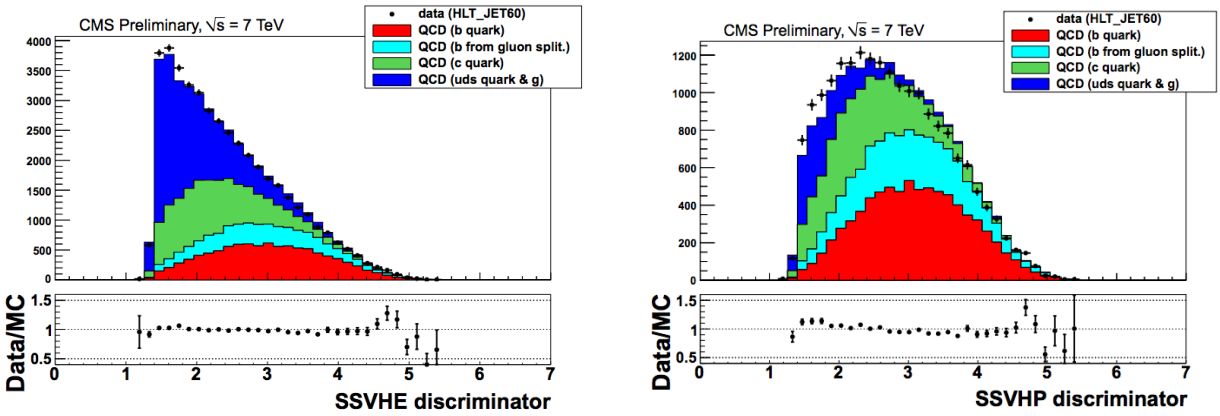


Figure 4.5: Distributions of the High Efficiency (left) and High Purity (right) Simple Secondary Vertex discriminators. Contributions are taken from multijet QCD simulation and renormalised to match the yields in data [89].

variables, such as reconstructed vertex mass and the number of tracks in the jet, are used to construct two likelihood ratios, for discriminating between b- and c-jets and between b- and light-flavour jets. The likelihood ratios are combined into a single CSV discriminator, the distribution of which is shown in Fig. 4.6.

The performance of all the algorithms described above can be seen in Fig. 4.7 where the misidentification probabilities (the efficiencies to incorrectly tag non-b-jets as b-jets) in simulation are shown as a function of the b-tagging efficiencies. For the loose working points (10% misidentification probability) a b-tagging efficiency of $70 - 85\%$ is achieved for light-flavour jets. In this region the JBP algorithm has the highest b-tagging efficiency. For tight selections (misidentification probability of 0.1%) the b-tagging efficiency values are $45 - 55\%$; as expected, the efficiency drops much lower for the high-efficiency algorithms at this high-purity working point, to approximately 15% . The non-negligible lifetime of c-hadrons makes it difficult to distinguish c-jets from b-jets, shown on the right in Fig. 4.7. As such, the c-jet misidentification probability is much higher for all b-tagging algorithms

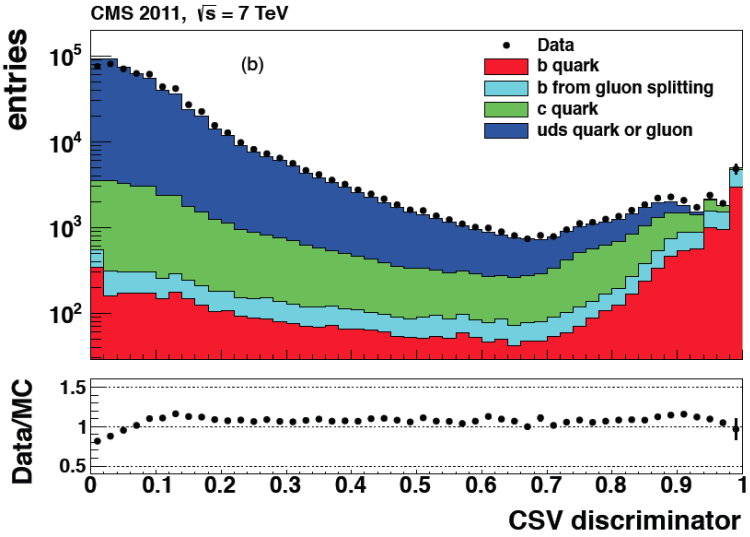


Figure 4.6: Distribution of the CSV discriminator, with contributions taken from multijet QCD simulation and renormalised to match the yields in data [88].

than for light flavour jets. The CSV algorithm, explicitly tuned for light-flavour and c-jet rejection, shows the best performance for the medium and tight working points.

The b-tagging algorithm used for this analysis is the Simple Secondary Vertex High Efficiency (SSVHE) algorithm, whose distribution is shown on the left of Fig. 4.5. It is applied at the medium operating point, corresponding to a discriminant value of > 1.74. The sample efficiency is favoured over purity, in light of the working luminosity of the LHC (corresponding to 4.9 fb^{-1} for the 2011 dataset) and the relatively low associated tW-channel production cross section (refer to Section 2.2). The medium working point corresponds to a reasonable trade-off in b-jet identification efficiency and purity. This algorithm was recommended by the CMS Top PAG for historical reasons. When early $t\bar{t}$ production cross section analyses were conducted the algorithm choice was between the TCHE and SSVHE algorithms, which performed similarly well at the medium working point. However, simulation

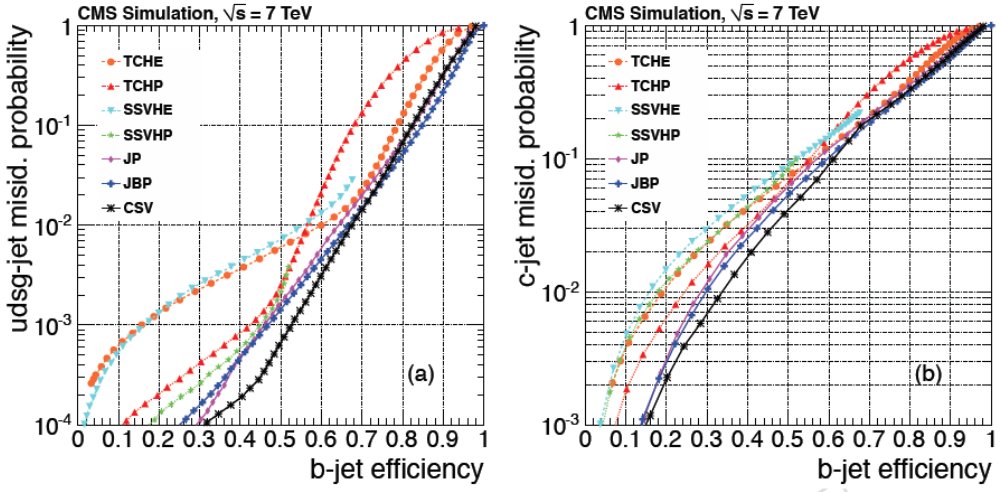


Figure 4.7: Performance curves obtained from simulation for the b-tagging algorithms described in the text. Light-flavour jet (left) and c-jet efficiencies (right) as a function of the b-jet efficiency [88].

studies conducted of possible tracker misalignment and pixel detector failure scenarios in early data suggested that the TC algorithms were potentially less robust, so were not selected for use. The details of the study can be found elsewhere [90].

Chapter 5

Event selection

This analysis imposes a set of kinematic requirements on data to extract a sub-sample enriched with tW signal events. The resulting event sample is then compared with the simulation predictions for the different background processes in a counting experiment assuming Poisson statistics. This is commonly referred to as a "cut and count" or "cut-based" method.

The prominent Standard Model backgrounds to the presented analysis are $t\bar{t}$ production with leptonically-decaying $W^{\pm}$ bosons and leptonically-decaying Z^0 bosons with additional jets (Z+jets). The other notable background sources are events with leptonically-decaying $W^{\pm}$ bosons with additional jets (W+jets), QCD multijet production (QCD) and single top production in the s- and t-channels. There are also additional small contributions from electroweak diboson production and photon production with additional jets; the latter is included with the QCD simulation sample and is not addressed separately.

The Top PAG within CMS provides a common event selection as a reference point for all of the top quark analyses performed by CMS; which is taken as the initial event selection for this analysis. The aim is to select the characteristics of a single top decay in the dilepton channels whilst reducing the contributions from the various background processes. The final require-

ment, on the H_T variable, is specific to this analysis. These requirements are described below in the order in which they are applied to the sample sets. The general principle of the order is to first check that the events are of good quality and then to reduce the more abundant backgrounds (in this case, Z+jets), thus increasing the sample purity.

The kinematic distributions of the main variables of the analysis are presented for data and simulation. The simulation is normalised using the theoretical cross-sections (see Section 4.1) to the integrated luminosity of the data (4.9 fb^{-1}). The Z+jets background in the distributions has been re-weighted to correct for high pile-up effects, unless otherwise stated; the derivation of these scale factors is explained in Section 5.3.

5.1 Trigger and event cleaning

Dilepton triggers are used to select data events containing a minimum of two leptons; a description of the trigger system is given in Chapter 3. This produces three distinct samples corresponding to each possible final state: electron-electron (ee), electron-muon ($e\mu$) and muon-muon ($\mu\mu$). Events that are accepted by the HLT are processed using the Particle Flow reconstruction scheme that is described in Chapter 4.

Tables 5.1, 5.2 and 5.3 list the HLT triggers used by this analysis over the course of the 2011 data-taking period, including their L1 trigger seeds and the run numbers they were active in.

In the trigger path names, *Mu, Ele* and *EG* denote muons, electrons and electromagnetic particles (electrons or photons), respectively, with the number suffix representing the energy threshold in GeV for that particle. The version number at the end of the HLT trigger name corresponds to the CMS trigger menu it was included in; it does not denote any (significant) changes to the actual HLT trigger.

As the LHC operating conditions changed over 2011 the rates of the

Table 5.1: HLT trigger paths for the $\mu\mu$ final state.

Trigger path	L1 seed	Up to Run
HLT_DoubleMu7_v1	L1_DoubleMu3	163261
HLT_Mu13_Mu8_v2	L1_DoubleMu3	167043
HLT_Mu13_Mu8_v4	L1_DoubleMu3	170053
HLT_Mu13_Mu8_v6	L1_DoubleMu3	173198
HLT_Mu13_Mu8_v7	L1_DoubleMu3p5	178380
HLT_Mu17_Mu8_v10	L1_DoubleMu3p5	180252

triggers became excessively high, necessitating them to be lowered. For the majority of 2011, the trigger rates were reduced by increasing the energy thresholds of the HLT triggers and their L1 seeds. However, the online energy requirements began to encroach on those used in offline analyses, so more complex triggers had to be created.

To this end, calorimeter- and tracker-based isolation (see Section 5.2) and identification requirements were introduced for the electrons in the ee and $e\mu$ final state triggers; the latter only using *Calo* requirements. These correspond to basic kinematic requirements on the ECAL cluster shape, total p_T (energy) of tracks (ECAL hits) neighbouring the electron and the angular matching between the ECAL supercluster and electron track. In the trigger path names, the *Id* and *Iso* letter-suffixes denote Tight (T), Loose (L) and Very Loose (VL) working points, representing the severity of these kinematic requirements.

Before use, data samples are required to be 'cleaned' of anomalous readings caused by known detector effects. Anomalous noise, independent of electronics (or pedestal) noise, is caused by instrumentation issues associated with the HCAL hybrid photodiodes (HPDs) and HF Readout BoXes (RBXs) - see section 3.4.1. This noise proves to be problematic as it can extend up to TeV energies with a rate of the order of several Hz. The HBHE (HCAL Barrel, HCAL Endcap) Noise Filter is a collection of algorithms designed to remove this anomalous noise by placing constraints on the quantities measured using these HCAL sub-systems [91]. A large amount of energy and/or

Table 5.2: HLT trigger paths for the $e\mu$ final state.

Trigger path	L1 seed	Up to Run
HLT_Mu17_Ele8_CaloIdL_v1	L1_Mu3_EG5	
HLT_Mu8_Ele17_CaloIdL_v1	L1_Mu3_EG5	161176
HLT_Mu17_Ele8_CaloIdL_v2	L1_Mu3_EG5	
HLT_Mu8_Ele17_CaloIdL_v2	L1_Mu3_EG5	163261
HLT_Mu17_Ele8_CaloIdL_v3	L1_Mu3_EG5	
HLT_Mu8_Ele17_CaloIdL_v3	L1_Mu3_EG5	164237
HLT_Mu17_Ele8_CaloIdL_v4	L1_Mu3_EG5	
HLT_Mu8_Ele17_CaloIdL_v4	L1_Mu3_EG5	165888
HLT_Mu17_Ele8_CaloIdL_v5	L1_MuOpen_EG5	
HLT_Mu8_Ele17_CaloIdL_v5	L1_MuOpen_EG5	166967
HLT_Mu17_Ele8_CaloIdL_v6	L1_Mu3_EG5	
HLT_Mu8_Ele17_CaloIdL_v6	L1_MuOpen_EG12	170053
HLT_Mu17_Ele8_CaloIdL_v8	L1_Mu7_EG5	
HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_v3	L1_MuOpen_EG12	173198
HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_v4	L1_Mu7_EG5	
HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_v4	L1_MuOpen_EG12	178380
HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_v7	L1_Mu12_EG5	
HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_v7	L1_MuOpen_EG12	180252

Table 5.3: HLT trigger paths for the ee final state.

Trigger path	L1 seed	Up to Run
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v1	L1_SingleEG12	161176
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v2	L1_SingleEG12	163261
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v3	L1_SingleEG12	164237
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v4	L1_SingleEG12	165888
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v5	L1_SingleEG12	167043
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v6	L1_SingleEG12	170053
HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v6	L1_SingleEG12	170759
HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v7	L1_DoubleEG_12_5	173198
HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v8	L1_DoubleEG_12_5	178380
HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v9	L1_DoubleEG_12_5	180252

hits observed in a specific detector region with nothing else nearby is classified as *isolated noise*. If an event has $E > 50$ GeV or $E_T > 25$ GeV identified as isolated noise the event is rejected. There are also upper limits placed on the number of hits seen in an HPD and on the number of channels identified as isolated noise for an event to be accepted as signal.

As proton bunches circle round the LHC beam pipe they sometimes collide with the beam collimators or residual gas particles, producing showers of secondary particles that can be picked up by the detector. These are known as 'beam scraping events' and are removed from the data samples by requiring that at least 25% of the good tracks measured by the CMS tracker system are of highPurity [92], see Section 4.2.1.

5.2 Lepton selection and veto

Exactly two isolated, oppositely-charged leptons originating from the same reconstructed primary vertex have to be present in the event for it to be accepted as a possible signal event.

