

Top Physics at the CMS Detector at the LHC

Jeson Jacob

March 2015



A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science, School of Physics, March 2015.

Word Count: XX,XXX words

Abstract

This is my last 3.5 years.

Acknowledgements

The author would like to thank...

his badass homies.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed:.....

Date:.....

Contents

1	Introduction	1
2	The Standard Model	3
3	The CMS Detector at the LHC	5
3.1	Introduction to the LHC	5
3.2	Overview of CMS	7
3.3	Sub-Detectors	9
3.3.1	Tracker	9
3.3.2	Electromagnetic Calorimeter	11
3.3.3	Hadronic Calorimeter	13
3.3.4	Superconducting Magnet	14
3.3.5	Muon Chambers	15
3.3.6	Trigger and Data Acquisition	16
3.4	CMS Computing	17
3.5	Event Data Model	18
3.6	Upgrades	19
3.7	Object Reconstruction and Identification	19
3.7.1	Electron Reconstruction	20
3.7.2	Muon Reconstruction	21
3.7.3	Jet Reconstruction	22
3.7.4	Track Reconstruction	24
3.7.5	Pileup Subtraction	25
4	B Tagging Studies	27
4.1	Efficiency	28
4.2	Algorithm Comparison	29
5	7 TeV and 8 TeV Differential Cross Section Measurement	31
5.1	Introduction	31

5.2	Data and Simulated Samples	32
5.2.1	Data	32
5.2.2	Simulated Samples	32
5.3	Event Selection	35
A	Datasets in cross section analysis	41

List of Figures

2 The Standard Model

2.1	Relative branching ratios of the $t\bar{t}$ system	4
-----	--	---

3 The CMS Detector at the LHC

3.1	Map of LHC location (top) and schematic of LHC and experiments (bottom). CMS is located at Point 5 on the LHC ring.	6
3.2	Cumulative luminosity delivered to CMS over time during 2010, 2011 and 2012 proton-proton data taking periods.	7
3.3	Diagrammatic sectional view of the CMS detector [28]	8
3.4	Cross-sectional diagram of the CMS strip tracker; each line represents a detector module [29]	10
3.5	Diagram of the ECAL showing barrel supermodules, endcaps and preshower detectors. [13]	12
3.6	A longitudinal schematic of the HCAL showing the location of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [9]	13
3.7	Bore of the CMS solenoid magnet in vertical position in the CMS assembly hall, SX5, prior to installation [9]	14
3.8	Schematic representation of a quarter of the CMS muon system [24]	15

4 B Tagging Studies

4.1	Discriminator values produced by the Combined Secondary Vertex algorithm for all 4 jet flavours after normalisation	28
4.2	Non b jet efficiencies as a function of b jet efficiency for the CSV algorithm	29
4.3	UDS jet efficiencies as a function of b jet efficiencies for all algorithms	30

List of Tables

A Datasets in cross section analysis

A.1	7 TeV data sets by run period with the corresponding integrated luminosities (L_{int}) and run numbers.	41
A.2	8 TeV data sets by run period with the corresponding integrated luminosities (L_{int}) and run numbers.	41
A.3	8 TeV data sets by run period with the corresponding integrated luminosities (L_{int}) and run numbers.	41

Chapter 1

Introduction

Chapter 2

The Standard Model

The top quark was discovered by the CDF and DØ collaborations in 1995 and is still one of the less well studied fundamental particles in the Standard Model. The top quark is the heaviest fermion with its mass currently placed at $173.29 \pm 0.23(\text{stat.}) \pm 0.92(\text{syst.}) \text{ GeV}/c^2$ [18]. Since the lifetime of the top quark is very short, approximately $5 \times 10^{-25} \text{ s}$, it is the only one of the quarks to decay before it hadronises, meaning that the bare quark properties can be investigated. These unique properties of the top quark within the Standard Model mean it is an interesting focus of study.

During 2011 and 2012, the LHC produced millions of top quark pair events with gluon-gluon fusion ($\sim 70\%$) or quarkantiquark annihilation ($\sim 30\%$) TODO:feynmann diagram of production mechanisms being the primary production mechanisms. Top quarks decay almost 100% of the time to a W-boson and a b flavour jet. The W-boson then decays either hadronically (into two jets) or leptonically (lepton + neutrino). Top pair events are characterised by the decay of the W-bosons:

- Leptonic - both W-bosons decay to a lepton and a neutrino. The event would consist of 6 jets. (10.5%)
- Hadronic - both W-bosons decay to two jets. The event would consist of 2 jets, 2 leptons and 2 neutrinos (which would show up as E_T^{miss} in the event). (45.7%)
- Semi-Leptonic - one W-boson decays to a lepton and a neutrino, the other decays to two jets. The event would consist of 4 jets, 1 lepton and 1 neutrino. (43.8%)

The branching ratios for each decay mode are quoted in brackets [26], and are represented graphically in Figure 2.1.

Gluon-gluon fusion contributes more at the LHC as a result of the gluon momentum fraction

Top Pair Decay Channels

$\bar{c}s$		electron+jets	muon+jets	tau+jets	all-hadronic	
$\bar{u}d$						
$e^- \tau^+$		$e\tau$	$\mu\tau$	$\tau\tau$	tau+jets	
$\mu^- \mu^+$		$e\mu$	$\mu\mu$	$\mu\tau$	muon+jets	
$e^- e^+$		ee	$e\mu$	$e\tau$	electron+jets	
W decay	e^+	μ^+	τ^+		$u\bar{d}$	$c\bar{s}$

dileptons

Figure 2.1: Relative branching ratios of the $t\bar{t}$ system

increasing at a higher rate than that carried by the sea quarks which would be required to produce a top pair.

Chapter 3

The CMS Detector at the LHC

3.1 Introduction to the LHC

The Large Hadron Collider (LHC) at the Organisation Européenne pour la Recherche Nucléaire (CERN) near the Swiss city of Geneva was constructed with the aim of investigating various aspects of the Standard Model of physics. The Standard Model has stood up well to scientific scrutiny for several decades. However, areas of current interest such as the Higgs mechanism and the role electroweak symmetry breaking plays in it, and physics beyond the standard model such as supersymmetry (explained in further detail in Section 2), require the acceleration of particles to high energies (of the order of TeV). The Compact Muon Solenoid (CMS) general-purpose detector is one of four detectors located around the LHC, approximately 100 m below ground level. The geographical site and location of the various experiments are shown in Figure 3.1.

The 27 km circumference LHC can collide two proton beams, each composed of 2808 bunches, at a design energy of 7 TeV per beam (meaning a 14 TeV centre of mass energy in collisions) with a luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ at a collision rate of 25 ns (40 MHz). In fact, in 2011 the machine ran at a centre of mass energy of 7 TeV at a bunch spacing of 50 ns (leading to 8 proton-proton interactions per bunch crossing) and CMS recorded a peak instantaneous luminosity of $3.7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. In 2012 the centre of mass energy was increased to 8 TeV (21 proton-proton interactions per bunch crossing) and CMS recorded a peak instantaneous luminosity of $7.7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. The design energy and luminosities mentioned previously are planned when data taking begins again after Long Shutdown 1 (currently scheduled to be around May 2015) and after Long Shutdown 2 (in 2019) respectively. 6.1 fb^{-1} and 23.3 fb^{-1} of data was delivered to CMS during the 2011 and 2012 data taking periods respectively; Figure 3.2 shows the increase in integrated luminosities delivered to CMS over time.

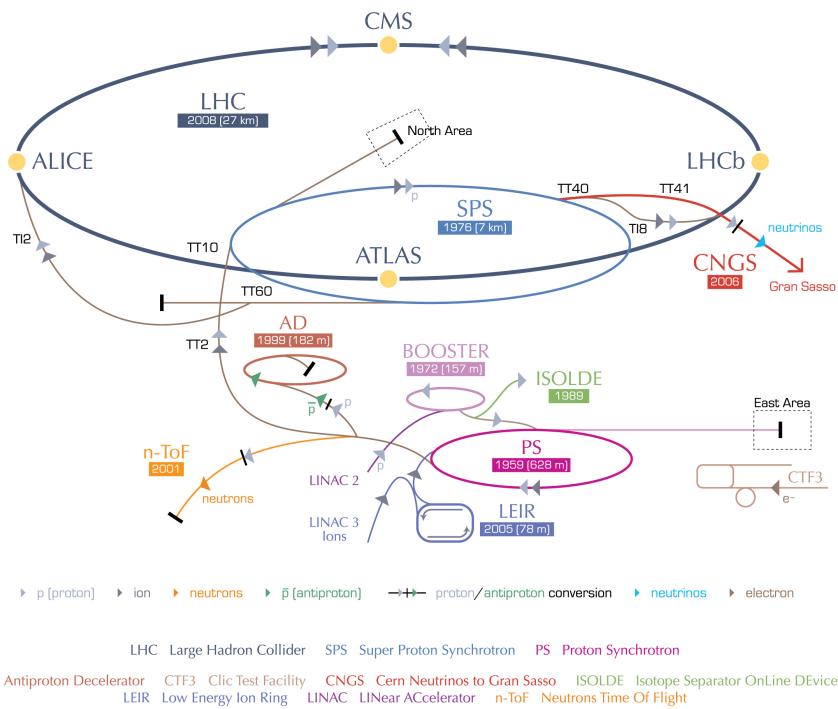
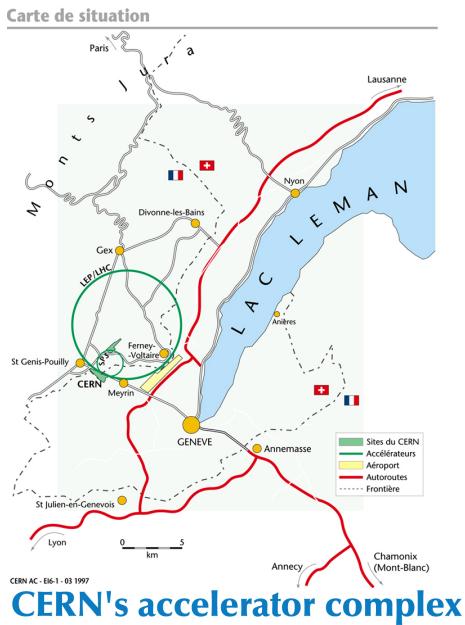


Figure 3.1: Map of LHC location (top) and schematic of LHC and experiments (bottom). CMS is located at Point 5 on the LHC ring.