Electron candidates are PF electrons that pass the following criteria:

- transverse energy $E_T > 20$ GeV;
- pseudorapidity $|\eta| < 2.5$;
- consistent with the primary vertex with $|\Delta z| < 1.0$ cm;
- transverse impact parameter between the electron track and the beam spot (the luminous region at the proton beam crossing point within CMS) < 0.02 cm;

Electron candidates undergo an additional set of identification requirements, known as 'Cuts in Categories' (CiC), optimised to select electrons

from Z^0 or $W^{\pm}$ decays and to reject fake electrons from jets or photon conversions [93]. CiC imposes requirements on the relative position and matching between the track, ECAL supercluster and HCAL energy deposits that make up the electron candidate, as well as their individual isolation (see equation 5.1 below); as well as constraints on the track vertex and ECAL supercluster shape. The name comes from the identification being divided into different categories of electrons, based on how sensitive they are to Bremsstrahlung energy loss from their trajectory, energy, etc.

The categories are defined using the fraction of radiated energy as measured from the innermost and outermost parts of the electron track and the ratio E/p between the ECAL supercluster energy and the measured track momentum at the vertex. These electron categories are further sub-divided in the ECAL barrel and endcaps, as well as into several regions of E_T, all with their own selection requirement thresholds optimised to give the best signal to background ratio for single electrons. In this analysis the CiC identification is applied at the 'supertight' working point, which corresponds to an electron selection efficiency of 89.3% in the barrel and 85.5% in the Endcaps [93].

Electrons originating from photon conversions represent a significant source of fake isolated electrons; As such, two techniques are used to distinguish them from prompt electrons originating from hadron collisions. The first identification technique uses missing hits from the electron track in the inner part of the tracker. If there are any layers of the tracker missing a hit from the electron, it is identified as a conversion electron and is vetoed. The second technique attempts to associate a partner track to the electron track which would be consistent with both tracks having originated from a photon conversion. The event is vetoed if a partner track is identified with an absolute distance (in the $r-\phi$ plane) between the tracks of less than 0.02 cm and with a difference in the cotangent of the polar angle of each track of less than 0.02.

Muon candidates are PF muons that pass both the Global and Tracker muon reconstruction (a muon spectrometer track extrapolated to a matching tracker system track, and vice versa - see Chapter 4). They must then also

satisfy the following criteria:

- transverse momentum $p_T > 20$ GeV/c;
- pseudorapidity $|\eta| < 2.4$;
- > 10 hits in the muon tracker system;
- transverse impact parameter of the muon with respect to the beam spot < 0.02 cm;
- normalised χ^2 (of the muon track fit) < 10;

The relative isolation of muons and electrons are quantified using Particle Flow Isolation, evaluated as the charged and neutral energy deposits of the hadrons and photons summed in a cone of $\Delta R < 0.3$ around the lepton, divided by the lepton p_T:

$$I_{PF} = \frac{chargedHadronIso + neutralHadronIso + photonIso}{P_T^l} \tag{5.1}$$

if the additional energy within the cone around the lepton is small, it is considered isolated. For both muons and electrons, the PF isolation is required to be < 0.15.

If any additional leptons are observed in an event it is vetoed and not considered further for the analysis. Any additional leptons are identified with looser requirements compared to the signal leptons. Events with extra leptons that satisfy $p_T^{electron} > 15$ GeV/c ($p_T^{muon} > 10$ GeV/c), PF isolation < 0.2 and a minimum $|\eta|$ of 2.5 are vetoed.

For loose electron candidates a simpler cut-based quality criteria is applied than for signal tight electrons. This identification scheme places four kinematic requirements involving the energy ratio with, and relative position to, the HCAL energy deposits and the ECAL cluster shapes. The simplicity is that there are only two categories here, barrel and endcap, each with different requirement thresholds. It is implemented at a relaxed working point corresponding to a 92.5% efficiency for electrons in the barrel and 86.4% in

the endcaps [93].

5.3 Invariant mass and Z^0 boson veto

Events where the invariant mass of the lepton pair $m_{ll} < 20$ GeV/c^2 are rejected to remove low mass Z+jets (Drell-Yan) and QCD events. In the same flavour final states (ee and $\mu\mu$), an additional veto is applied for events with m_{ll} between 81 and 101 GeV/c^2, corresponding to the mass of the Z^0 boson. This cut is intended to reduce the background from Z^0 boson decays, primarily Z+jets, but also ZZ and WZ diboson events, whose invariant mass spectra also peak in this region, as illustrated in Fig.5.1.

Under high pile-up conditions, like those present in the 2011 data taking period, the E_T^{miss} resolution becomes degraded, causing disagreement between the simulation and data; this disagreement can be clearly seen in Fig. 5.2. The cause of this degradation is particles from separate pile-up collisions being mis-assigned to the collision being analysed, imbalancing the transverse-momentum-vector sum used to calculate E_T^{miss}. The events removed by the Z^0 boson veto are used as a Z+jets enriched control sample to study this effect. The simulated E_T^{miss} distributions are then corrected by scaling to match the E_T^{miss} distribution in data observed in this enriched control sample. The relative differences between the event counts before and after scaling is assigned as the uncertainty for this process. In the case of the $e\mu$ final state, where the Z^0 boson invariant mass cut is not required, the average of the two derived scale factors is applied. The same-flavour final state scale factors can be found in Table 5.4, along with statistical uncertainties for the calculated ee and $\mu\mu$ final states.

Another selection requirement specific to the ee and $\mu\mu$ final states is made on the missing transverse energy of the event, which is required to be $>$ 30 GeV. The intent of this requirement is to exclude Z+jets and QCD events with E_T 'lost' or mis-identified by the detector and mistaken for legitimate

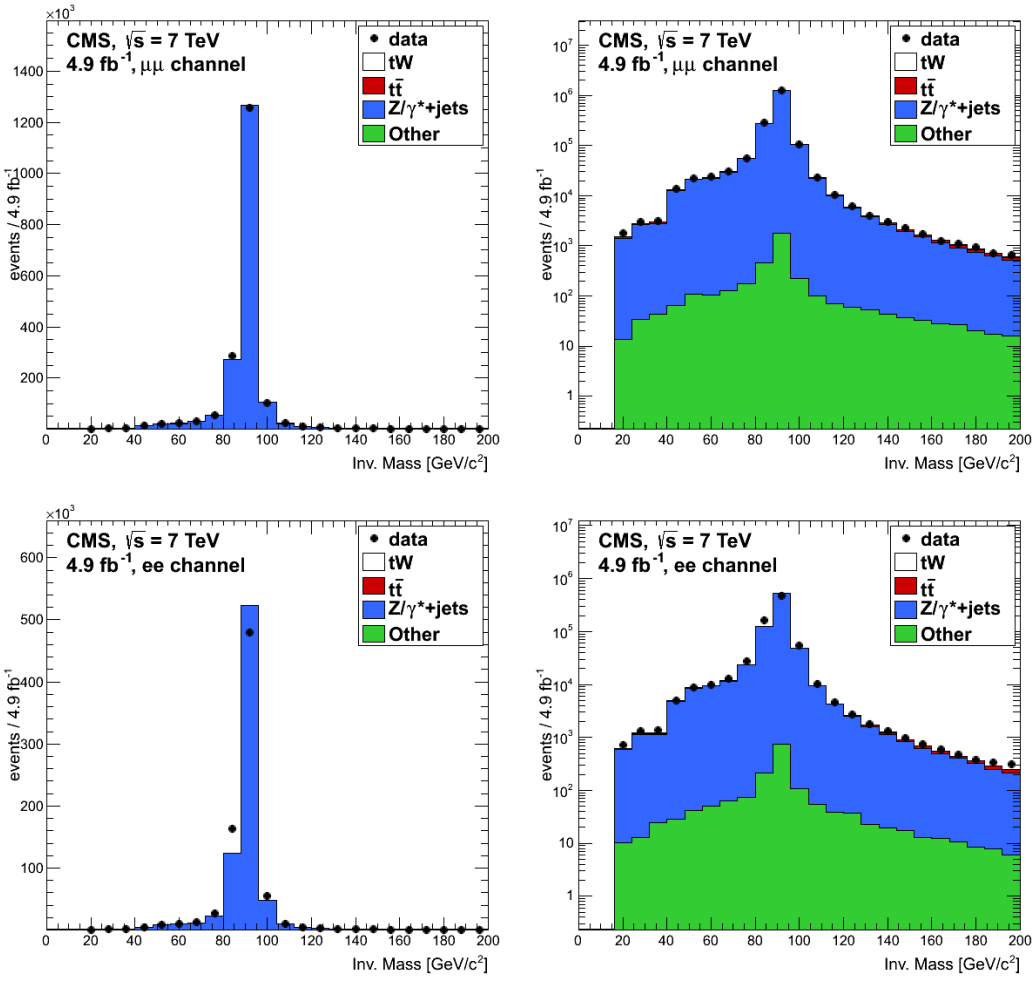


Figure 5.1: Invariant mass of the lepton pair together for data and simulation in the $\mu\mu$ (top row) and ee (bottom row) decay channels, after lepton selection, lepton veto and $m_{ll} > 20$ GeV/c^2.

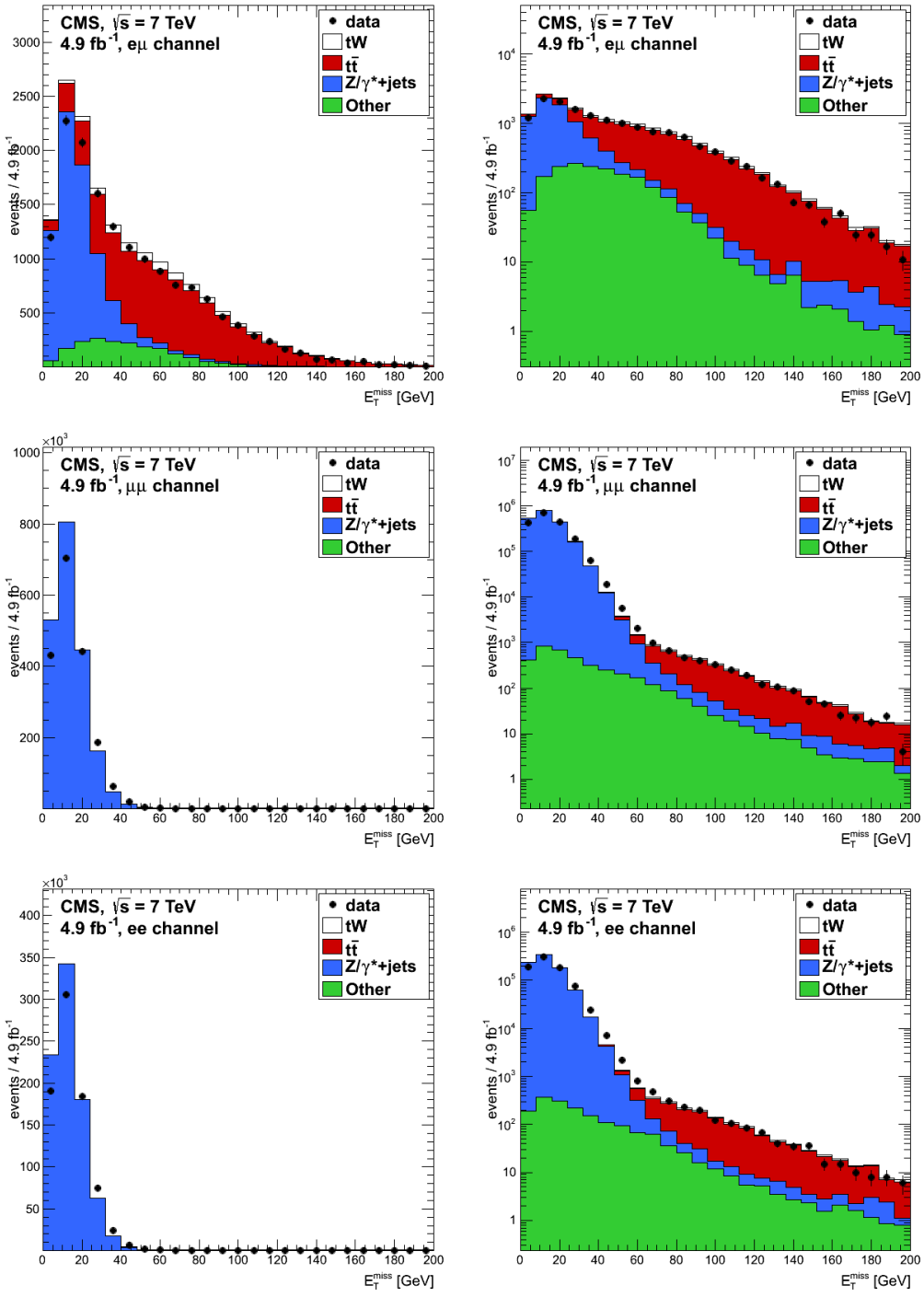


Figure 5.2: Unscaled missing transverse energy distribution together for data and simulation in the $e\mu$, $\mu\mu$ and ee decay channels, after lepton selection, lepton veto and $m_{ll} > 20$ GeV/c^2.

Table 5.4: Simulation scale factors as a function of the E_T^{miss}, with statistical uncertainties for the calculated same-flavour final states.

E_T^{miss}	Scale factor $\mu\mu$	Scale factor ee	Scale factor $e\mu$
< 5 GeV	0.8001±0.0007	0.792±0.001	0.80
< 10 GeV	0.8224±0.0006	0.823±0.001	0.82
< 15 GeV	0.8798±0.0007	0.893±0.001	0.88
< 20 GeV	0.948±0.001	0.966±0.002	0.96
< 25 GeV	1.039±0.002	1.073±0.003	1.06
< 30 GeV	1.144±0.003	1.176±0.005	1.16
< 35 GeV	1.259±0.005	1.287±0.008	1.27
< 40 GeV	1.400±0.009	1.43±0.01	1.41
< 45 GeV	1.50±0.01	1.51±0.03	1.50
< 50 GeV	1.61±0.03	1.80±0.05	1.70
< 60 GeV	1.79±0.04	1.86±0.08	1.82
> 60 GeV	1.41±0.09	1.68±0.13	1.54

E_T^{miss} from neutrinos. The E_T^{miss} distributions, scaled to correct for pile-up effects, are shown in Fig. 5.3 for the ee and $\mu\mu$ final states.

Although the majority of the Z+jets background is removed by the m_{ll} requirements the large cross section for this process means that the contribution is still relatively large, prompting this additional constraint.

However, to improve the signal to background performance of the E_T^{miss} constraints in events with pile-up, a new quantity called *Tracker* E_T^{miss} was developed by the CMS $H \to WW$ sub-group [94]. The distribution of this variable, which is calculated using only the charged particle candidates associated to the primary vertex, is shown in Fig. 5.4. As can be seen from the distribution plots, the data and MC are in good agreement and, as such, the variable requires no correction scaling. Comparing the distributions of the two E_T^{miss} quantities shows that the tracker E_T^{miss} has a more significant tail in Z+jets events. However, the two quantities are weakly correlated for processes with no genuine E_T^{miss}, such as the Z+jets background, and strongly correlated for processes with genuine E_T^{miss}, like the tW signal. Therefore, the E_T^{miss} requirement imposed in this analysis is made on the smaller of the two E_T^{miss} quantities, $min(E_T^{miss}, TrackerE_T^{miss}) > 30$ GeV (distribution shown in

Fig. 5.5), in the ee and $\mu\mu$ final states only.
By looking at the various E_T^{miss} distribution plots, it can be seen that the imposed 30 GeV requirement could be optimised to remove more background events; specifically, increasing the requirement to 40-50 GeV would remove more Z/γ^* events, a large background to this analysis. Unfortunately, due to time constraints for publication, no optimisation could be performed and the recommended 30 GeV requirement was kept. Instead, the focus, through the E_T^{miss} rescaling and introduction of tracker E_T^{miss}, was to produce a reliable E_T^{miss} variable to place requirements on.