The acceleration process begins with a tank of hydrogen. After stripping electrons off the hydrogen atoms using an electric field, the remaining protons are inserted into the LINAC2 which accelerates them up to 50 MeV. From here, the proton beam is injected into the Proton Synchrotron Booster (PSB) which increases the energy to 1.4 GeV, followed by the Proton Synchrotron (PS) which accelerates the beam to 25 GeV and the Super Proton Synchrotron (SPS) where the beam energy reaches 450 GeV. Finally, the protons are injected into the LHC where two beampipes carry the beams in a clockwise and an anti-clockwise direction while accelerating them up to the required collision energy. Filling the LHC rings takes approximately 4 minutes, followed by approximately 20 minutes until the beams reach energies of 4 TeV (during 2012 data taking). Once the collision energies have been attained, the counter-circulating beams are brought to collide at the four main LHC experiments CMS, ATLAS, LHCb and ALICE. ALICE is specifically designed to investigate heavy ion collisions (as opposed to protons), while LHCb investigates b-meson physics.

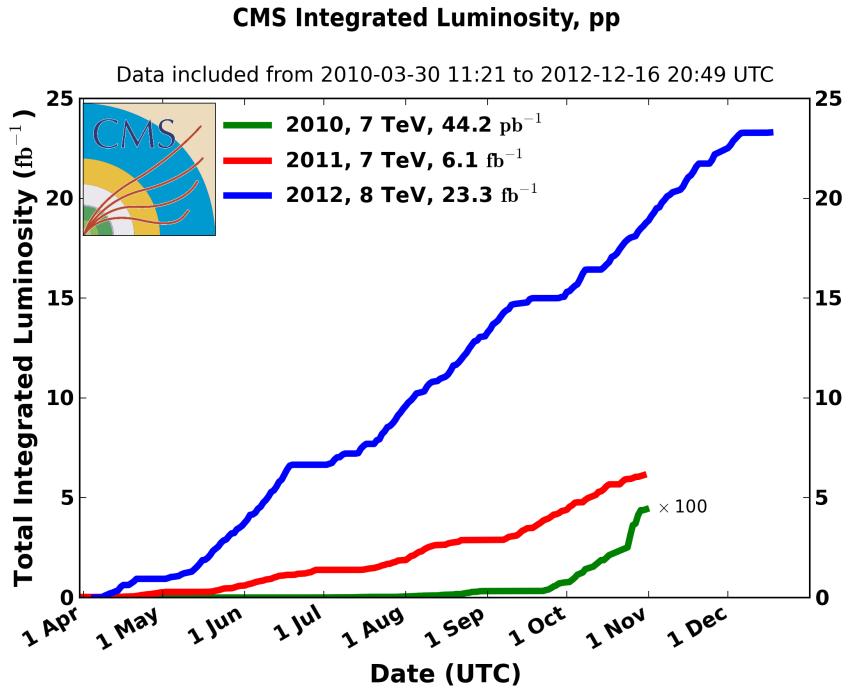


Figure 3.2: Cumulative luminosity delivered to CMS over time during 2010, 2011 and 2012 proton-proton data taking periods.

3.2 Overview of CMS

Compact Muon Solenoid is both the name of one of the detectors at the LHC, and of the collaboration of people worldwide who work to build, operate, maintain and upgrade it and to analyse the data recorded from it.

CMS is a general-purpose detector, designed to be efficient at detecting new physics with a wide range of signals. The detector measures 14.6 m in diameter, 21.6 m in length and weighs 12500 tonnes [9]. In general, the different components of CMS are arranged in concentric layers around the interaction point in the beam pipe, which is the point at which the two beams of proton bunches collide. As the products of any collisions that occur in the bunch crossing travel outwards from the interaction point, they pass through the various sub-detectors leaving signals as they do so. A subset of the information from some sub-detectors (the calorimeters and muon chambers, described in Section 3.3) is then processed by a trigger while information from other sub-detectors is buffered on the detector.

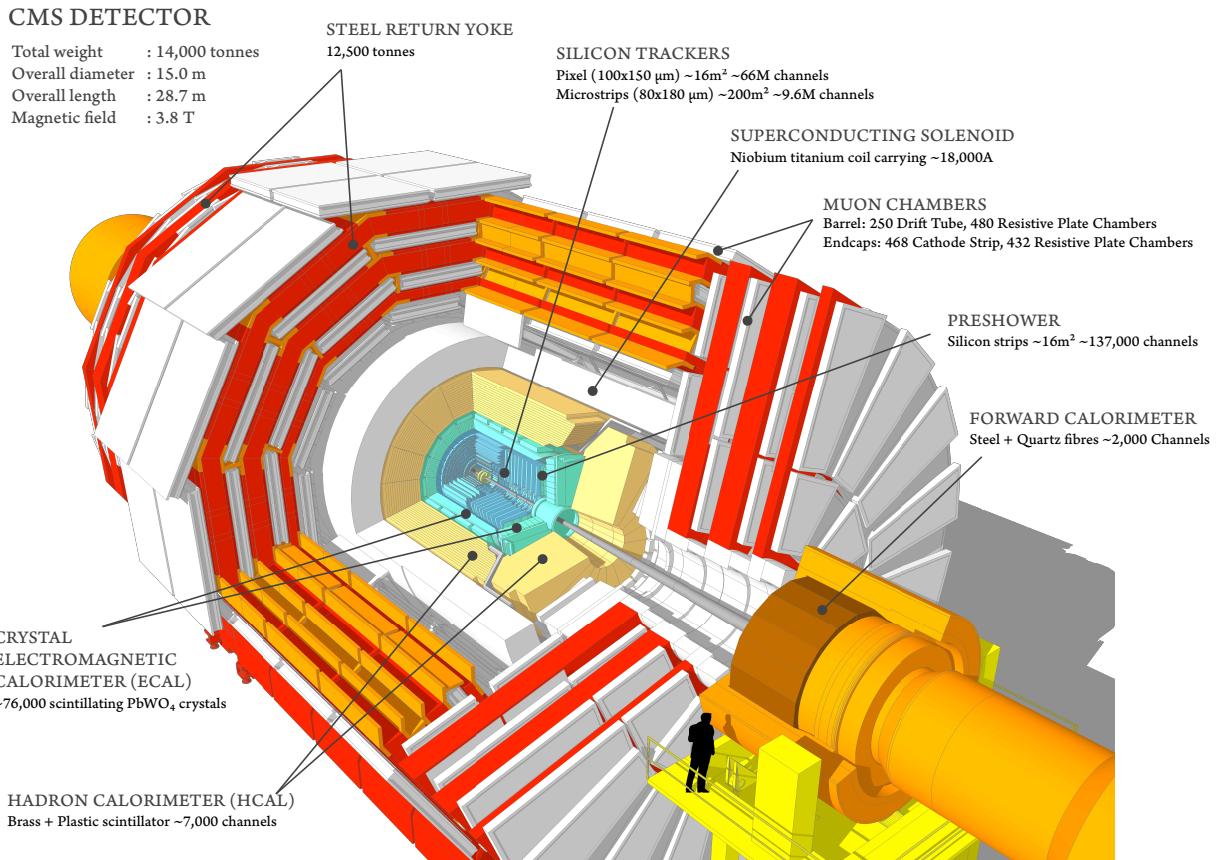


Figure 3.3: Diagrammatic sectional view of the CMS detector [28]

The design of CMS is based firstly around a superconducting solenoid magnet. Around and inside this magnet are the pixel and strip tracker, the electromagnetic calorimeter, the hadronic calorimeter, the return yoke for the magnet and the muon detectors. The coordinate system of the detector is designated with the origin as the intersection point of the two beams in the beam pipe, with the x-axis positive towards the centre of the LHC ring, with the y-axis positive vertically upwards and with the z-axis pointing along the beampipe in an anticlockwise direction. The angle ϕ is defined as the angle in the x-y plane, transverse to the beampipe starting from the x-axis. θ , the polar angle is defined as the angle in the z-y

plane starting from the z-axis (the beampipe). θ is related to the pseudorapidity, η , by the formula $\eta = -\ln(\tan(\theta/2))$, so that η ranges from 0 at $\theta = 90^\circ$ to ∞ at $\theta = 0^\circ$ [29]. The two ends of the detector, at high η values, are called the endcap regions while the central area at lower η values is called the barrel region.

3.3 Sub-Detectors

3.3.1 Tracker

Each event that is recorded in the CMS detector is one crossing of the bunches of protons of which the beams are comprised. Any charged particles produced as a result of the proton-proton collisions in these bunch crossings need to be recorded for later use in reconstruction of particle information such as trajectories and charges. This is achieved using the tracker section of CMS which consists of silicon sensors in two forms: pixels and strips. As the charged particles travel through these silicon sensors, electron-hole pairs are produced in the silicon which are processed by readout chips. This readout is carried out by the Analogue Pipeline Voltage (APV) chip (a new chip called the CMS Binary Chip (CBC) is intended to replace the APV in the HL-LHC after Long Shutdown 3, which is currently scheduled for 2023-2024).

The purpose of the tracker within the CMS detector is to track the trajectory of charged particles as they travel out from the interaction point. The tracker is split into two parts, with an inner silicon pixel detector and an outer silicon strip detector. The inner pixel detector consists of three layers in the barrel region at distances of 4.4 cm, 7.3 cm and 10.2 cm from the beam line with two endcap discs at each end. Each pixel measures $100 \mu\text{m}$ by $150 \mu\text{m}$ and the entire pixel detector covers an area of only approximately 1 m^2 but contains 66 million pixels in total.

The strip tracker is comprised of ten layers in the barrel region at distances ranging from 20 cm to 1.1 m from the beam pipe. The four inner layers (at distances of 26 cm, 34 cm, 42 cm and 50 cm from the beampipe) are collectively named the Tracker Inner Barrel (TIB), while the outer six layers (at distances of 61 cm, 70 cm, 78 cm, 87 cm, 97 cm and 108 cm from the beampipe) are known as the Tracker Outer Barrel (TOB). The endcaps consist of twelve discs, three of which correspond to the Tracker Inner Discs (TID) at z-distances from 75 cm to 100 cm, and nine to the Tracker End Caps (TEC) at z-distances between 120 cm and 280 cm [27]. A diagrammatic view of the CMS tracker, including both pixels and strips, is shown in Figure 3.4. The largest silicon detector ever constructed, the strip tracker contains about ten million channels and covers an area of approximately 200 m^2 [9, 29]. The

operating temperature of the tracker during 2011 and 2012 has been 0 °C for the pixels and +4 °C for the strips but will be lowered for both after Long Shutdown 1 [7].

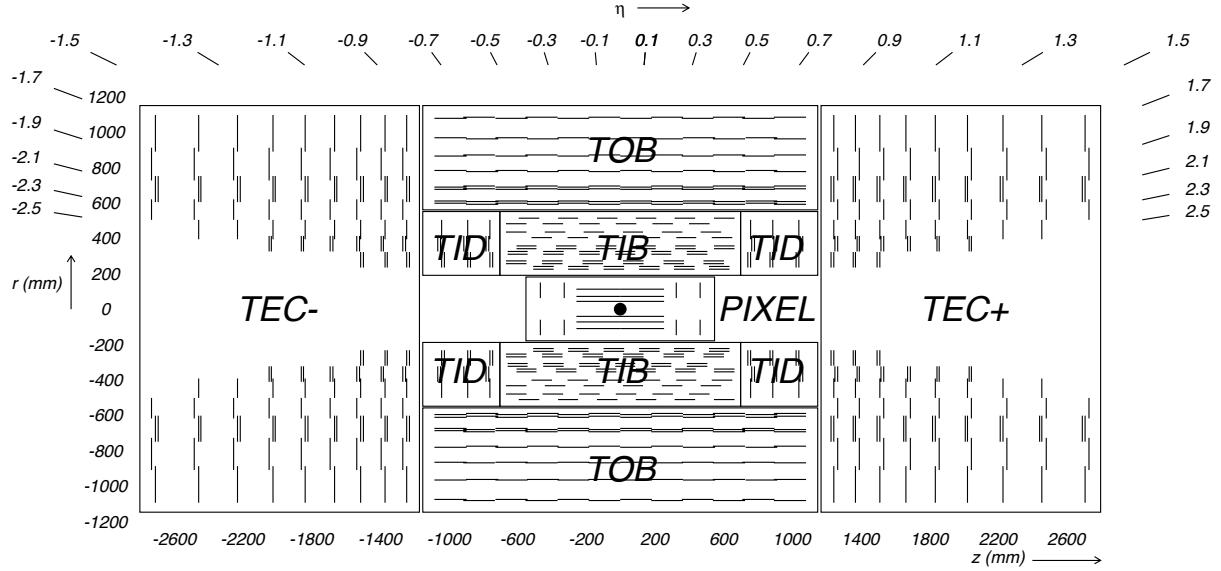


Figure 3.4: Cross-sectional diagram of the CMS strip tracker; each line represents a detector module [29]

As a charged particle passes through the various layers of silicon in these detectors, electrical signals are created which are then read by readout electronics. Offline algorithms then use the data relating to which strips and sensors received 'hits' to reconstruct the tracks of charged particles and this information is later used for particle reconstruction and identification [9, 25].

Neutral particles could also leave signals in the tracker in a small proportion of cases if the particle interacts with a silicon atom and/or converts to charged particles such as a photon conversion to two electrons. In the majority of cases, the presence of neutral particles must be detected from their energy deposits in the calorimeters. For physics analyses, the premise employed is that the curving of particle tracks within a magnetic field allows the calculation of the particle momentum and charge. An important requirement of the tracker is that it must allow the above whilst not affecting the energy of the particle being tracked. The silicon in the tracker sensors also does not stop particles, thereby allowing them to pass through this innermost layer of the detector and interact with outer layers.

The tracker readout chips must be able to distinguish from which bunch crossing a particle was produced (bunch-crossing identification).

The required tracker characteristics of high granularity, good resolution, fast response to process the high data rate without losing potentially valuable information, radiation hardness and keeping detector material to a minimum to minimise interactions of the particles with

detector matter, resulted in silicon being chosen to construct the tracker. One of the advantages of using silicon over gaseous detectors is that silicon produces more charge carriers per unit of length travelled by a particle. This therefore means the silicon thickness can be reduced, typically around $300 \mu\text{m}$ per layer. Silicon also allows fast signal transfer, of the order of about 10 ns, as the aforementioned charge carriers travel quickly through silicon. Furthermore, the spacing between strips in a silicon sensor, known as the strip pitch, can be manufactured to be extremely small, down to the order of around $10 \mu\text{m}$. The silicon sensors that make up the strip tracker in the CMS detector vary in thickness depending on their location. Sensors of $320 \mu\text{m}$ thickness are utilised in the TIB and TID, $500 \mu\text{m}$ in the TOB, and a mixture of both thicknesses is used in the TEC. The strip pitches and their lengths vary such that there are fifteen distinct types of sensor geometries in the strip tracker [10]. Thicker sensors are used in the outer sections of the strip tracker in order to maintain a high signal to noise ratio, since the increased strip lengths in this region lead to increased electronics noise due to higher capacitance [9]. Furthermore, in general the occupancy needs to be low (the fraction of channels with a hit in an event was desired to be 1 % or lower at the nominal LHC luminosity) so that pattern recognition can be carried out efficiently and so the amount of simultaneous data to be read out is manageable. In order to maintain a low occupancy, a high granularity is maintained in the strip tracker with strip lengths less than 20 cm and strip pitches less than $205 \mu\text{m}$ [27]. While high granularity is beneficial for high precision tracking, it also results in a high number of channels, which in turn requires large amounts of electronics for readout and leads to high heat load.

On a global scale, the tracker data is read out using a combination of interoperating systems. The data from the Analogue Pipeline Voltage (APV25) tracker readout chips is taken via optical fibres to 440 off-detector Front End Drivers (FEDs) (which make up 63% of the total number of FEDs in CMS). The FEDs then forward the data to the online Data Acquisition System (DAQ). In addition, off-detector Front End Controllers (FECs) control the front-end electronics by means of approximately 350 control rings, including triggers, clock and monitoring [9, 19].

3.3.2 Electromagnetic Calorimeter

The electromagnetic calorimeter, often referred to as the ECAL, is a hermetic, homogeneous subdetector constructed of scintillating crystals of lead tungstate (PbWO_4), and is the next layer outside the tracker. Any electromagnetic particles such as electrons or photons are absorbed by the ECAL crystals. The energy of the particles is absorbed by the crystals which then scintillate emitting a blue-green coloured light. These signals are then detected by connected photodetectors and processed by readout electronics.

The high density crystals used (8.28 g/cm^3), short radiation length ($X_0 = 0.89\text{cm}$) and small Molière radius of 2.2 cm (EXPLAIN THIS? OR REMOVE IT?) lead to a compact calorimeter with fast response time, high granularity and of course capable of withstanding the radiation levels within CMS. The barrel region of the ECAL consists of 61,200 crystals and extends up to a pseudorapidity, $\eta \pm 1.479$. Each individual barrel crystal is $25.8X_0$ thick and has cross-sectional dimensions of $22 \times 22 \text{ mm}^2$, which equates to 0.0174×0.0174 in the $\eta - \phi$ plane. These crystals are divided into 36 supermodules, with each consisting of 4 modules. The endcaps are comprised of 7,324 crystals, cover the range $1.479 \leq \eta \leq 3.0$ and are split into two halves known as “Dees” and is divided into operating segments of 40° each. The endcap crystals have slightly larger dimensions, having a thickness of $24.7X_0$ and cross-sectional dimensions of $28.62 \times 28.62 \text{ mm}^2$ [9, 6]. Figure 3.5 shows the layout of the ECAL.