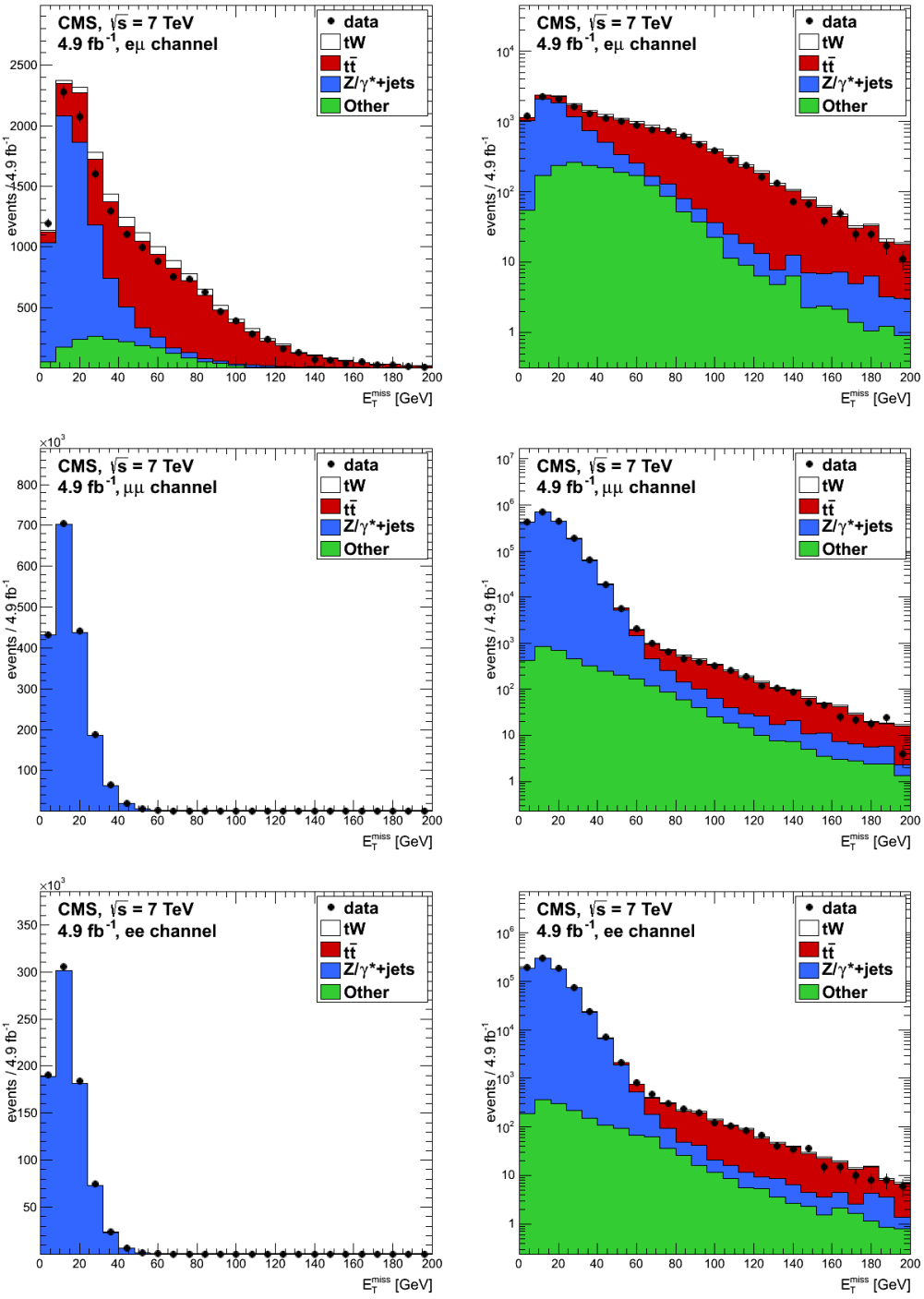


Figure 5.3: Scaled missing transverse energy together for data and simulation in the $e\mu$ (top row), $\mu\mu$ (middle row) and ee (bottom row) decay channels, after lepton selection, lepton veto and $m_{ll} > 20$ GeV.

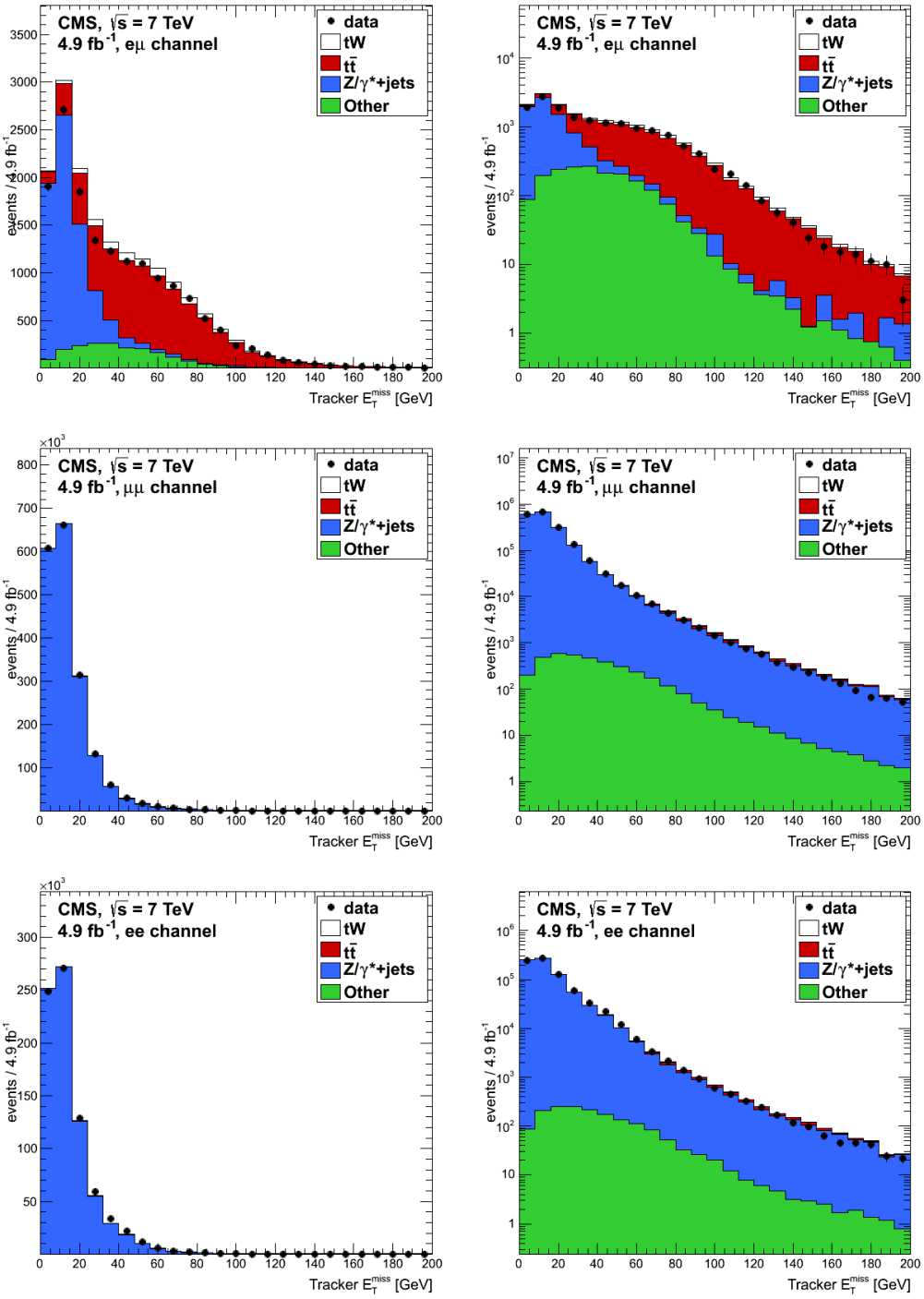


Figure 5.4: Tracker missing transverse energy together for data and simulation in the $e\mu$ (top row), $\mu\mu$ (middle row) and ee (bottom row) decay channels, after lepton selection, lepton veto and $m_{ll} > 20$ GeV.

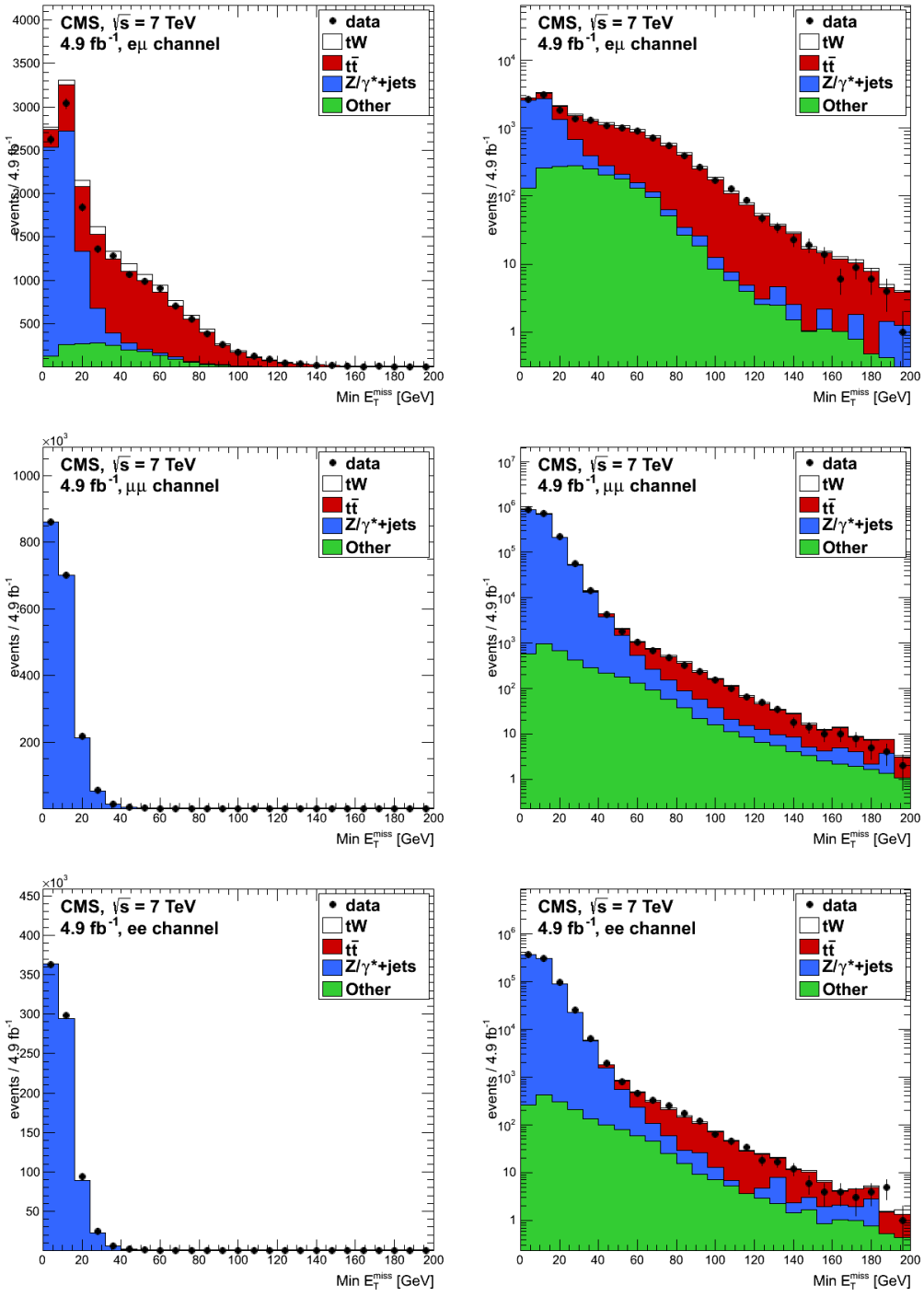


Figure 5.5: Minimum of the E_T^{miss} and tracker E_T^{miss} together for data and simulation in the $e\mu$ (top row), $\mu\mu$ (middle row) and ee (bottom row) decay channels, after lepton selection, lepton veto and $m_{ll} > 20$ GeV.

5.4 Jet selection & b-jet requirements

The identified jets in each event are ordered in terms of their momentum, with the highest p_T jet referred to as the 'leading jet'. To reduce $t\bar{t}$ contamination in the signal sample, events are required to have exactly one jet, by default the leading jet, with $|\eta| < 2.4$ and $p_T > 30$ GeV/c. The p_T distribution of the leading jet is shown in Fig. 5.6 for the $e\mu$, $\mu\mu$ and ee final states.

In these distributions there are no jets with p_T ¡ 25 GeV. This is due to the 'skimming' process, where loose lepton and jet cuts were applied to the data and MC samples to reduce their size, and thus the analysis run-time.

The jet must also have an angular separation, between it and the nearest lepton of $\Delta R > 0.3$. If ΔR drops below this value the lepton is associated with the reconstructed jet and they are not considered independent. Similar to the case with electrons, CMS has developed jet quality criteria for PF jets. It requires that the energy of any given jet be distributed amongst the ECAL and HCAL detectors for charged and neutral particles, meaning fake jets from false energy deposits in a single sub-detector system can be identified and removed from the sample [82]:

- The PF jet must have a charged particle multiplicity > 0;
- The PF jet must have more than 1 constituent;
- The fraction of jet energy deposited in the ECAL must be < 0.99 for both charged and neutral electromagnetic particles;
- The fraction of jet energy deposited in the HCAL must be < 0.99 for neutral hadrons and > 0 for charged hadrons;

These requirements correspond to the loose jet ID, which is fully efficient ($>$ 99.9%) for real, high p_T jets [82].