The signals in the crystals are collected by avalanche photodiodes in the barrel region and vacuum phototriodes in the endcaps. The numbers of photoelectrons produced is dependent on temperature, with increasing temperature resulting in a decrease in the number of electrons at a rate of $-3.8 \pm 0.4\%/\text{ }^\circ\text{C}$. A cooling system is thus employed which maintains a stable operating temperature of the ECAL system to within $\pm 0.05\text{ }^\circ\text{C}$, with a nominal operating temperature of $18\text{ }^\circ\text{C}$ [9]. The energy resolution of the ECAL has been shown to follow $\sigma_E/E = 2.8\%/\sqrt{E} \oplus 12\%/E \oplus 0.3\%$ where the three constant terms come from stochastic fluctuations such as photostatistics, electronics noise and temperature stability and calibration uncertainties [13].

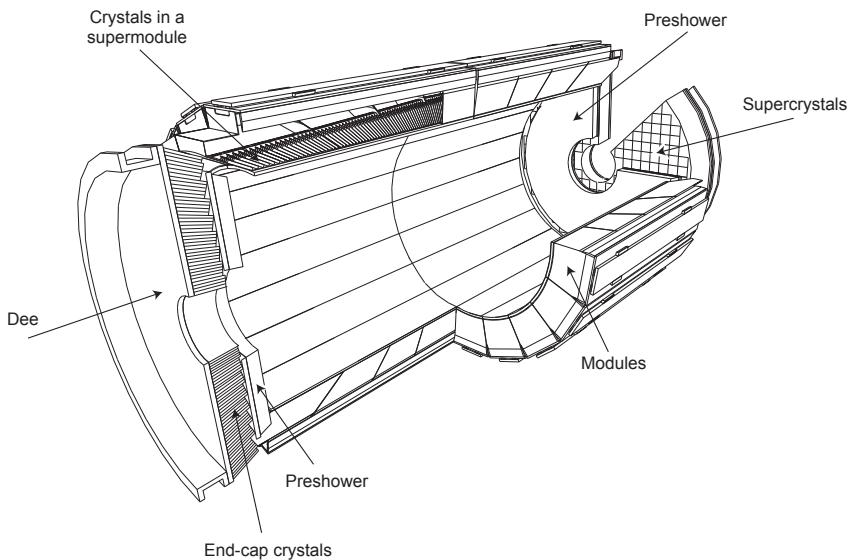


Figure 3.5: Diagram of the ECAL showing barrel supermodules, endcaps and preshower detectors. [13]

3.3.3 Hadronic Calorimeter

The next subdetector outside the ECAL is the hadronic calorimeter (HCAL) which works in a similar way to the ECAL, in this case by absorbing hadron jets. The HCAL is located at a distance of 1.77 m to 2.95 m from the beamline, with an additional outer hadronic calorimeter also installed outside the magnet due to radial restrictions imposed by the magnet coil. The HCAL barrel (HB), outer (HO) and endcaps (HE) cover the η range up to 3.0, with the forward HCAL (HF) placed at η up to 5.2 increasing the coverage. Figure 3.6 shows the a quarter of the HCAL endcap.

The HB The HCAL is a sampling calorimeter, meaning it is composed of alternating absorber layers (made of brass) and scintillator layers (made of plastic). A hadronic particle produces a shower of secondary particles when it strikes an absorber layer; this process is repeated as these secondary particles themselves pass through successive absorber layers. The scintillator layers in between absorb the energy of these particles, emitting light as they do which is fed to readout electronics via optical fibres to measure the particle energies.

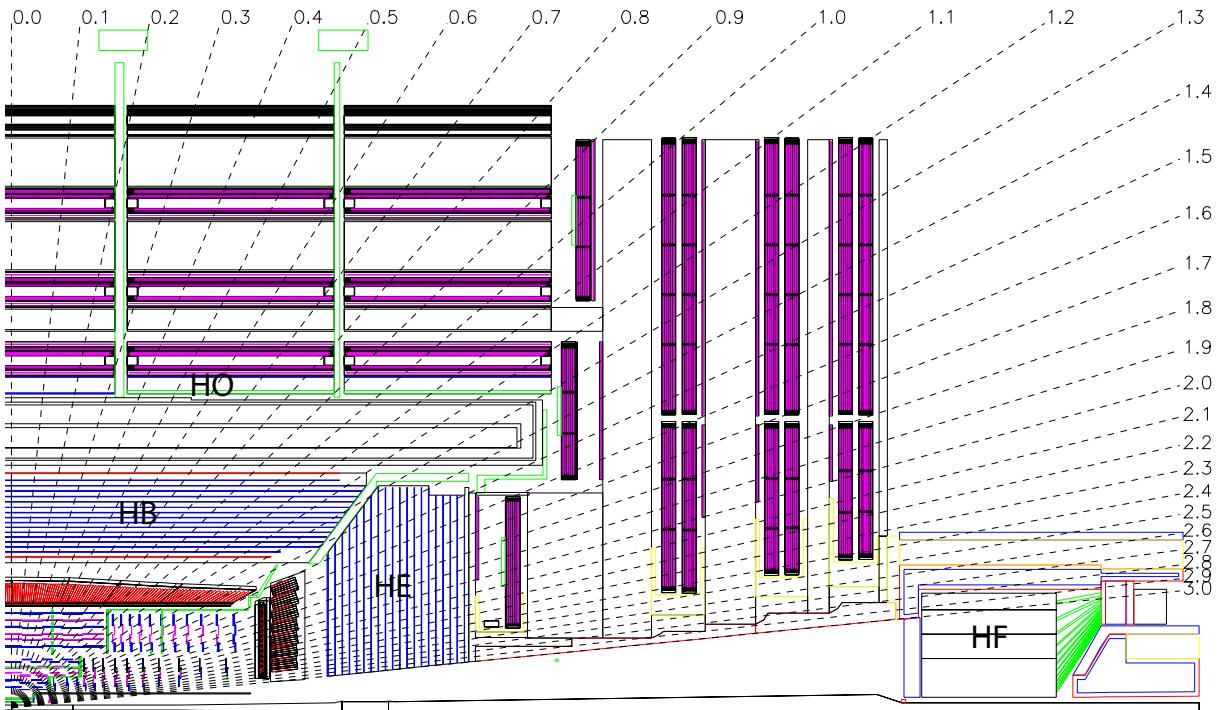


Figure 3.6: A longitudinal schematic of the HCAL showing the location of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [9]

3.3.4 Superconducting Magnet

The magnet of the CMS detector is the largest and highest field strength superconducting solenoid ever constructed. It can produce a magnetic field up to 4 T (a field strength of 3.8 T is used in running) with a stored energy of 2.7 GJ [29]. The cold bore, cooled to 4.5 K using liquid helium, has dimensions of 12.5 m length and 6 m diameter, and weighs 220 tonnes [20]. Due to design constraints and structural requirements, the magnet itself provides some of the structural strength both to support itself and to withstand its own magnetic bursting force on the coil [9].



Figure 3.7: Bore of the CMS solenoid magnet in vertical position in the CMS assembly hall, SX5, prior to installation [9]

The cold bore, pictured in Figure 3.7 in the CMS assembly hall, is enclosed in a steel return yoke weighing 10,000 tonnes with the aim of containing and returning the magnetic field. The yoke is interspersed in the muon chambers and is made up of 5 barrel segments and 6 endcap discs. The magnet was designed in a manner such as to facilitate assembly at ground level prior to lowering into the CMS experimental cavern, UX5. Despite the challenges involved in constructing such a powerful magnet, the high track-curving power created by the high magnetic field was desirable in order to provide good momentum resolution of the tracking components.

3.3.5 Muon Chambers

Muons pass through all the previous inner sub-detectors, losing very little energy as they traverse them (around 1 MeV/mm), leading to the muon chambers being located outermost in the detector. There are, in fact, three types of muon detectors in use in CMS. The endcaps contain Cathode Strip Chambers (CSCs), the central barrel regions contain Drift Tubes (DTs), and both regions are reinforced with Resistive Plate Chambers (RPCs). These systems, in combination with the silicon tracker, are used to determine the momentum of muons by taking advantage of their curved tracks due to the magnetic field. The different technologies are used due to the different numbers of particles expected in different areas of the detector and because of technological considerations regarding the physical areas to be covered [29]. Figure 3.8 shows a diagrammatic representation of one quarter of the CMS muon detectors. The CSCs are used in the endcaps to cover $|\eta|$ values between 1.2 and 2.4 that experience high muon rates and where the magnetic field of the solenoid is high [29]. They take the form of four disc layers each made of 2 (inner layer) or 3 (outer 3 layers) concentric rings. They consist of volumes of gas in which are found positively charged wires placed at right angles to negatively charged copper strips. To relate these to the name given to these detectors, the positive wires are the anodes and the negative strips are the cathodes. As a charged particle passes through the gas it ionises gas atoms and the electrons that are knocked out travel towards the anode wires. At the same time the resulting positively charged ions in the gas travel towards the cathode strips. Since the wires and strips are at right angles to each other, the CSCs provide two position co-ordinates for the passing muon. The CSC detection mechanism is a fast process so their signals are used for muon triggering [9].

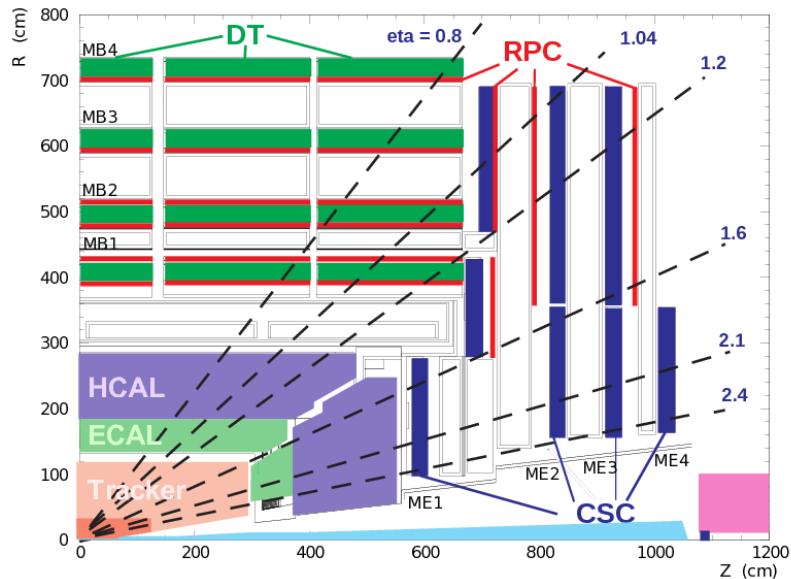


Figure 3.8: Schematic representation of a quarter of the CMS muon system [24]

The barrel region DTs cover $|\eta|$ values of less than 1.2, that encounter low rates of low transverse momentum muons [29]. They are arranged in four cylindrical layers among the layers of the magnet's iron return yoke and RPC layers. The layers are slightly staggered to ensure that even a muon of high transverse momentum will be detected by at least three of the four layers. The drift tubes are gas tubes containing a stretched wire and enclosed by aluminium plates. As a particle with a charge traverses the volume of gas, atoms in the gas are ionised and the resulting free electrons travel along electric field lines towards the positively charged wire. By using the position of the electron along the wire and the time taken for the wire to detect the electron, two coordinates of the muon position can be deduced. In comparison to the DTs and CSCs, the RPCs that complement them provide a better resolution in terms of time (1ns) but a worse resolution in terms of position, and so provide additional trigger information. As their name suggests, they are constructed of plates, one negatively charged (cathode) and one positively charged (anode) and made of a plastic with high resistance. In a similar process to the other muon detectors, a gas that makes up the volume in between the plates is ionised by a passing charged muon. The resulting electrons create an avalanche of electrons by, in turn, ionising other gas atoms; this avalanche of electrons moves towards the anode and metal readout strips outside the plastic anode detect the signal for readout. The hit strips pattern allows a calculation of the muon momentum which is then fed to trigger algorithms.

Neutrinos, like muons, also pass through all the layers of the detector, but cannot be detected directly and their presence must therefore be deduced by the presence of missing energy in the transverse plane (perpendicular to the beam directions).

3.3.6 Trigger and Data Acquisition

The data acquisition system (DAQ) is the name given to the combination of the trigger and data recording systems. At the design bunch crossing frequency of 40 MHz and at the design luminosity of $10^{34} \text{ cm} - 2\text{s} - 1$, the expected number of proton-proton collisions per second is approximately 10^9 , translating to 21 proton-proton collisions per bunch crossing on average. Each bunch crossing produces approximately 1 MB of data, leading to 40,000 GB/s of data. This extremely large amount of data, which is impractical to record, means that a trigger system is required to filter the events in order to record only data that are of interest for physics analyses.

There are two levels to the CMS trigger, named Level 1 Trigger (L1) and High Level Trigger (HLT). These work in sequence with the electronics and readout systems in place in the detector to filter the events to a practically manageable amount for processing. In combination, these triggers reduce the data rate by at least a factor of 10^6 [9].

The L1 Trigger is a pipelined deadtimeless system comprised of calorimeter triggers, muon triggers and a global trigger that works on a combination of data from the former two. The decisions made by the L1 Trigger are carried out by ‘custom hardware processors’, such as Field Programmable Gate Arrays (FPGAs). The data from a bunch crossing are held in pipeline buffers in the electronics at the front end whilst the information for the triggers is processed in the CMS service cavern (USC) and the decisions transmitted back to the front end. The maximum time allowed for this process for each event is $3.2 \mu\text{s}$, which is the time taken for the signals to be transmitted via optical fibres from the detector to the processors and back again, known as the latency [29]. The L1 Trigger outputs data to the HLT at a rate of 100 kHz [9].

As already mentioned, the trigger uses calorimeter and muon chamber data to reach a decision on an event within the required timeframe. Tracker data are not currently used since track reconstruction exceeds the amount of time allowed for the L1 trigger decision. Good trigger performance is related to the quality of the calorimeters and muon systems. Factors such as good momentum resolution of high momentum muons, good charge determination of muons, good ECAL energy resolution and good missing transverse energy resolution are required for the trigger to select interesting events. The trigger was designed as described in order to allow the CMS experiment to meet the goals of the LHC physics programme [9].

Unlike the Level 1 Trigger, the High Level Trigger is software-based and run offline in separate processor farms [29]. If the Level 1 produces an accept decision, the data which was stored in buffers in the front end electronics is read into readout buffers from where the DAQ system accesses it. The L1 output rate of 100 kHz corresponds to a data rate of approximately 100 GB/s. The software making up the HLT processes these events in a computer farm that carries out fast processing of offline algorithms such as selections and object reconstructions to reduce the rate down to approximately 100 Hz. This accepted data are then stored on tape.

All sub-detector readout systems consist of front-end systems into which data is fed from the detector. When the L1 trigger accepts an event, the data corresponding to that event is extracted from these front-end buffers and front-end drivers (FEDs) enters the DAQ system. A more detailed description of the DAQ system can be found in [9] and [29].

3.4 CMS Computing

The CMS offline computing takes on the workload of transferring accepted data from the triggers to both permanent and temporary storage, and the processing of this data for

subsequent analysis, in addition to the production of simulated CMS data. The resources needed to process the high volumes of data involved require a distributed computing system; the Worldwide LHC Computing Grid (WLCG) infrastructure, an international collaboration of LHC experiments and computing centres, is employed to carried out these tasks.

The CMS computing resources are primarily divided into three tiers. Tier 0 (T0) consists of only one site, at CERN itself. The T0 centre has as its main aims to take accepted data from the detector and transfer it to permanent storage on tape. Tier 0 computing is also responsible for reconstructing the RAW data into RECO data formats (see Section 3.7 for more details on data formats). From the T0 centre, the copies of the data in RECO and RAW format is transferred to T1 centres around the world (there is also a Tier 1 (T1) centre at CERN, known as the CERN computing centre). Owing to its important role in ensuring the robust transfer of RAW and RECO data, the T0 resources are not available for analysis by CMS users. There are seven Tier 1 centres located in various countries within the CMS collaboration. In the UK, the T1 centre is located at the Rutherford Appleton Laboratory in Harwell, near Oxford. T1 centres provide reliable computing resources for data storage and processing; RAW data is spread between them, providing a second copy of the RAW data stored at CERN. The second reconstruction step is also carried out at T1 centres, in addition to the production of simulated data. These can then be provided to any of the Tier 2 centres at CMS institutes (typically universities) where they may be temporarily stored. Typically, T2 centres are used to run users' final analyses and produce simulations [9, 29].

3.5 Event Data Model

Reconstructed data from CMS uses a data model based around an event, called the Event Data Model (EDM), where one event is one crossing of proton bunches at the centre of CMS which passes the triggers. This model is created and manipulated within a CMS software framework (- ADD REFERENCE?) written in C++, with the event and related objects in the object oriented data analysis framework ROOT.

Everything in an event is passed through a sequence of framework modules to produce subsequent versions of the data. The arrangement of these data takes the form of layers, with the first of these being RAW. This level contains the full information from the event in CMS and occupies approximately 1.5-2 MB/event. More information than is necessary for user analyses is included at this level, and so the majority of CMS users will not use this data format. Reconstructed level data (RECO) is slightly smaller in size (approximately 0.5 MB/event) and is essentially a compressed subset of the RAW data after modules performing reconstruction have been run. These modules produce physics objects (like

electrons and jets) using algorithms to reconstruct tracks in the silicon tracker, clusters of deposits in the calorimeters, primary and secondary vertices, determine particle identification and to correct for detector characteristics such as non-functioning components.

The third level, Analysis Object Data (AOD), is the smallest of the data formats, requiring approximately 100 kB/event, which is small enough to allow the entire AOD data to be stored at computing centres worldwide. AOD format is a subset of the RECO data, and is produced by performing a skim on RECO leaving only high level physics objects (e.g. electrons, jets) which is adequate for most physics analyses.