It is then required that the single identified jet have originated from a b-quark decay. The Simple Secondary Vertex High Efficiency b-tagging algorithm, operating at the medium working point, is used to identify b-jets

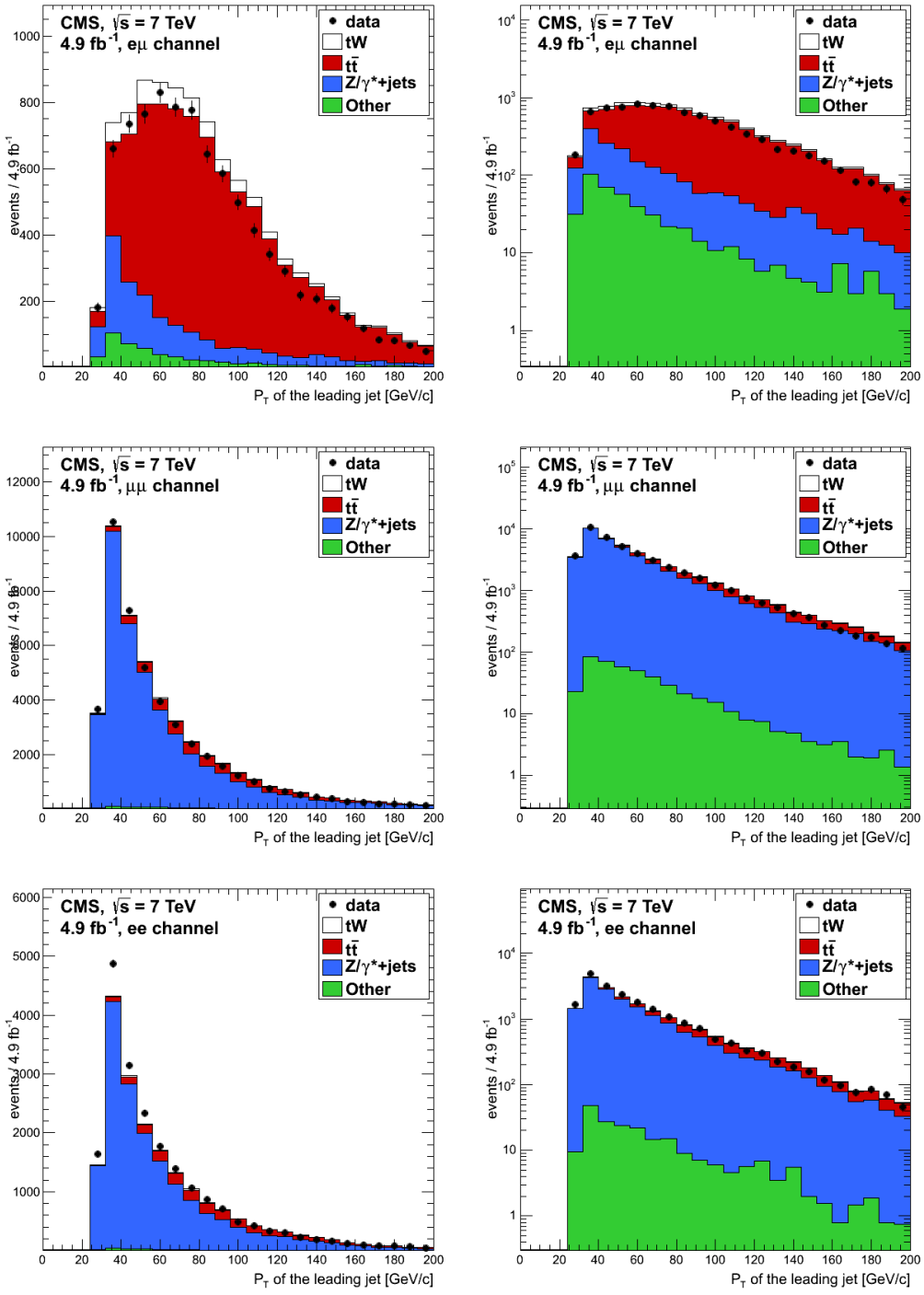


Figure 5.6: p_T of the leading jet in events with at least one jet for data and simulation in the $e\mu$ (top row), $\mu\mu$ (middle row) and ee (bottom row) decay channels, after the invariant mass requirements.

(see Section 4). Events with extra b-tagged jets are vetoed; for that, jets are defined as previously described but with a lower momentum threshold of $p_T > 20$ GeV/c.

5.5 H_T veto

For the same-flavour dilepton final states the majority of the Z+jets background has been removed by this stage in the selection; although events do persist in the sample due to the large cross section for this process. The Z+jets contamination in the $e\mu$ final state, however, is primarily due to $Z \to \tau\tau$ events, and therefore requires bespoke kinematic requirements to remove it. The kinematic variable H_T is defined as the scalar sum of the transverse momenta of the leptons, jet and E_T^{miss} in the event, and is required to be greater than 160 GeV for the event to be accepted as signal in the $e\mu$ final state only. This requirement takes advantage of the differing topology of Z+jets events compared to tW events, specifically the generally lower E_T^{miss}, which causes the H_T distribution to peak at lower values, as seen in Fig. 5.7.

A requirement is not placed directly on the E_T^{miss} distribution as this would reduce the signal selection efficiency too much, as can be seen in Fig. 5.3.

5.6 Background estimations

The estimated number of events in simulation for the signal and all background processes after each kinematic requirement are given, with statistical uncertainties, in Tables 5.5, 5.6 and 5.7 for the three leptonic final states. Table 5.8 shows the event yield in data, compared with the combined total for simulation, for each individual final state. Entries with zero events are denoted with "-".

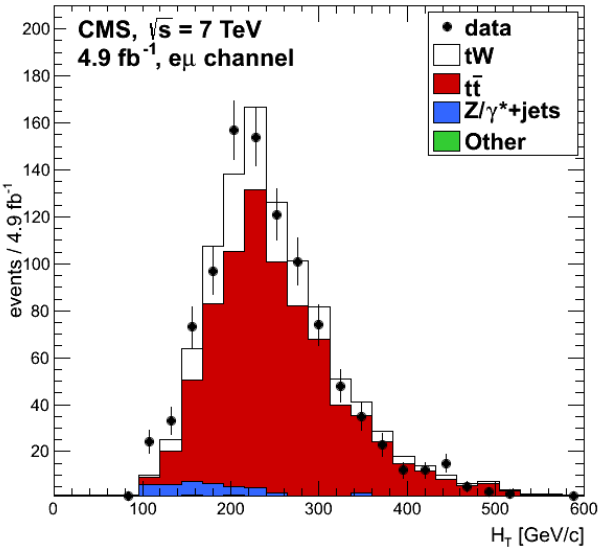


Figure 5.7: Distribution of the variable H_T in the $e\mu$ decay channel, after the b-tagging kinematic requirements.

Table 5.5: Event yields in simulation for the ee final state normalised to 4.89 fb^{-1}, with statistical uncertainties.

	tW	$t\bar{t}$	diboson	W+jets	Z+jets	single top	QCD
Lepton Sel.	251 ± 4	2576 ± 28	1586 ± 8	106 ± 18	775593 ± 659	2.2 ± 0.4	–
Inv. Mass	211 ± 4	2127 ± 25	571 ± 6	85 ± 16	95952 ± 262	1.7 ± 0.3	–
E_T^{miss}	153 ± 3	1549 ± 21	283 ± 4	35 ± 10	1569 ± 44	1.1 ± 0.3	–
1 Jet	77 ± 2	320 ± 10	63 ± 2	6 ± 4	569 ± 26	0.3 ± 0.2	–
b-tagging	42 ± 2	146 ± 7	1.7 ± 0.3	≤ 8	24 ± 5	0.06 ± 0.06	–

Table 5.6: Event yields in simulation for the $e\mu$ final state normalised to 4.90 fb^{-1}, with statistical uncertainties.

	tW	$t\bar{t}$	diboson	W+jets	Z+jets	single top	QCD
Lepton Sel.	754 ± 7	7699 ± 48	1699 ± 11	185 ± 24	6674 ± 64	3.8 ± 0.5	35 ± 27
Inv. Mass	754 ± 7	7699 ± 48	1699 ± 11	185 ± 24	6674 ± 64	3.8 ± 0.5	35 ± 27
1 Jet	379 ± 5	1581 ± 22	345 ± 5	37 ± 11	1131 ± 29	1.2 ± 0.3	≤ 54
b-tagging	203 ± 4	762 ± 15	9.5 ± 0.8	≤ 21	35 ± 6	0.4 ± 0.2	–
H_T	189 ± 4	718 ± 15	7.9 ± 0.7	–	20 ± 5	0.3 ± 0.1	–

Table 5.7: Event yields in simulation for the $\mu\mu$ final state normalised to 4.92 fb^{-1}, with statistical uncertainties.

	tW	$t\bar{t}$	diboson	W+jets	Z+jets	single top	QCD
Lepton Sel.	558 ± 6	5739 ± 41	3726 ± 12	21 ± 8	1836690 ± 1006	2.4 ± 0.4	22 ± 22
Inv. Mass	479 ± 6	4826 ± 38	1354 ± 9	21 ± 8	231343 ± 405	2.0 ± 0.4	22 ± 22
min(E_T^{miss})	347 ± 5	3468 ± 32	678 ± 7	13 ± 7	4551 ± 71	1.1 ± 0.3	≤ 45
1 Jet	173 ± 3	717 ± 15	144 ± 3	6 ± 5	1772 ± 46	0.3 ± 0.1	–
b-tagging	91 ± 2	355 ± 10	4.4 ± 0.5	3 ± 3	69 ± 10	0.04 ± 0.03	–

Table 5.8: Event yields in data and combined simulation for the three final states for an integrated luminosity of 4.9 fb^{-1}, with statistical uncertainties.

	ee channel (4.89 fb^{-1})		$e\mu$ channel (4.90 fb^{-1})		$\mu\mu$ channel (4.92 fb^{-1})	
	Data	Simulation	Data	Simulation	Data	Simulation
Lepton Sel.	790846	780114 ± 659	15581	17051 ± 89	1859028	1846760 ± 1008
Inv. Mass	108700	98947 ± 264	15581	17051 ± 89	243287	238047 ± 408
min(E_T^{miss})	3931	3590 ± 50	15581	17051 ± 89	8573	9059 ± 79
1 Jet	1183	1035 ± 29	3137	3476 ± 39	2644	2812 ± 48
b-tagging	235	214 ± 8	992	1010 ± 17	472	522 ± 15
H_T	235	214 ± 8	899	935 ± 16	472	522 ± 15

After all the kinematic requirements are placed on the data, two background processes still contribute non-negligible yields in all three decay channels: Z+jets and $t\bar{t}$. The Z+jets contribution is the smaller of the two, having been substantially reduced by the invariant mass, E_T^{miss} and H_T requirements. However, as can be seen in Tables 5.5 and 5.7, it is still relatively large compared to the signal yield in the same-flavour final states. This is due to a long E_T^{miss} tail for Z+jets processes in these final states, as illustrated in Fig. 5.5, which cannot be cut out without removing tW signal events. As such, the $E_T^{miss} > 30$ GeV requirement, whilst not sufficient to completely remove the Z+jets contribution, has reduced it to acceptable levels that are less than the signal process.

The main background source for this analysis is $t\bar{t}$ production for two key reasons. Firstly, with a cross-section of 163 ± 14 pb^{-1} [3] $t\bar{t}$ events are produced almost 10 times as often as the signal process. Secondly, and more importantly, is its similarity to the signal tW topology, resulting in many of the kinematic requirements favouring both $t\bar{t}$ events and the signal. These two factors result in $t\bar{t}$ production being a substantial background even after applying all kinematic requirements. This is illustrated in Tables 5.5, 5.6, and 5.7, where, although the $t\bar{t}$ simulation is largely reduced by the kinematic requirements, its final event yield is approximately three times that of the signal process.

Due to this, and to reduce the dependence on the simulation modelling of the $t\bar{t}$ process, two additional sample regions are defined. Two control regions enriched in $t\bar{t}$ events, and lacking in tW events, are produced in data and simulation by selecting events with 2 jets, with either one (2j1t) or both (2j2t) of the jets being tagged b-jets. All other kinematic requirements are unchanged. There are 16 times as many $t\bar{t}$ events than tW in the 2j1t region, and 33 times as many in the 2j2t region; for comparison, in the 1j1t region the tW event total is approximately a quarter of that from $t\bar{t}$. The distribution of the number of jets and the number of b-tagged jets per event, in each decay channel, are shown in Figs. 5.8 - 5.10.

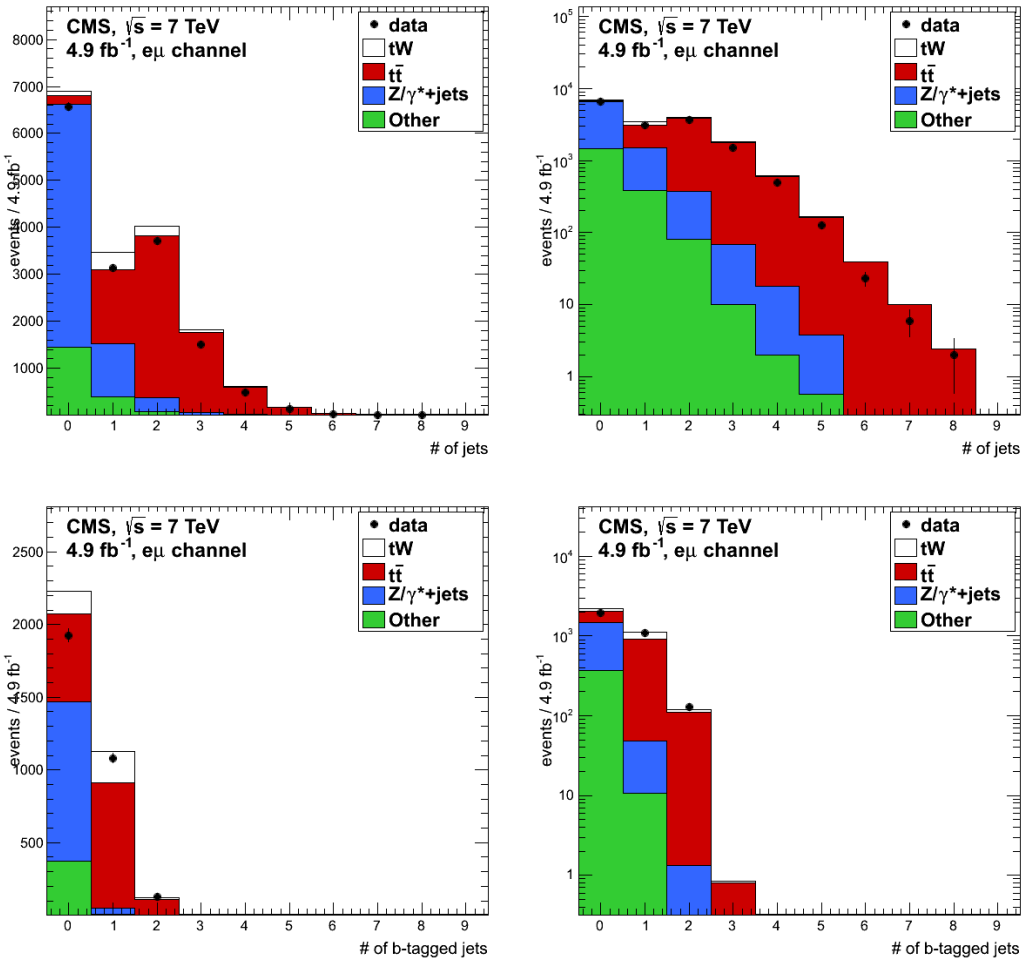


Figure 5.8: Number of jets (top row) and the number of b-tagged jets (bottom row) together in data and simulation in the $e\mu$ decay channel, after the invariant mass requirements.