In the different cross section analysis presented in this thesis, as in the majority of CMS physics analyses, this AOD data is processed using the Bristol Top Group’s NTupleProduction code (REFERENCE TO NTT CODE ON GITHUB) to produce private ntuples which are yet again smaller in size, at approximately 3 kB/event. These ntuples are then converted to simple ROOT histogram files after applying the required selections and corrections in the BristolAnalysisTools (REFERENCE TO BAT CODE ON GITHUB). Scripts written in Python in DailyPythonScripts (REFERENCE TO DPS CODE ON GITHUB) are then used to produce final plots and tables.

3.6 Upgrades

It is worth noting at this point, that the CMS experiment, along with the LHC and the other detectors are in a long term programme of upgrades and maintenance. By 2023 the luminosity provided by the LHC is expected to be $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [16]. The various runs and shutdowns until 2023 are collectively referred to as Phase 1. In phase 2, or after 2023, long shutdown 3 is also planned to bring further improvements and upgrades to the performance of the LHC, after which the luminosity of the LHC is expected to reach $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The machine in this state will be known as the high luminosity LHC, HL-LHC (also sometimes referred to as Super LHC, SLHC).

Naturally, the above dates and schedules are the latest best estimates and are liable to change over time, particularly those further in the future, as work progresses.

3.7 Object Reconstruction and Identification

The process of producing a physics object such as an electron, photon or jet, from the data recorded by CMS is known as reconstruction and is carried out by “modules” known as

EDProducers within CMSSW. The three step process of reconstructing high level objects consists of local reconstruction within a sub detector, global reconstruction using data from the whole CMS detector, and a final stage combining reconstructed objects from both of these. The reconstruction technique used in the majority of CMS analyses is called Particle Flow (PF) [1]. PF uses information from all of the sub-detectors of CMS to identify and reconstruct particles produced from a proton-proton collision.

3.7.1 Electron Reconstruction

ECAL local reconstruction algorithms calculate the time of arrival, position and the energy deposited by electromagnetic objects. After grouping together deposits in neighbouring crystals to form clusters, deposits are then matched to deposits in the HCAL, forming a Calo Tower. Electrons are completely stopped in the ECAL and deposit their energy in a narrow cluster of crystals.

However, electrons can interact with the material between the interaction point and the ECAL, emitting a photon via bremsstrahlung radiation. Similarly, photons can convert to e^+e^- . Both of these processes result in ECAL deposits with a larger spread in ϕ because of the strong magnetic field in the inner section of CMS with the tracker. In the case of photons, several clusters are grouped together to form superclusters, which are then corrected for their energies to obtain the energies of the original photon [23].

Seeds for electrons are produced in the ECAL by two methods. The first matches superclusters with a trajectory compatible with two or three pixel detector hits and the interaction point. The second matches the supercluster to tracker tracks to identify electrons (and in the case of electrons emitting bremsstrahlung radiation, tracks with a low number of hits) [2]. Combining the seeds from the two methods, a Gaussian-Sum Filter, a generalisation of the Kalman Filter algorithm, is used to reconstruct electron paths [5].

Since other objects can leave similar signatures in the detector to electrons, such as jets or electrons from photon conversions, candidates are required to satisfy additional requirements of identification and isolation. Several electron identification methods exists and are used in CMS analyses. The main top cross sections analysis in this thesis uses the multivariate identification (MVA ID). As the name suggests, this approach uses a multivariate analysis, with track, vertex and supercluster variables as input, to produce a discriminator value, with higher values indicating a higher likelihood for a candidate to be a real electron. The MVA ID algorithm is optimised for identifying electrons from W and Z boson decays [2].

The isolation of an electron is defined as the activity within a cone surrounding the electron. Isolation is used in parallel with identification to select electrons, in particular to distinguish

electrons promptly produced in a proton-proton collision. Such isolated electrons would have less activity in its vicinity than electrons from within a jet, which could originate from leptonically decaying b hadrons, and jets faking electrons. Two methods exist in CMS of calculating the isolation of a particle: detector based isolation and particle based isolation. The detector based method is defined in each sub detector as the sum of the momenta or energies in a cone of $\Delta R = 0.3$ around the electron. The particle based method uses the total transverse energy of PF reconstructed particles within a cone of $\Delta R = 0.3$ and can remove activity coming from collisions other than the interaction point. By normalising this isolation to the momentum or energy of the electron, a relative isolation is obtained, relating the cone activity to the electron. This relative isolation is the variable used in the top cross sections analysis to select electrons.

In order to avoid the selection of electrons originating from a photon conversion, a veto can be placed on a second electron in the event. However, since the two electrons in a photon conversion may not necessarily have equal transverse momentum, such a veto may be insufficient, and so further techniques to identify conversion events are used. Firstly, since an electron from a conversion would be produced at some distance from the interaction point and in the detector material, eliminating candidates with missing hits in the pixel tracker helps to distinguish such electrons from promptly produced electrons. In events in which the conversion occurs in the beam pipe or if the electron is matched to unassociated pixel hits, this method can also be insufficient, so an additional track matching step is used. Tracks are matched in pairs and following geometrical cuts, can be removed if they appear to originate from a conversion [2].

3.7.2 Muon Reconstruction

Local reconstruction in the muon chambers provide hit position and time of arrival of a muon. This information from the DTs and CSCs is then amalgamated to create muon track hits and segments. These are used by the muon global reconstruction algorithms to reconstruct “standalone” muons. An inner detector segment is used as a seed for a Kalman Filter (EXPLAIN THIS?) [22] and possible trajectories are generated. By removing hits which are not likely to have come from the track in question, the likely trajectory is constructed layer-by-layer. A final fit is carried out, including an extrapolation to the interaction point for greater momentum resolution.

The magnetic field in the muon system is only 2 T compared to 3.8 T in the tracker. Tracks reconstructed in the tracker can therefore be combined with the aforementioned muon chamber information to improve the p_T resolution of muons, as seen in Figure

Two methods are employed to combine the information from the two sub-detectors. *Global muon reconstruction* matches a tracker track to a standalone muon track and carries out a fit of the resulting *global muon* track. The second method, *tracker muon reconstruction*, tracker tracks are extrapolated outwards to the muon chambers and accepted as a muon candidate if a DT or CSC matching track is found [12].

In terms of triggering, the p_T of a muon is first estimated using the information available at Level 1 trigger stage from all three types of muon detectors. At HLT level, the muon candidates from Level 1 are further refined using track finding and fitting, but still using only information from muon chambers, leading to Level 2 muons. As mentioned in Section 3.3.6, due to the time constraints required of the trigger, full tracker data are not currently used. However, tracks of Level 2 muons are extrapolated into the tracker systems and a localised track finding algorithm is run to identify only nearby tracker hits. A track matching that of a Level 2 muon leads to a Level 3 muon.

3.7.3 Jet Reconstruction

As quarks produced in proton-proton interactions (except the top quark) hadronise, they form jets of particles in the direction of travel of the quark from which it originates. The time of arrival, position and the energy deposited by hadronic objects are locally reconstructed in the HCAL. If the deposit matches an ECAL deposit, a Calo Tower is formed for later use in jet reconstruction algorithms.

The particle flow (PF) algorithm performing the reconstruction of particles in the jet, and the $anti - k_t$ algorithm used to perform the clustering of these particles into jets. The $anti - k_t$ algorithm, explained in detail in [8] is one of several jet algorithms that exist in CMS to combine reconstructed particles into jets. It defines a distance d_{ik} between reconstructed particles as

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right) \frac{(y_i - y_j) + (\phi_i - \phi_j)^2}{R^2} \quad (3.1)$$

$p_{T,i}$ and $p_{T,j}$ are the transverse momenta of the two particles i and j , y_i and y_j are the rapidities and ϕ_i and ϕ_j are the azimuth angles and R is the radius of the jet cone. $anti-k_t$ iteratively clusters together particles with the smallest $d_{i,j}$ between them until all jets are reconstructed and there are no particles remaining. Events will usually consist of a small number of high- p_T (hard) particles and a large number of low- p_T particles. The distance between hard particles is typically small, and the distance between softer particles is larger. Soft (low- p_T) particles tending to cluster around hard (high- p_T) particles, before clustering

with other soft particles.

While the PF anti- k_t jets show high jet matching efficiency on Monte Carlo simulation samples, corrections are applied based on the jet p_T and η to correct for mismeasurements in the detector and thereby improve the agreement between generated and reconstructed particle flow jets. The factored approach to CMS jet energy corrections is comprised of three parts:

- L1 Pile Up: corrects for additional energy from charged particles from pile-up in the event 3.7.5.
- L2 Relative Jet Correction: corrects the reconstructed energy to match the generated jet with respect to η .
- L3 Absolute Jet Correction: corrects the reconstructed energy to match the generated jet with respect to p_T .
- L2L3Residuals: reduces any residual differences between the reconstructed and generated jet due to simulation not being perfectly tuned to data. This correction is applied to data only.

In order to identify jets in the differential cross sections analysis, further identification criteria are used to reduce electronic noise, to reduce the number of electrons mis-identified as jets and, in so doing, to ensure the selection of high quality jets. The requirements, which are known as the losse PF Jet ID, are:

- at least one constituent particle
- the neutral hadron energy fraction (NHF) must be < 0.99
- for jets with abseta $|> 2.4$, the charged hadron energy fraction (CHF) must be > 0
- the neutral electromagnetic energy fraction (NEF) must be < 0.99
- for jets with abseta $|> 2.4$, the charged electromagnetic energy fraction (CEF) must be < 0.99
- for jets with abseta $|> 2.4$, the number of charged hadronic constituents (NCH) must be > 0

Jet energy resolution corrections are also applied (known as jet smearing) to account to the fact that jet energy resolution has been measured to be worse in data than in simulation. These PF jet corrections are obtained using 0.8fb^{-1} of 2011 dijet data and 19.7fb^{-1} of 2012 dijet data [11, 17].