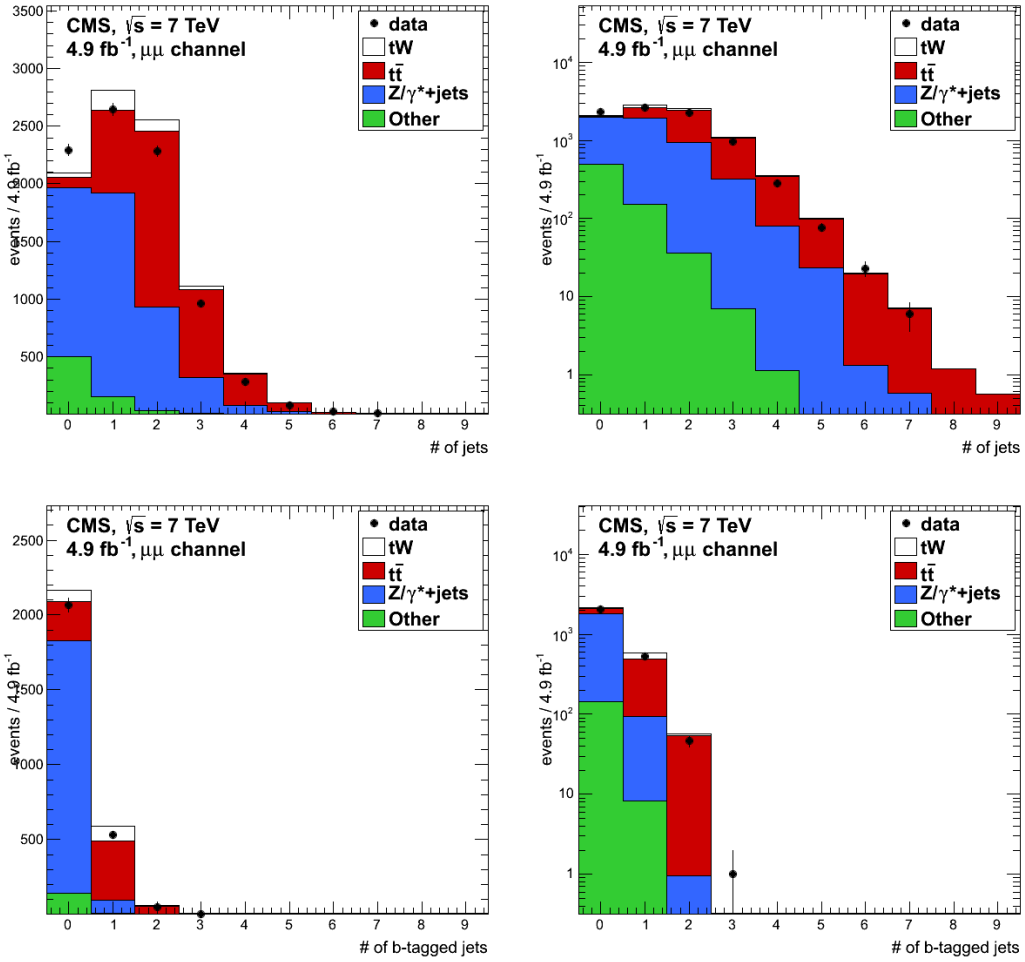


Figure 5.9: Number of jets (top row) and the number of b-tagged jets (bottom row) together in data and simulation in the $\mu\mu$ decay channel, after the invariant mass and E_T^{miss} requirements.

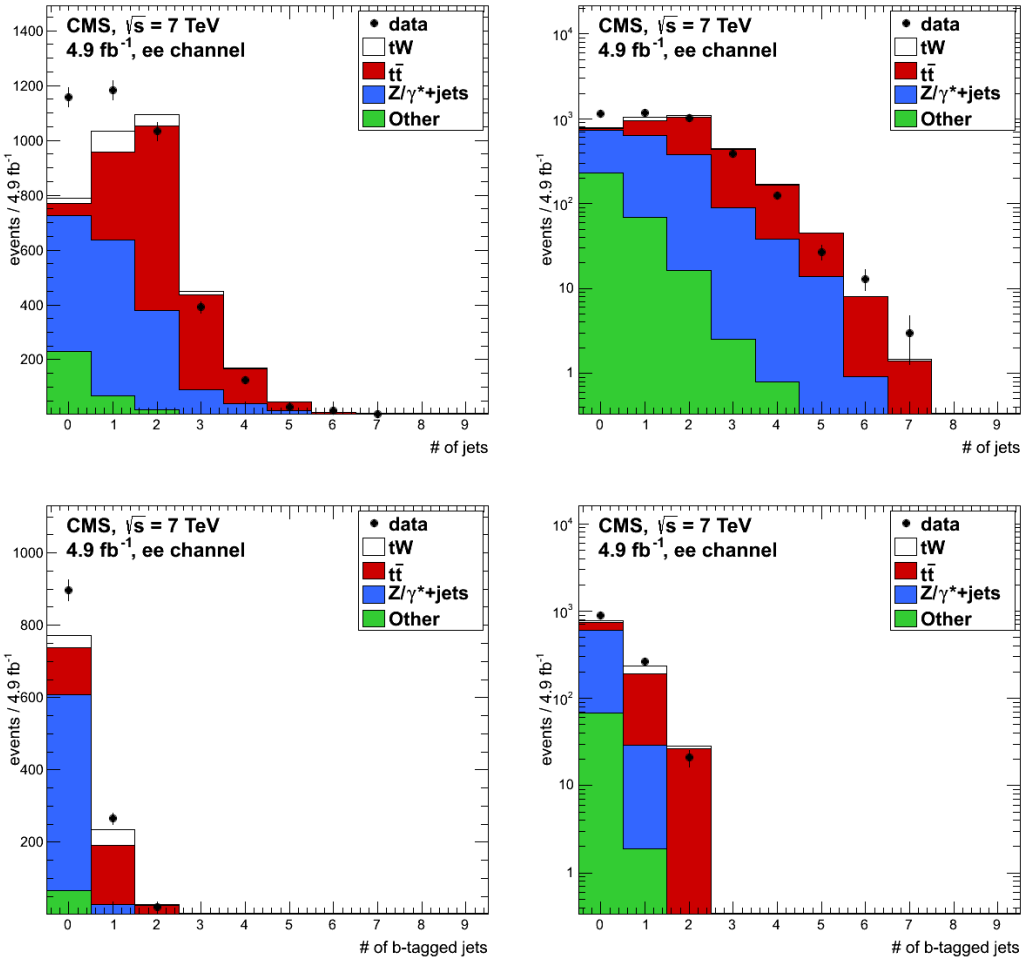


Figure 5.10: Number of jets (top row) and the number of b-tagged jets (bottom row) together in data and simulation in the *ee* decay channel, after invariant mass and E_T^{miss} requirements.

All three sample regions, the signal region (1j1t) and the two control regions (2j1t & 2j2t), are defined for each dilepton final state. The $t\bar{t}$ background is constrained by using these control regions simultaneously in the statistical fit to derive the cross section result, detailed in Chapter 7. The other backgrounds, including Z+jets, are estimated using simulation only. The event yields for data and simulation in the three sample regions are shown in Fig. 5.11 for each final state, clearly illustrating the abundance of $t\bar{t}$ events in all regions, and the persisting Z+jets background in the 1j1t signal region.

The contributions from all the other background processes, such as QCD, W+jets and diboson events, are expected to be, combined, no greater than 20% of the tW event count, from applying the kinematic requirements to simulated samples, as shown in Tables 5.5, 5.6, and 5.7. As such simulation is deemed sufficient to estimate their contribution in data.

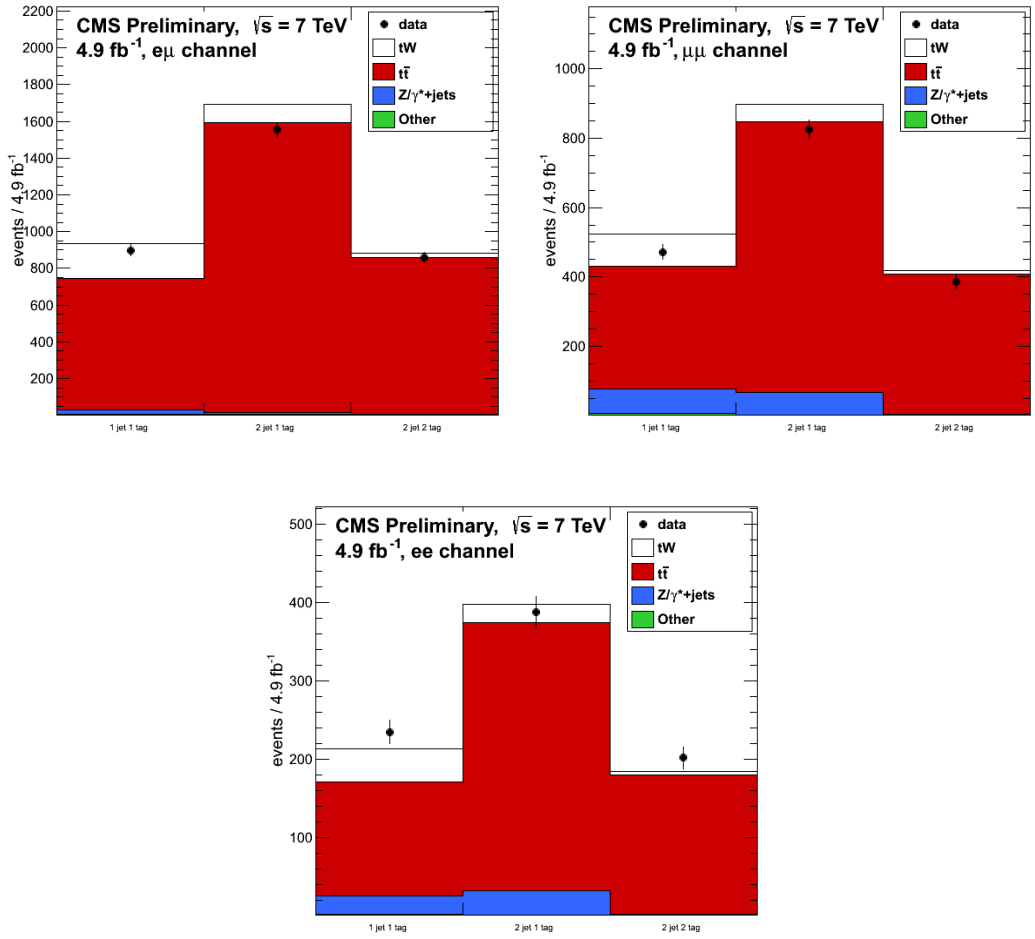


Figure 5.11: Final event yields together in data and simulation in the three sample regions for the $e\mu$, $\mu\mu$ and ee decay channels, after all kinematic requirements.

Chapter 6

Systematic uncertainties

Due to the relatively low cross section for the tW production process the systematic uncertainties on the measurement are of comparable scale to the statistical uncertainties. It is thus vital for a robust result that a comprehensive study of the systematic uncertainties is made. The systematics are broadly divided into two categories:

- *Flat-rate uncertainties* due to limitations of the detector or reconstruction algorithms. Unless specified, they are universal uncertainties, independent of the analysis, and as such are derived by external groups;
- *Scale-factor uncertainties* which are derived by varying adopted scale factors up and down by one standard deviation, with the difference with respect to the nominal result taken as the uncertainty;

All uncertainties are included as nuisance parameters in the statistical fit model, see Section 7.1. The statistical uncertainty, derived for each final state, is also considered.

6.1 Flat-rate uncertainties

The HF sub-detector was used as the CMS luminometer, for both online and offline measurements, until 2011, when the effects of event pile-up and cal-

ibration shift prompted the development of a new measurement technique. The offline **luminosity** measurement in this analysis is made using this new procedure, known as the *Pixel Cluster Counting Method* (PCC) [41]. This method assumes the probability of any given pixel, of the 7×10^7 pixels in the inner part of the CMS detector, being hit by two different tracks from the same bunch crossing to be very small. Therefore, the number of hit pixel clusters per crossing can be taken as a linear function of the number of interactions per crossing and thus a very good measure of luminosity. The measured rates are calibrated using a Van der Meer scan [95]. The PCC method returns a luminosity measurement of 4.9fb^{-1} with a total systematic uncertainty of 2.2% for the entire 2011 data sample [41]. The luminosity uncertainty in this analysis applies to the normalisation of the simulated background estimates, which are scaled to the measured luminosity to match the data.

The **lepton efficiency** measurements are made using the tag-and-probe method to analyse $Z \to l^+l^-$ events, which provide an unbiased and highly pure lepton sample [96]. In the tag-and-probe approach, lepton pairs are considered, where one candidate, called the "tag", satisfies tight identification and isolation requirements, whilst the second lepton, called the "probe", is a probable candidate selected with looser criteria. Tighter identification requirements are then placed on the probe collection, depending on the efficiency being measured. (Although the passing-probe requirements are typically less strict than those for the tag, in all the efficiency measurements described below the tag and passing-probe requirements are the same). The result is two sub-samples: tags with passing probes and tags with failing probes. The efficiency is defined as the fraction of probe leptons which pass the selection criteria being considered.

For electrons the tag-and-probe efficiencies are measured separately for the ECAL barrel and endcaps. These studies required, for sample purity, that the reconstructed mass of the lepton pair be close to the nominal Z^0 boson mass, falling within the 60 - 120 GeV/c^2 range; a broader mass window than that used in the kinematic requirements of this analysis (see Section 5.3).

The total lepton efficiency measurement is broken down into three separate component measurements:

- the efficiency of the **trigger** system to identify probable leptons in proton-proton collisions;
- the efficiency of the **reconstruction** algorithms, described in Section 4.2, to reconstruct leptons from detector information;
- the efficiency of the analysis selection requirements, described in Section 5.2, to **identify** and **isolate** electrons and muons;

To measure the **trigger efficiency** for either lepton type, the probe leptons, passing all other tight lepton kinematic requirements, are then required to pass the HLT trigger. The electron trigger efficiency is measured to be 98.9% and 99.2% in the barrel and endcaps [93], respectively, and 88.3% for muons [58].

The electron **reconstruction efficiency** is defined as the efficiency for an ECAL supercluster (probe) to successfully seed an ECAL-driven electron (tag) - see Section 4.2.5. The probe definition requires clusters to be within the acceptance of the tracking system and ECAL and to have a reconstructed $E_T > 20$ GeV. In data, the reconstruction efficiency is found to be 99.3% in the barrel and 96.8% in the endcaps [93].

For the muon reconstruction efficiency, the triggered muon candidate (probe) must pass all the tight requirements imposed by the muon tracking and spectrometer systems, qualifying as both a global muon and a tracker muon (refer to Section 4.2.4 for the explanation of these muon categories). In data, the reconstruction efficiency is found to be 96.4% for the inner-tracker-derived tracker muons and 99.1% for the spectrometer-derived global muons [97]. These uncertainties were calculated by the CMS Muon and EGamma Physics Object Groups (POGs), for use in all physics analyses.

The electron **identification & isolation efficiency** is defined as the efficiency for a GSF electron track (probe) to pass the kinematic requirements of a *tight* electron (tag) as defined by the Top Physics PAG group and as used in this analysis. In data, the identification & isolation efficiency was found to be 79.1% in the barrel and 69.2% in the endcaps [97]. Similarly for muons, the probes must pass the tight muon kinematic requirements. For muons the isolation efficiency is measured to be 98.5%; there is no identification efficiency for muons, as this is specific to electrons (see Section 4.2.5).

The uncertainties on the efficiencies for reconstruction and identification have been combined into a single systematic uncertainty for each lepton type. The lepton reconstruction and identification efficiency uncertainty is taken to be 1% for muons and 2% for electrons [94]. The trigger efficiency is quoted as an individual systematic, taken as 1.5% for both electrons and muons [94]. It should be noted that the quoted values used in this analysis are not always identical to the referenced papers, as the analyses they describe have progressed and subsequently their estimates have changed. The values used here were accurate at the time this analysis was conducted and have additionally been validated with the authors of the relevant studies.

MC generators produce events with momentum fractions and energies taken from PDFs derived from data collected from many different experiments. These PDFs are therefore subject to the uncertainties coming from these analyses, which must be propagated into any measurements that use the resulting PDFs. In order to assign systematic uncertainties to the result presented here, the re-weighting method [98], defined by the CMS Generator Group, is used which is consistent with the latest PDF4LHC recommendations [99].

The 22 parameters in the CTEQ6.6 PDF set have error PDFs associated with them. A dedicated tool called LHAPDF (*Les Houches Accord PDFs*) [100] is used to generate 44 error PDFs for each event from these 22 parameters, corresponding to two error PDFs per parameter, varied up/down by one standard deviation.

The re-weighting method assumes that the PDF can be varied within its uncertainties whilst not affecting the kinematics nor the topology of the event, i.e. there is no change to the available phase space nor, for example, to the jet multiplicity. A set of $2n + 1$ weights, w, are calculated for each event, defined as the ratio of the PDF values for each of the two protons in a proton-proton collision, evaluated for the different error PDFs, PDF^j, for each of the 22 parameters, n, with respect to the nominal PDF, PDF^0 [98]:

$$w^j = \frac{PDF^j(x_1, f_1, Q) \cdot PDF^j(x_2, f_2, Q)}{PDF^0(x_1, f_1, Q) \cdot PDF^0(x_2, f_2, Q)} \text{ for } 0 \leq j \leq 2n \qquad (6.1)$$

where f is the parton flavour, with momentum fraction x, at the factorisation scale Q, and 1,2 represent the two protons. Following this definition w^0, for the nominal PDF, is equal to unity, whilst all other values vary around it. In order to determine the uncertainty on a given variable, X, it is calculated in turn with each of the $2n + 1$ weights returned by equation 6.1, giving different values for that variable, $X_0 ... X_{2n}$. The largest difference obtained with respect to the nominal X_0 is quoted as the PDF uncertainty.

The PDF uncertainty for this cross section measurement, was found to be 2%, and is one of the smaller systematic uncertainties.

6.2 Scale-factor uncertainties

6.2.1 Pile-up uncertainty

An incorrect modelling of pile-up in the simulation samples can bias the expected number of signal events when applying the selection requirements of the analysis. However, the LHC setup changes fairly regularly; for example an increase in the number of protons per bunch or an improvement in the collimation of the beams can lead to increased proton-proton interactions per bunch crossing. Therefore, to be able to use the simulation to describe the data properly, the vertex multiplicity of the samples are generated with a flat

distribution of primary vertices [101]. The observed vertex distribution in the simulation is then re-weighted for a given scenario, to match the distribution seen in data, using recommendations from the CMS pile-up re-weighting PAG. The remaining differences between the data and simulation are then covered by the pile-up multiplicity systematic uncertainty.

The uncertainty on this process is calculated by varying the re-weighting scale factors by ±5% to cover the uncertainties due to the underlying modelling of pile-up in the simulation; an additional ±3% is then required to cover all of the modelling and physics aspects of the pile-up simulation that have not been properly studied, e.g. differences in the angular distributions or energy spectra with respect to the data [102]. The largest variation with respect to the nominal value is then taken as the systematic uncertainty. The re-weighting procedure only slightly affected the event yields, with the largest variation being ±2% for the signal process, seen in the *ee* channel.

6.2.2 Jet uncertainties

As explained in Section 4.4.1 the jet reconstruction includes corrections to account for differences in jet energy between reconstructed and particle-level jets. Jet energy calibration studies were performed to produce these correction factors [103]. This leads to an important systematic uncertainty; as changes in the Jet Energy Scale (JES) will affect the number of jets that pass or fail the kinematic requirements, potentially changing the final result. By shifting the correction factor by one standard deviation up/down, including in the calculation of the E_T^{miss} of the event, the uncertainty on the JES is derived. The JES uncertainty is asymmetrical, with the maximum of +2.5%/-1.8% seen in the $\mu\mu$ signal sample.

The Jet Energy Resolution (JER) is defined as the standard deviation of a Gaussian fitted to the jet response. Jet asymmetry measurements have shown that the JER in data is approximately 10% broader than in simulation, with an uncertainty of the same order [104]. To account for this all

simulated jet 4-momenta are increased/smeared by 10% to better agree with the distribution seen in data. For the systematic estimate no correction is applied for the systematic scaled-down sample, equivalent to 0%, and a 20% correction for the systematic scaled-up sample. The symmetrical JER uncertainty peaks at ±1.9% in the *ee* signal sample, where it is of a similar magnitude to the JES uncertainty.

6.2.3 E_T^{miss} uncertainties

Events with neutrinos in the final state exhibit a small uncertainty arising from how the E_T^{miss} is modelled. The E_T^{miss} distribution is already affected by the JES and JER uncertainties due to how it is derived. However, the contribution from *unclustered energy* deposits, those not identified as belonging to a lepton or jet, must also be considered. By being unclustered these deposits are effectively uncorrected, unlike those belonging to a lepton or a jet. To compensate for this a conservative estimate of this uncertainty is made by varying the unclustered energy component of the E_T^{miss} by ±10%. This systematic is still applicable in the $e\mu$ channel, as although there is no E_T^{miss} requirement applied, it is used in constructing the H_T variable. The E_T^{miss} modelling uncertainty is one of the lowest contributing uncertainties, peaking at only ±0.3% in the signal sample, for the *ee* channel.

As explained in Section 5.3, to correct for pile-up effects the E_T^{miss} distributions in simulation are scaled to match the data using a Z+jets enriched control sample. As such, the relative difference between the event yields before and after E_T^{miss} scaling is taken as a systematic uncertainty for the Z+jets background process; referred to as *Background normalisation* in Tables 6.1 - 6.3 below. The related uncertainty is one of the largest of all the processes studied, peaking at +36% for the $e\mu$ channel; approximately a factor of 7 larger than the dominant signal process uncertainty.

6.2.4 B-tagging uncertainty

The performance of the b-tagging algorithms has been studied by the b-tag working group, measuring their efficiencies and misidentification probabilities. The differences in efficiencies observed between the data and simulation are used to produce a scale factor, used to correct the simulation performance to match that in data. The performance measurements were carried out separately with $t\bar{t}$ samples [105] and with inclusive multijet samples [88] (i.e. events with one or two charged leptons and ≥ 2 jets). To measure the b-tagging systematic uncertainty, the scale factor was varied up/down within its uncertainties and the effect propagated to the final result. The scale factor and uncertainties used for $t\bar{t}$ processes is 0.956 ± 0.03, from the multijet measurements [88], and 0.960 ± 0.04, from the $t\bar{t}$ measurements [105], is used for the remaining processes. Although the two derived uncertainty estimates are almost identical, the $t\bar{t}$ measurement is still used because of the importance of this background process in the analysis.

The b-tagging efficiency is one of the largest systematic uncertainties for this analysis, with peak absolute values ranging between 4.1% and 4.6% for all three lepton channels in the signal process, and rising as high as +6.3% for the $t\bar{t}$ background.

6.2.5 Simulation modelling uncertainties

Uncertainties on the **QCD renormalisation and factorisation scales** in the $t\bar{t}$ and tW MC samples are determined by varying the scale for hard scattering, $Q^2 = m_{top}^2 + \sum p_T^2$, and the subsequent parton showering simultaneously. Q is varied by a factor of 0.5 and 2.0 with respect to the nominal value. The resulting uncertainties are less than +2% for the signal process.

Hard scattering at a hadron collider can potentially produce gluon radiation from both the incoming and outgoing partons, called initial-state (ISR) and final-state radiation (FSR), respectively. The additional uncertainty due to this process is no longer determined separately for this analysis, as the variation of the Q^2 uncertainty already accounts for this effect.

As explained in Section 4.1, generating the $t\bar{t}$, Z+jets and W+jets simulation samples involves matching the MadGraph matrix-element production of the hard processes to the partons and shower-generated jets simulated by PYTHIA. This matching process is governed by the MLM **parton level matching threshold** scale [63]. In the nominal $t\bar{t}$ simulation sample, the threshold for interfacing the matrix elements and parton showering is 20 GeV. Additional samples are produced with the threshold set at 10 and 40 GeV, to allow the related uncertainty to be estimated, referred to as *ME/PS* in Tables 6.1 - 6.3 below. This is measured for $t\bar{t}$ only, as it is the dominant background process, where it peaks at +5.9% as one of the largest uncertainties.

The difference between the default **Diagram Removal** (DR) **and Diagram Subtraction** (DS) schemes used in the production of the tW simulated samples is also included as a systematic uncertainty for signal (see to Section 2.3 for an explanation of the two schemes). The relative difference in final event yields between the samples is taken as the uncertainty, found to be +3%.

6.3 Summary of systematic uncertainties

The impact of all considered systematic uncertainty sources on the event yields are summarised as percentages for the signal region in Table 6.1, and the two $t\bar{t}$ enriched control regions in Tables 6.2 and 6.3. The systematic uncertainties are calculated for the control regions as well as the signal region as they are used in the statistical treatment, described in Section 7.1, to evaluate the final cross section result. If two numbers are listed for a single uncertainty, the upper number is the effect on the event yield when the systematic uncertainty source is scaled up and the lower for when it is scaled down. In the cases where the systematic was found to be symmetric between the scaled up and scaled down effects only one value is listed. Where

a certain systematic does not apply is denoted with "-", and in the case that a systematic is negligible when compared to others for the same process or by itself is denoted with "$\star$".

Table 6.1: Event yield impact of systematic uncertainties for the signal region (1j1t), given as a percentage.

Systematic Uncertainty ($ee/e\mu/\mu\mu$) [%]	signal tW	$t\bar{t}$	Z/γ^*	other
Luminosity	2.2	2.2	2.2	2.2
Pile-up multiplicity	1.9/0.4/0.1	0.8/0.6/0.5	1.4/2.2/1.8	$\star$
Trigger efficiency	1.5	1.5	1.5	1.5
Muon reconstruction and identification	-/1/1	-/1/1	-/1/1	-/1/1
Electron reconstruction and identification	2/2/-	2/2/-	2/2/-	2/2/-
JES	$^{+0.7}_{+1.5}$ / $^{-0.4}_{+0.6}$ / $^{-1.8}_{+2.5}$	$^{-0.9}_{+6.3}$ / $^{-4.5}_{+3.4}$ / $^{-4.9}_{+5.8}$	$\star$	$\star$
JER	1.9/0.1/1.1	2.1/2.3/2.2	$\star$	$\star$
B-tagging	$^{-4.5}_{+4.6}$ / $^{-4.1}_{+3.9}$ / $^{-4.3}_{+4.6}$	$^{-1.8}_{+3.5}$ / $^{-3.4}_{+2.5}$ / $^{-2.8}_{+3.3}$	$\star$	$\star$
Factorisation/Normalisation Scale (Q^2)	$^{+0.3}_{+1.8}$	$^{-5.6}_{+12}$	-	-
ME/PS matching thresholds	-	$^{+0.3}_{+5.9}$	-	-
DR/DS scheme	3	-	-	-
E_T^{miss} modelling	0.3/0.2/0.2	0.7/0.1/0.1	10/1.7/20	$\star$
PDF uncertainties	2	$\star$	-	-
Background normalisation	-	-	33/36/27	-
Theoretical cross sections	-	6	5	$\star$
Simulation statistics	4.0/1.9/2.7	4.5/2.0/2.9	19/25/14	17/8.9/42

Table 6.2: Event yield impact of systematic uncertainties for the signal region (2j1t), given as a percentage.

Systematic Uncertainty ($ee/e\mu/\mu\mu$) [%]	signal tW	$t\bar{t}$	Z/γ^*	other
Luminosity	2.2	2.2	2.2	2.2
Pile-up multiplicity	1.5/0.1/1.5	0.7/0.1/0.7	1.3/3.4/5.4	$\star$
Trigger efficiency	1.5	1.5	1.5	1.5
Muon reconstruction and identification	-/1/1	-/1/1	-/1/1	-/1/1
Electron reconstruction and identification	2/2/-	2/2/-	2/2/-	2/2/-
JES	$^{+1.7}_{-6.4}$/$^{+1.4}_{-3.2}$/$^{-1.8}_{+2.5}$	$^{-0.8}_{-1.5}$/$^{+0.3}_{-0.4}$/$^{-0.1}_{+0.2}$	$\star$	$\star$
JER	3.2/0.6/0.8	2.7/1.6/0.2	$\star$	$\star$
B-tagging	$^{+1.0}_{-1.5}$/$^{+1.9}_{-1.9}$/$^{+0.5}_{-3.4}$	$^{-0.3}_{-0.8}$/$^{-0.8}_{+0.4}$/$^{-0.1}_{+0.6}$	$\star$	$\star$
Factorisation/Normalisation Scale (Q^2)	$^{+2.6}_{-4.3}$	$^{+5.3}_{-4.3}$	-	-
ME/PS matching thresholds	-	$^{+0.8}_{-7.1}$	-	-
DR/DS scheme	-11	-	-	-
E_T^{miss} modelling	0.5/0.1/0.5	$\star$/0.1/0.1	7.7/4.1/1.1	$\star$
PDF uncertainties	2	$\star$	-	-
Background normalisation	-	-	32/30/28	-
Theoretical cross sections	-	6	5	$\star$
Simulation statistics	5.3/2.5/3.5	2.9/1.4/2.0	20/29/12	18/12/14

Table 6.3: Event yield impact of systematic uncertainties for the signal region (2j2t), given as a percentage.