B Jets

The process of identifying jets coming from b-quarks is known as b-tagging and is very important in top quark physics due to the decay of the top to a W-boson and a b-quark. B-tagging can therefore help to appreciably reduce background processes in signal. A description of the several algorithms available in CMS is given in 4. The algorithm used in the differential cross sections analysis is the combined secondary vertex tagger (CSV). The relatively long lifetime of b quarks means that they travel a significant distance (of the order of a few cm) before decaying. This leads to events with b-jets possessing a secondary vertex distinguishable at a significant distance from the primary interaction vertex. This secondary vertex information, together with track based information such as the number of tracks in the jet, the invariant mass of the secondary vertex and the impact parameter significance of each jet track is used by the CSV algorithm to produce a discriminator. TODO: EXPLAIN IMPACT PARAMETER SIGNIFICANCE WITH A DIAGRAM Three working points are used in CMS: tight, medium and loose. The medium working point is used in the different cross sections analysis, which carries a 1 % mis-tag rate and approximately 70 % b-tag efficiency. The tight and loose working points have mis-tag rates of 0.1 % and 10 % respectively [14].

3.7.4 Track Reconstruction

Algorithms performing local reconstruction scan tracker modules in each layer with higher than a threshold signal and constructing clusters by adding adjacent strips or pixels to the original seed strip or pixel. In order to reconstruct complete tracks to obtain the position and momentum of the charged particle, algorithms based on specific requirements such as high or low transverse momentum tracks are used. These algorithms in CMS are collectively known as the Combinatorial Track Finder (CTF).

Multiple passes of the CTF reconstruction software are carried out to reconstruct tracks, in a process called iterative tracking. Earlier iterations identify tracks that are easy to find such as high p_T originating near the interaction point. As these tracks are reconstructed, the corresponding hits are removed from consideration in subsequent iterations, making it simpler for later iterations to identify tracks that are more difficult to find such as those of displaced particles or with low p_T .

Six iterations are carried out in total, and each iteration can be split into four steps. Seeds are created using 2 or 3 hits to produce initial track candidates. The seed gives an estimate of the trajectories of the potential track candidates. A Kalman Filter (EXPLAIN THIS?) [22] based track finding algorithm then looks for further hits along an extrapolated path of

the seed trajectory. A track fitter is then run using information from the previous steps to produce final values for trajectory parameters. The fourth and final step then rejects tracks which fail specified quality checks [15].

3.7.5 Pileup Subtraction

When reconstructing an event in CMS, all vertices in the event must be reconstructed. By ordering the vertices in by the sum of the transverse momenta of their tracks, it is possible to identify the vertex of interest for physics analyses, known as the primary vertex (PV) as the vertex with the largest transverse momentum. The particle flow algorithm reconstructs objects, starting with those coming from the primary vertex, followed by other vertices (known as pileup). The reconstructed objects from the PV can be affected by the number of other vertices present in the event. For example, the jet momentum and lepton isolation could both increase with high pileup. This can, in turn, lead to signal events not passing selection requirements because a truly isolated lepton from the PV may appear not to be isolated. In addition, a larger number of events from background processes may pass selection requirements due to low energy jets appearing to have a higher energy. Hence, pileup subtraction, the removal of charged particles coming from vertices other than the PV is implemented to reduce these effects.

Neutral particles, however, pose a more difficult problem since they leave no tracker information for the reconstruction algorithms to easily identify their origin. One method of removing such particles from an event known as the $\Delta\beta$ correction, uses the fact that the estimated average energy in an event from neutral particles is half that from charged particles. Thus, it can be estimated that 0.5 times the charged particle energy comes from neutral particles. The second method, known as ρ correction, subtracts an average transverse momentum coming from pileup per unit area. While the two methods produce similar results, the ρ correction is used to correct the electron isolation and the $\Delta\beta$ correction is used to correct the muon isolation in the differential cross sections analysis presented in this thesis, as recommended by the CMS TOP physics analysis group.

Chapter 4

B Tagging Studies

In order to study $t\bar{t}$ events, event selections often use the reconstruction of jets as a key part of the event signature. More specifically, b tagging is used to identify those jets which come from a b quark. There are several algorithms which carry out b tagging in CMS: Combined Secondary Vertex, CombinedSecondaryVertex, CombinedSecondaryVetexMVA, JetBProbability, JetProbability, SimpleSecondaryVertex (High Efficiency), SimpleSecondaryVertex (High Purity), SoftMuon, SoftMuonByPt, SoftMuonByIP3d, TrackCounting (High Efficiency), TrackCounting (High Purity). These algorithms perform analyses on the jet in question and produce a discriminator output (simply a number). In all cases, a more positive discriminator value indicates a jet that is more likely to be a b flavour jet. The performance of the different algorithms vary depending on which jet characteristics are used to calculate the discriminator.

The current CMS recommendation is to use the Combined Secondary Vertex for physics analyses. This algorithm reconstructs the event vertices using the Trimmed Kalman Vertex Finder (MORE INFO ABOUT THIS, SEE PRESENTATION) and applies cuts to these vertices in order to find a secondary vertex, the b decay vertex. Depending on how many candidate secondary vertices are found, different combinations of variables such as impact parameter, flight distance significance between primary and secondary vertices, kinematics and secondary, kinematics and secondary vertex information are used to calculate a discriminator value [30].

The JProbability algorithms take all tracks, calculate the signed impact parameter significance and calculate the discriminator based on the negative log of the confidence level that all tracks are from the primary vertex. The Simple Secondary Vertex uses an adaptive vertex finder to reconstruct the secondary vertex and calculate the disciminator based on variables such as the decay length significance. The Soft Muon algorithms use the detection of a muon

from the semi-leptonic decay of a b quark and a neural net analysis. The Track Counting algorithms simply take all the tracks within a jet and order them by decreasing order of signed impact parameter significance, with the discriminator being the significance of the second track (high efficiency) or the third track (high purity).

EXPAND ON ALL THESE ALGORITHMS: INCLUDE A MORE DETAILED DESCRIPTION OF EACH ALGORITHM

Analysing a Fall 2011 $t\bar{t}$ Madgraph Monte Carlo sample (/TTJets_TuneZ2_7TeV-madgraph-tauola/) simulated in CMSSW 44X, all algorithms produced higher discriminator values for the b jets in the sample than the light jets (up, down and strange flavour), gluon jets and c flavour jets. All histograms were normalised to unity in order to facilitate shape comparison. Figure 4.1 shows these normalised histograms produced by the CSV algorithm.

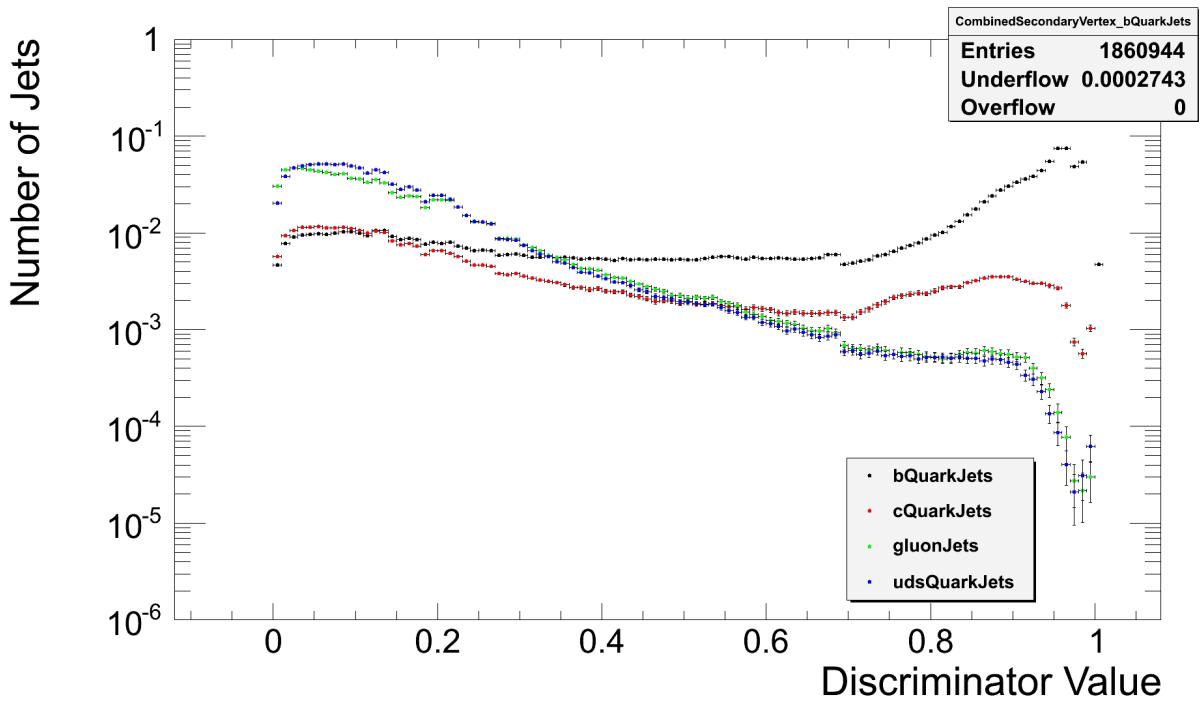


Figure 4.1: Discriminator values produced by the Combined Secondary Vertex algorithm for all 4 jet flavours after normalisation

4.1 Efficiency

Cutting on a desired discriminator value allows the efficiency of that cut value to be found by taking the area of the histogram to the right of the cut value, as a proportion of the total histogram area. Cuts were made through the whole range of discriminator values in order to create plots of b tag efficiency as a function of cut value. In practice, the aim is to achieve a

high efficiency for b jets and a low efficiency for all other jet flavours. Figure 4.2 shows how the efficiency for light, gluon and c jets vary as a function of b jet efficiency for cuts made on the CSV discriminator.