Systematic Uncertainty ($ee/e\mu/\mu\mu$) [%]	signal tW	$t\bar{t}$	Z/γ^*	other
Luminosity	2.2	2.2	2.2	2.2
Pile-up multiplicity	1.2/0.1/1.4	1.2/0.4/1.3	1.1/4.9/3.0	$\star$
Trigger efficiency	1.5	1.5	1.5	1.5
Muon reconstruction and identification	-/1/1	-/1/1	-/1/1	-/1/1
Electron reconstruction and identification	2/2/-	2/2/-	2/2/-	2/2/-
JES	$^{-3.5}_{+11.0}$ / $^{+1.0}_{+2.4}$ / $^{+6.8}_{+6.5}$	$^{+0.9}_{+1.1}$ / $^{-0.9}_{-0.4}$ / $^{-0.3}_{+1.0}$	$\star$	$\star$
JER	3.9/3.0/9.1	2.8/0.1/2.0	$\star$	$\star$
B-tagging	$^{+13.0}_{-9.6}$ / $^{+8.7}_{-9.2}$ / $^{+15.0}_{-7.6}$	$^{+6.5}_{-5.4}$ / $^{+6.7}_{-6.7}$ / $^{+6.5}_{-6.8}$	$\star$	$\star$
Factorisation/Normalisation Scale (Q^2)	$^{+3.4}_{-5.7}$	$^{+15}_{-3}$	-	-
ME/PS matching thresholds	-	$^{+1.5}_{+13.0}$	-	-
DR/DS scheme	-14	-	-	-
E_T^{miss} modelling	1.3/0.1/1.0	0.7/0.1/0.5	6.3/$\star$/22.0	$\star$
PDF uncertainties	2	$\star$	-	-
Background normalisation	-	-	27/28/33	-
Theoretical cross sections	-	6	5	$\star$
Simulation statistics	11.0/4.9/7.5	4.1/1.9/2.7	62/100/48	47/90/40

Chapter 7

Results

A likelihood function is used to extract the cross section from the cut-based analysis and to evaluate the statistical significance of the result. The statistical model is composed of one parameter of interest, the tW signal cross section, and various nuisance parameters, such as the background event yields and systematic uncertainties. The signal cross section, and 68% confidence level, is determined using a profile likelihood method [106], for the reason that it is well suited to produce confidence levels from small samples, as is expected in this analysis.

To quantify the excess of signal events over the expected background, a maximised likelihood ratio is used [107], comparing a purely background model to the combined signal and background model. This measures the significance of the cross section result.

Once the cross section has been measured it is compared to the theoretical cross section prediction to directly estimate the CKM matrix element $|V_{tb}|$.

7.1 Significance and cross section

The extraction of the cross section and statistical significance uses a binned likelihood function of the event counts assuming a Poisson distribution. The function is binned in terms of the three dilepton final states, $ee/e\mu/\mu\mu$, and the three sample regions: the signal region (1j1t) and two control regions (2j1t and 2j2t). The inclusion of all three sample regions enables the data-driven, simultaneous estimation of the $t\bar{t}$ background contribution. The expected event count for bin i, x_i, is given by the sum of all background processes and the tW signal:

$$x_i = \mu S_i + \sum_k B_{k,i} \tag{7.1}$$

where k runs over all background processes, $B_{k,i}$ is the background contribution from process k in bin i, and S_i is the signal contribution, scaled by the signal cross section μ, which is itself scaled by the SM predicted cross section.

The event yields, x_i, are affected by the presence of systematic uncertainties, which were discussed in Chapter 6. For each independent source of uncertainty u, a nuisance parameter δ_u is introduced, with a Gaussian prior of mean 0 and width 1. The inclusion of the nuisance parameters allows the corresponding event yield to vary by up to ± 1 standard deviation; including for the parameters that exhibit asymmetric event yield changes. Similar models with the same treatment of systematic uncertainties are used in other multi-channel counting experiments, such as Higgs searches [94].

The signal cross section and associated 68% confidence level is evaluated using the profile likelihood method. The profile likelihood function depends only on the signal cross section μ, and is defined by maximising the full likelihood function with respect to the nuisance parameters δ_u, estimating the optimal parameter values with respect to the data sample. The likelihood function is given by the product of the Poisson probabilities in each bin i, using x_i from equation 7.1, and the Gaussian priors for the parameters δ_u:

$$L(\mu, \delta_u) = \sum_i \frac{x_i^{n_i} e^{-x_i}}{n_i!} \cdot \frac{1}{\sqrt{2\pi}} exp(-\frac{\delta_u^2}{2}) \tag{7.2}$$

where n_i is the number of observed data events in bin i and x_i depends on both μ and δ_u. The cross section measurement is obtained by maximising the likelihood function in equation 7.2. The 68% confidence level for this calculated value of μ is then defined by a ratio of maximised likelihood:

$$C.L. = -2\ln \frac{max_{\delta_u} L(\mu, \hat{\hat{\delta}}_u)}{max_{\mu,\delta_u} L(\hat{\mu}, \hat{\delta}_u)} \tag{7.3}$$

where in the numerator the likelihood function μ is fixed whilst the nuisance parameters $\hat{\hat{\delta}}_u$, that maximise it, are allowed to float. In the denominator both parameters are allowed to float in the maximisation, with the constraint of $\hat{\mu} > 0$.

To calculate the significance of the measurement, a ratio of maximised likelihood is used:

$$\lambda = -2\ln\left(\frac{L_{s+b}}{L_b}\right) = -2\ln \frac{max_{\mu,\delta_u} L(\mu, \delta_u)}{max_{\delta_u} L(\mu = 0, \delta_u)} \tag{7.4}$$

where in the numerator, the signal cross section parameter μ is allowed to float, whereas in the denominator, for the background-only case, it is fixed to zero. Using toy simulations with no signal ($\mu = 0$), the distribution of λ for the background-only hypothesis is estimated and is quoted as the observed significance. To evaluate the expected significance, toy simulations for the signal plus background hypothesis, assuming the SM signal cross section $\mu = 1$, are performed. The median and central 68% of the values obtained are quoted as the expected significance and its related uncertainty.

For the CMS single top t-channel cross section analysis [11], which uses

the same statistical treatment and 2011 data, it was believed that the scale-factor uncertainties might be over-constraining the fit, leading to an under-estimation of their effects on the result. Therefore, when maximising the profile likelihood functions, the nuisance parameters corresponding to the scale-factor uncertainties were fixed to the nominal value, and not allowed to vary. Instead, to calculate each of these uncertainties, two specific toy MC samples were generated with the corresponding nuisance parameter shifted by $\pm 1\sigma$. The resulting differences in signal event yield using these samples were taken as the systematic uncertainties and added in quadrature to the uncertainty on the profile likelihood interval. This conservative approach led to a slight reduction in the significance and the cross section estimate, but is a more robust approach. It is therefore also used for this analysis.

The resulting significance, seen in data, obtained using the method described is 3.5σ, whilst the expected significance, from simulation, is $3.2\sigma \pm 0.9\sigma$.

The obtained cross section value and 68% confidence level is $14.9^{+5.0}_{-5.1}$ (stat $\oplus$ syst) pb. This is in agreement with the Standard Model prediction of $15.6 \pm 0.4^{+1.0}_{-1.2}$ pb [3].

The significance calculation is also repeated using the 1j1t signal region only. This calculation does not constrain the $t\bar{t}$ background with the two control regions, instead relying solely on the simulation modelling in the signal region. The observed significance obtained is 1.3σ, while the expected significance is $1.7\sigma \pm 0.8\sigma$, resulting in a cross section value of $15.6^{+6.2}_{-9.4}$ (stat $\oplus$ syst) pb. This method sees a pronounced drop in the significance and increase in the uncertainties of the cross section measurement, validating the data-driven approach used to handle the $t\bar{t}$ background.

7.2 $|V_{tb}|$ estimation

The cross section measurement can be used to test CKM matrix unitarity under the assumption that $|V_{td}|$ and $|V_{ts}|$ are much smaller than $|V_{tb}|$, which is inline with the current measurements of these matrix elements (see Section 2.1). Since tb coupling is present in the single top production vertex, as well as in top decay, the central value of $|V_{tb}|$ can be directly obtained here from the cross section measurement, which is proportional to the vertex function squared, and so to $|V_{tb}|^2$. As such, $|V_{tb}|$ is calculated from:

$$|V_{tb}| = \sqrt{\frac{\sigma_{tW}}{\sigma_{tW}^{th}}} = 0.98^{+0.16}_{-0.17}(\text{exp.})^{+0.03}_{-0.04}(\text{th.}) \tag{7.5}$$

in which σ_{tW} is the measured cross section and σ_{tW}^{th} the cross section prediction assuming $|V_{tb}| = 1$.

This result is compatible with the Tevatron value of $|V_{tb}| = 0.88^{+0.07}_{-0.07}$, obtained from the combination of D0 and CDF single top cross section measurements [28]. The result is also compatible with the measurement of $|V_{tb}| = 1.14 \pm 0.22$ [3], derived by the Particle Data Group from the recent CMS single top t-channel cross section measurement [11].

7.3 Other cross section measurements at the LHC

In addition to the analysis reported in this thesis, there have been two other published measurements of the tW process at the LHC. The ATLAS and CMS detector experiments have both produced cross section measurements using a boosted decision trees (BDT) method.

7.3.1 ATLAS analysis

The ATLAS experiment's first measurement of the W-associated single top quark production cross section was made using 2.05 fb^{-1} of $\sqrt{s} = 7$ TeV pp collisions, taken during the first half of the 2011 period [24]. They measure a cross section of $16.8 \pm 2.9(\text{stat.}) \pm 4.9(\text{syst.})$ pb, with an observed (expected) significance of 3.3σ (3.4σ). Whilst using approximately half as much data as in this analysis a comparable result is achieved by implementing a multivariate analysis method; using a similar signal event selection, a BDT approach was used to improve the discrimination of tW signal events from the backgrounds, particularly $t\bar{t}$. The BDT was constructed using 22 different input variables. Excluding the analysis method itself, the two key differences between the analyses are as follows:

- **b-tagging** - the ATLAS analysis does not have any requirements for the jets to originate from b-quarks, as they deemed this to not offer significant background rejection. By not implementing b-tagging, the ATLAS group increased the size of the sample of signal events, at the cost of a reduced purity. However, the BDT method compensated for this by making a more comprehensive use of the selected data; which is evident in the comparable cross section result achieved. This decision also benefits the ATLAS analysis by losing the related systematic uncertainty, which at approximately 4.5% dominates the CMS cut-based analysis;

- **LO single top simulation** - the tW interference with $t\bar{t}$ is handled by using the (non-interfering) LO-approximation for the tW simulation samples. The differences between the LO and NLO tW calculation is considered as a relatively large modelling uncertainty of 10%. For comparison, the equivalent systematic in the cut-based analysis, the uncertainty due to differences in the DR/DS modelling schemes, is only 3%. The largest cut-based uncertainty, due to b-tagging, is less than half the size of the ATLAS modelling uncertainty.

It is hard to directly compare the specifics of the two analyses, due to the aforementioned method and data sample differences, but also due to the uncertainties being handled rather differently; a notorious complication when combining results from the two experiments, such as for the Higgs boson searches [34]. The achieved cross section results, however, are compatible with each other and with the SM expectation.

The cut-based cross section is slightly closer to the SM, with marginally smaller uncertainties, and a larger observed significance. The ATLAS result also presents larger individual systematic errors, ranging between $5 - 16\%$, which are more conservative than the cut-based analysis. However, these small improvements are not proportional to the significant increase in data used by the cut-based analysis; the similarity in results demonstrates that the simplistic cut-based approach does not utilise the data as efficiently as the BDT, as one reasonably expects.

The ATLAS measurement obtains a value of $|V_{tb}| = 1.03^{+0.16}_{-0.19}$, using the same calculation as presented here. This result is compatible with the cut-based estimate.

7.3.2 CMS BDT analysis

For the tW CMS publication [22], the cut-based analysis described in this thesis was presented as the cross-check to a main BDT analysis. The preliminary cut-based work showed the result to be statistics limited, in as much as it was projected to not be possible to exceed a significance of 5σ before the planned 2013 LHC shutdown. Therefore, the decision was made to also conduct a multivariate analysis, which would make greater use of the information in the available data and simulation samples, leading to an improved result.

A simple BDT was constructed from just four variables, which had the

highest discriminating power, using the same statistical treatment and almost identical kinematic requirements as the cut-based analysis. The four BDT variables were the transverse momentum of the leading jet, the difference in angle η between the E_T^{miss} and the closest of the two leptons, the H_T of the system and the total p_T of the system; where the system refers to the signal event topology of two leptons, E_T^{miss} and a single b-tagged jet.

A cross section measurement of 16^{+5}_{-4} pb was obtained, with an observed (expected) significance of 4.0σ (3.6σ). This is in good agreement with the cut-based result and the SM expectation. Although the BDT is relatively simple, the achieved result is closer to the SM expectation and has a larger significance than the cut-based analysis, although with no reduction in the related uncertainties. These improvements are as expected of a multivariate method. The CMS BDT cross section measurement is, at the time of writing, the current leading result on W-associated single top quark production.

The BDT result was also used to determine a compatible value of $|V_{tb}| = 1.01^{+0.16}_{-0.13}(\text{exp.})^{+0.03}_{-0.04}(\text{th.})$, using the exact same formula as the cut-based measurement.

Chapter 8

Conclusion

In summary, using 4.9 fb^{-1} of data collected with the CMS experiment at the LHC, evidence has been found for the associated production of a single top quark and $\text{W}^{\pm}$ boson in pp collisions at $\sqrt{s} = 7$ TeV. Focusing on the leptonic final states of the process, a cut-based analysis has been defined. The observed significance of the signal process over the backgrounds was found to be 3.5σ, while the expected significance is $3.2\sigma \pm 0.9\sigma$. The cross section measured is $14.9^{+5.0}_{-5.1}$ (stat $\oplus$ syst) pb, in agreement with the Standard Model expectation.

The aim of this analysis was to produce a baseline measurement for the process using a relatively basic, but robust, cut-based method. It was to form the initial CMS measurement of the process, to be improved upon by more complex multivariate techniques, whilst remaining as a useful cross-check. Although the method is less complex than the other two LHC measurements, which both implement boosted decision trees, the results presented here are highly consistent, in both the measurement and the related uncertainties.

This analysis does exhibit some large sources of systematic uncertainty, placing an error of approximately 30% on the cross section measurement; comparable to those of the initial CMS t-channel cross section measurements [11]. It should be noted that this tW analysis is the first measurement of

a less common single top process, with complex backgrounds, and that the achieved results are consistent with theory.

8.1 Future measurements

A follow-up analysis is already underway on CMS implementing a BDT, with the cut-based method as a cross-check, using the $\sqrt{s} = 8$ TeV pp collision data recorded in 2012. This corresponds to 23.27 fb^{-1} [108], a significant increase, of approximately a factor of five, in the amount of data compared to 2011. This will allow a cut-based analysis to impose tighter kinematic requirements in the signal event selection, whilst retaining a statistically significant number of events (and similarly for future analyses at $\sqrt{s} = 14$ TeV). The increased purity and size of the signal samples will reduce the systematic uncertainty from the backgrounds, leading to an improved result.

However, some uncertainties will become more pronounced as the centre-of-mass energy increases, namely those due to pile-up effects. As the energy increases so will the number of jets produced, and the high luminosity will lead to multiple simultaneous events. The methods used to measure the related systematic uncertainties, such as pile-up multiplicity, will need to be validated at the new LHC running specifications, and more than likely require improvements. Especially crucial to this analysis is the modelling of the Z+jets background process, which saw noticeable disagreement between the data and simulation due to pile-up, requiring correction scale factors and the use of two $\mathrm{E_T^{miss}}$ variables to exclude the background. The current method of luminosity calculation will need to be reviewed when the LHC reaches design specifications, as it is based on the assumption of one track per pixel hit; whilst reasonable at the centre-of-mass of this analysis, it may not hold true under the 'cluttered' conditions of $\sqrt{s} = 14$ TeV collisions.

Two limiting systematic uncertainties, (relatively) independent of the LHC running conditions, will persist in future analyses: the JES correc-

tion and b-tagging uncertainty. These two uncertainties dominate top quark physics searches, the latter being the largest uncertainty for the signal process in this analysis.

The cut-based analysis method itself will not go unchanged in future analyses. For example, the CMS Top PAG, have now recommended the use of Type 1 corrected PF E_T^{miss} to correct Jet Energy Scale differences between the measured calorimeter jet energy and the final state particle-level jet. However, it is the BDT method that will see the most modifications, as it is the principle method and has the greater scope for development. The BDT construction in the 2011 analysis was relatively simple, utilising only 4 variables, compared to the 22 used in the ATLAS measurement. The aim for the 2012 result is to construct the BDT from a larger number of variables to improve the discriminating power; not to imply that using more variables leads to an inherently better BDT. Additionally, the method of training the BDT is being investigated, as the 2011 result was hindered by the limited statistics available; something which has improved with the $\sqrt{s} = 8$ TeV running.

The first aims of future measurements are to improve the precision of the existing cross section measurement, reducing the uncertainties, and to increase the significance of the result above 5σ, the statistical significance associated with *observation* for a process. When this has been achieved further properties of the tW-channel will be studied to confirm theoretical predictions, such as a top mass measurement.

Bibliography

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

[2] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *CMS-HIG-12-028, submitted to Phys. Lett. B*, 2012.

[3] J. Beringer et al. (Particle Data Group). Review of particle physics, 2012. URL `http://pdg.lbl.gov`.

[4] A. V. Bednyakov V. A. Bednyakov, N. D. Giokaris. On Higgs mass generation mechanism in the Standard Model. *Phys. Part. Nucl.*, 39: 13–36, 2008.

[5] D. Griffiths. *Introduction to Elementary Particles.* WILEY-VCH, second, revised edition edition, 2008.

[6] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49 (2):652–657, 1973.

[7] 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.

[8] D0 Collaboration. Measurement of the single top quark production cross sections and $|V_{tb}|$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *FERMILAB-PUB-11-380-E, submitted to Phys. Rev. D*, 2011.

[9] CDF Collaboration. Observation of Single Top Quark Production and Measurement of $|V_{tb}|$ with CDF. *Phys. Rev. D.*, 82(112005), 2010.

[10] CDF Collaboration. Search for single top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV in the missing transverse energy plus jets topology. *Phys. Rev. D.*, 81(072003), 2010.

[11] The CMS Collaboration. Measurement of the t-channel single top quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. Lett.*, 107(091802), 2011.

[12] URL `http://www.quantumdiaries.org/author/emily-thompson/`.

[13] F. Margaroli (for the CDF and D0 collaborations). Top quark pair production at the Tevatron. *PoS ICHEP2010:064*, 2010.

[14] The ATLAS collaboration. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV. *European Physical Journal C*, 71(1577), 2011.

[15] The CMS Collaboration. First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B.*, 695:424–443, 2011.

[16] N. Kidonakis and B. D. Pecjak. Top-quark production and QCD. *European Physical Journal C*, 2011.

[17] N. Kidonakis. Single top quark production at the Fermilab Tevatron: Threshold resummation and finite-order soft gluon corrections. *Phys. Rev. D.*, 74(114012), 2006.

[18] N. Kidonakis. Next-to-next-to-leading logarithm resummation for s-channel single top quark production. *Phys. Rev. D.*, 81(054028), 2010.

[19] N. Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-. *Phys. Rev. D.*, 82(054018), 2010.

[20] N. Kidonakis. Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution. *Phys. Rev. D.*, 82(114030), 2010.

[21] N. Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys. Rev. D.*, 83(091503), 2011.

[22] The CMS Collaboration. Evidence for associated production of a single top quark and W boson in *pp* collisions at 7 TeV. *submitted to Phys.Rev.Lett.*, 2012. URL `arXiv:1209.3489 [hep-ex]`.

[23] P. Motylinski B.Webber C. D. White S. Frixione, E. Laenen. Single-top hadroproduction in association with a W boson. *JHEP*, 29, 2008.

[24] The ATLAS collaboration. Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV. *submitted to Phys. Lett. B*, 2012. URL `arXiv:1205.5764`.

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

[26] F.Tramontano J. Campbell. Next-to-leading order corrections to Wt production and decay. *Nucl.Phys.B726*, 726:109–130, 2005.

[27] M. Nowakowski and A. Pilaftsis. ON GAUGE INVARIANCE OF BREIT-WIGNER PROPAGATORS. *Z. Phys. C*, 60:121–126, 1993.

[28] CDF and D0 Collaboration. Combination of CDF and D0 Measurement of the Single Top Production Cross Section. URL `arXiv:0908.2171`.

[29] CDF Collaboration. Observation of Electroweak Single Top Quark Production. *Phys. Rev. Lett.*, 103(092002), 2009.

[30] D0 Collaboration. Observation of Single Top-Quark Production. *Phys. Rev. Lett.*, 103(092001), 2009.

[31] The ATLAS collaboration. Search for s-channel single top-quark production in pp collisions at $\sqrt{s} = 7$ TeV. August 2011. ATLAS-CONF-2011-118.

[32] 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. B*, 2012.

[33] T. Aaltonen for the CDF Collaboration. Precise measurement of the W-boson mass with the CDF II detector. 2012. URL `arXiv:1203.0275 [hep-ex]`.

[34] The ATLAS and CMS collaborations. Combined Standard Model Higgs boson searches with up to 2.3 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV at the LHC. *ATLAS-CONF-2011-157, CMS-PAS-HIG-11-023*, 2011.

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

[36] CERN Press Office. Incident in LHC sector 3-4. *CERN Press Release PR09.08*, 2008. URL `http://press-archived.web.cern.ch/`.

[37] The CMS Collaboration. LHC Commissioning. URL `http://lhc-commissioning.web.cern.ch/lhc-commissioning/`.

[38] URL `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

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

[40] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3(S08004), 2008.

[41] The CMS Collaboration. Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update. *CMS PAS SMP-12-008*, 2012.

[42] The CMS Collaboration. CMS Tracking Performance Results from Early LHC Operation. *CMS PAPER TRK-10-001*, 2010.

[43] The CMS Collaboration. Description and performance of the CMS track and primary vertex reconstruction. *CMS PAPER TRK-11-001*, 2012.

[44] P. Lecoq et al. Lead tungstate ($PbWO_4$) scintillators for LHC EM calorimetry. *Nucl. Instrum. Meth. A*, 365:291–298, 1995.

[45] The CMS Collaboration. CMS Physics : Technical Design Report Volume 1: Detector Performance and Software. *CERN-LHCC-2006-001*, 2006.

[46] The CMS Collaboration. Energy resolution of the barrel of the CMS Electromagnetic Calorimeter. *JINST*, 2(P04004), 2007.

[47] The CMS Collaboration. CMS: The Electromagnetic Calorimeter Technical Design Report. *CERN-LHCC-97-033*, 1997.

[48] The CMS Collaboration. Electromagnetic calorimeter calibration with 7 TeV data. *CMS PAS EGM-10-003*, 2010.

[49] N. Akchurin et al. Quartz fibers as active elements in detectors for particle physics. *Rev. Sci. Instrum.*, 74(2955), 2002.

[50] The CMS Collaboration. Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data. *JINST*, 5(T03012), 2010.

[51] The CMS-HCAL Collaboration. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. *European Physical Journal C*, 55:159–171, 2008.

[52] The CMS-HCAL Collaboration. Design, performance, and calibration of CMS forward calorimeter wedges. *European Physical Journal C*, 53: 139–166, 2008.

[53] The CMS Collaboration. The CMS muon project: Technical Design Report. *CERN-LHCC-97-032*, 1997.

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

[55] The CMS Collaboration. Fine synchronization of the CMS muon drift-tube local trigger using cosmic rays. *JINST*, 5(T03004), 2010.

[56] The CMS Collaboration. Aligning the CMS muon chambers with the muon alignment system during an extended cosmic ray run. *JINST*, 5 (T03019), 2010.

[57] The CMS Collaboration. Performance of CMS muon reconstruction in cosmic-ray events. *JINST*, 5(T03022), 2010.

[58] The CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *CMS-MUO-10-004*, 2012.

[59] The CMS Collaboration. Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. *CMS PAS PFT-10-001*, 2010.

[60] M. Oreglia (editors) H. Aihara, P. Burrows. SiD Letter of Intent. *arXiv:0911.0006*, 2009.

[61] S. Mrenna T. Sjöstrand and P. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05, 2006.

[62] J. Alwall et al. MadGraph/MadEvent v4: the new web generation. *JHEP*, 06(128), 2011.

[63] F. Piccinini et al. M. L. Mangano, M. Moretti. Matching matrix elements and shower evolution for top-pair production in hadronic collisions. *JHEP*, 01, 2007.

[64] J. H. Kuhn S. Jadach and Z. Was. TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons. *Comput. Phys. Commun.*, 64(2), 1991.

[65] Qing-Hong Cao Joey Huston Jon Pumplin Daniel Stump Wu-Ki Tung Pavel M. Nadolsky, Hung-Liang Lai and C.-P. Yuan. Implications of CTEQ global analysis for collider observables. *Phys. Rev. D.*, 78 (013004), 2008.

[66] GEANT4 Collaboration. GEANT4: a simulation toolkit. *Nucl. Instrum. Meth.*, A506(250), 2003.

[67] J. Allison et al. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1), 2006.

[68] F. Petriello et al. R. Gavin, Y. Li. FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order. *Comput. Phys. Commun*, 182, 2011.

[69] F. Petriello et al. R. Gavin, Y. Li. W physics at the LHC with FEWZ 2.1. 2012. URL `arXiv:1201.5896 [hep-ph]`.

[70] URL `http://mcfm.fnal.gov/`.

[71] M. Czakon and A. Mitov. Inclusive heavy flavour hadroproduction in NLO QCD: The exact analytic result. *Nucl. Phys. B.*, 824:111–135, 2010.

[72] The CMS Collaboration. Track reconstruction in the CMS Tracker. *CMS PAS TRK-09-001*, 2009.

[73] The CMS Collaboration. Tracking and Primary Vertex Results in First 7 TeV Collisions. *CMS PAS TRK-10-005*, 2010.

[74] The CMS Collaboration. Adaptive Vertex Fitting. *CMS Note 2007/008*, 2007.

[75] The CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}. *CMS PAS PFT-09-001*, 2009.

[76] A. Strandlie T. Todorov W. Adam, R. Fruhwirth. Reconstructions of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31, 2005.

[77] P. Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nucl. Instrum. Meth. A*, 294 (1-2), 1990.

[78] The CMS Collaboration. Commissioning of the particle-flow event reconstruction with leptons from J\Psi and W decays at 7 TeV. *CMS PAS PFT-10-003*, 2010.

[79] P.R Byron et al. Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification. *Nucl. Instrum. Meth. A*, 12(018), 2005.

[80] G. P. Salam M. Cacciari and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 4, April 2008.

[81] D. E. Soper S. D. Ellis. Successive Combination Jet Algorithm For Hadron Collisions. *Phys. Rev. D.*, 48(3160), October 1993.

[82] The CMS Collaboration. Jet Performance in *pp* Collisions at $\sqrt{s} = 7$ TeV. *CMS PAS JME-10-003*, 2010.

[83] The CMS Collaboration. Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS. *CMS PAS JME-09-002*, 2009.

[84] The CMS Collaboration. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6(11002), 2011.

[85] The CMS Collaboration. Performance of Jet Algorithms in CMS. *CMS PAS JME-07-003*, 2007.

[86] The CMS Collaboration. Performance of Track-Corrected Missing E_T in CMS. *CMS PAS-JME-09-010*, 2009.

[87] The CMS Collaboration. Missing transverse energy performance of the CMS detector. *CERN-PH-EP-2011-051, submitted to JINST*, 2011.

[88] The CMS Collaboration. Identification of b-quark jets with the CMS Experiment. *CMS PAS BTV-12-001*, 2012.

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

[90] The CMS Collaboration. Impact of Tracker Misalignment on the CMS b-Tagging Performance. *CMS PAS BTV-07-003*, 2007.

[91] The CMS Collaboration. HCAL Noise Library, January 2011. URL `https://twiki.cern.ch/twiki/bin/viewauth/CMS/HcalNoiseInfoLibrary`.

[92] CMS Collaboration. Recipes for 2010 collisions, March 2011. URL `https://twiki.cern.ch/twiki/bin/viewauth/CMS/Collisions2010Recipes`.

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

[94] The CMS Collaboration. Search for the standard model Higgs boson decaying to a W pair in the fully leptonic final state in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 710(91), 2012.

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

[96] The CMS Collaboration. Measurements of inclusive W and Z cross sections in *pp* collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 2011.

[97] J. Alcaraz Maestre et al. Updated Measurements of the Inclusive W and Z Cross Sections at 7 TeV. *CMS AN-2010/264*, 2010.

[98] The CMS Collaboration. Parton Distribution Uncertainty Determination within CMSSW. *CMS AN-2009/048*, 2009.

[99] URL `http://www.hep.ucl.ac.uk/pdf4lhc/`.

[100] D Bourilkov M R Whalley and R C Group. The Les Houches Accord PDFs (LHAPDF) and LHAGLUE. 2005. URL arXiv:hep-ph/0508110v1.

[101] M. Hildreth. Pileup Reweighting for 2011 Data and Monte Carlo, . URL https://twiki.cern.ch/twiki/bin/view/CMS/Pileup_2011_Reweighting?rev=4.

[102] M. Hildreth. Estimating Systematic Errors Due to Pileup Modeling, . URL https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors.

[103] The CMS Collaboration. Determination of the Jet Energy Scale in CMS with *pp* Collisions at $\sqrt{s} = 7$ TeV. *CMS PAS JME-10-010*, 2010.

[104] The CMS Collaboration. Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV. *CMS PAS JME-10-014*, 2011.

[105] The CMS Collaboration. Measurement of the b-tagging efficiency using $\mathrm{t\bar{t}}$ events. *CMS PAS BTV-11-003*, 2012.

[106] Jan Conrad Wolfgang A. Rolke, Angel M. Lopez. Limits and Confidence Intervals in the Presence of Nuisance Parameters. *Nucl. Instrum. Meth. A*, 551:493–503, 2005.

[107] A. L. Read. Presentation of search results: the CL_S technique. *J. Phys. G*, 28:2693–2704, 2002.

[108] URL https://lhc-statistics.web.cern.ch/LHC-Statistics/.