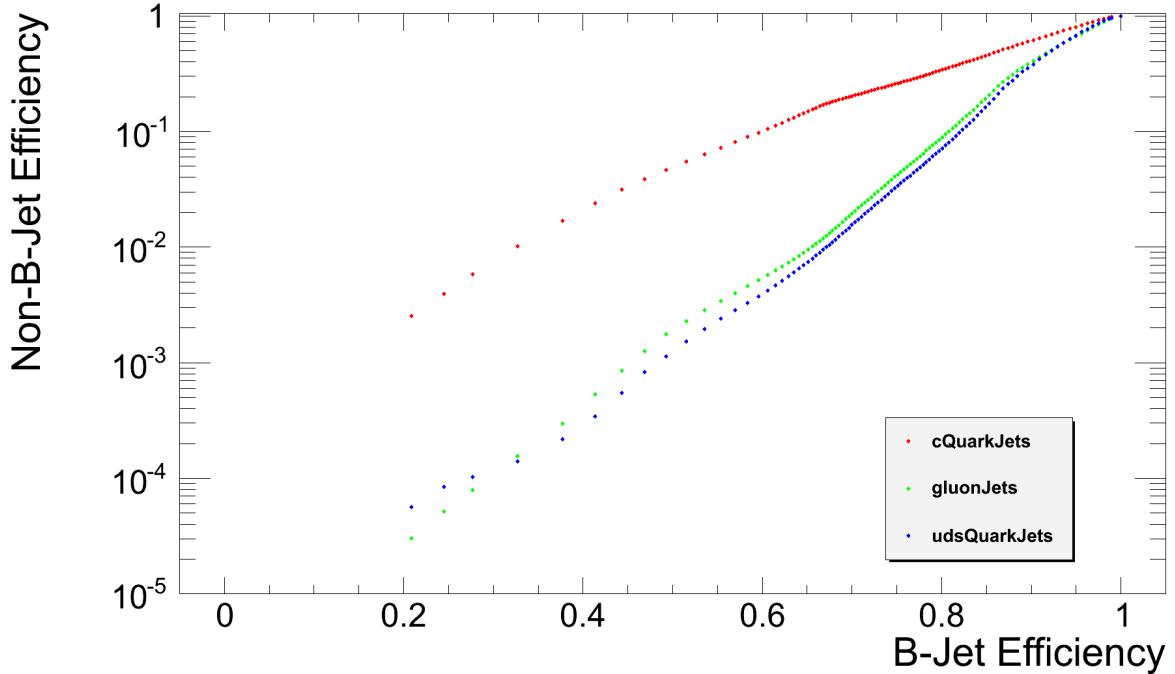


Figure 4.2: Non b jet efficiencies as a function of b jet efficiency for the CSV algorithm

4.2 Algorithm Comparison

The performances of the various algorithms can be compared in Figure 4.3.

Algorithms which reach closest to the lower lower right of the plot show better performance, according to the requirements of high b jet efficiency and low light jet efficiency. It can be seen that not all algorithms reach 100% efficiency for b jets due to some being inherently limited by their methods EXPAND ON THIS. For instance, the soft muon algorithms all show low maximum b jet efficiencies due to the low b hadron semi leptonic branching ratio to muon of approximately 11% (or 20% when further decays are included) [21]. The 2011 and 2012 CMS recommended b tagger is the Combined Secondary Vertex with operating point cuts of 0.244 (loose), 0.679 (medium) and 0.898 (tight) corresponding to 10%, 1% and 0.1% light jet and gluon jet efficiencies respectively. These cuts are indicated by the horizontal lines on Figure 4.3. It can be seen that for the tight and medium cuts, the CSV MVA algorithms provides highest b jet efficiency and lowest light jet efficiency, followed closely by the CSV algorithm. At approximately 3% light jet efficiency, there is a convergence of many

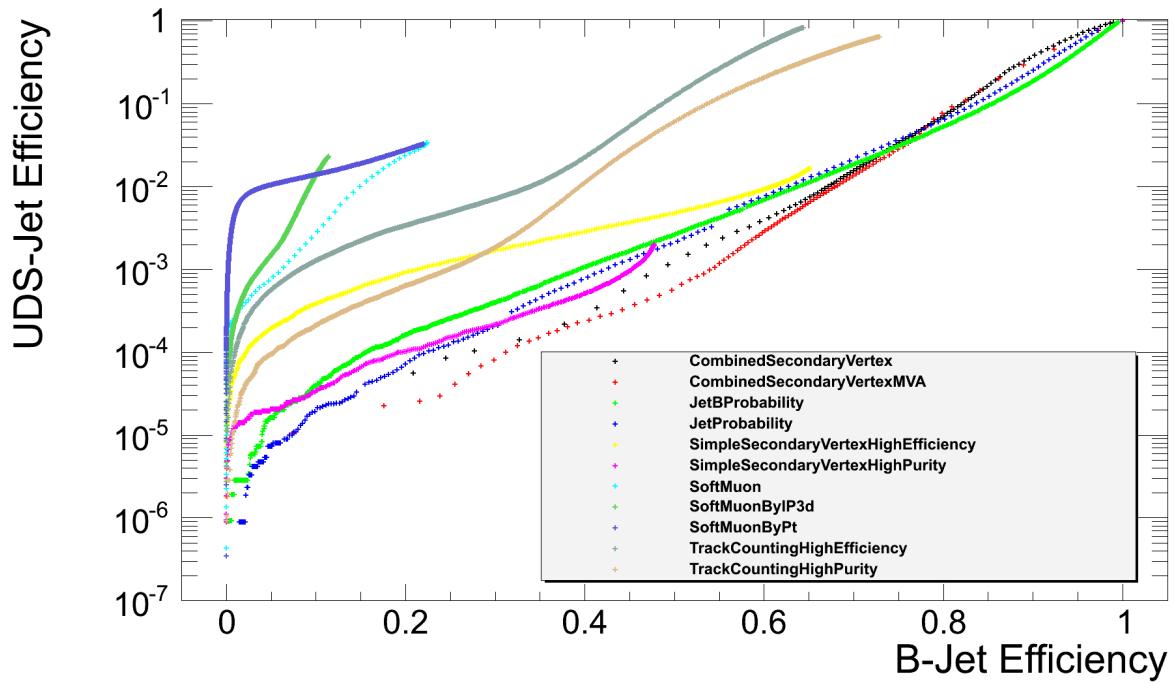


Figure 4.3: UDS jet efficiencies as a function of b jet efficiencies for all algorithms

b taggers which all provide similar performance, above which the JetBProbability algorithm gives marginally higher b jet efficiency.

Chapter 5

7 TeV and 8 TeV Differential Cross Section Measurement

5.1 Introduction

This analysis measures the normalised differential $t\bar{t}$ cross section in the semi-leptonic channel and is carried out on 2011 and 2012 data recorded from the CMS detector. The measurement is carried out with respect to the following global variables is carried out:

- E_T^{miss} , the missing transverse energy in an event
- H_T , the sum of the jet transverse momenta in an event
- S_T , the sum of the observed transverse momenta in an event
- M_T^W , the transverse mass of the leptonically decaying W boson
- p_T^W , the transverse momentum of the leptonically decaying W boson

Previous analyses using 7 TeV data investigating E_T^{miss} and using 8 TeV data investigatin all the global variables listed above can be found in [3] and [4] respectively.

This investigation is motivated primarily by the importance of understanding $t\bar{t}$ events since they are a significant background in many new physics analyses. It is also helpful in understanding QCD and event generators. Rare Standard Model processes such as $t\bar{t} + W \rightarrow l\nu$ or $t\bar{t} + Z \rightarrow \nu\bar{\nu}$ would appear in E_T^{miss} distribution tail, and $t\bar{t} + X$ where X is massive would appear in the H_T and S_T distributions. There are also possible new physics scenarios such as stop pair production, $\tilde{t}\bar{\tilde{t}} \rightarrow t\tilde{\chi}_0\bar{t}\tilde{\chi}_0$ which could show hints of dark matter.

5.2 Data and Simulated Samples

5.2.1 Data

Data collected in 2011 at a centre-of-mass energy of 7 TeV and in 2012 at a centre-of-mass energy of 8 TeV are used. The datasets are determined by the triggers that were used to record them. For 7 TeV, the ElectronHad dataset is used in the electron channel. It was recorded with triggers that select based on a single, isolated electron and additional jets. At 8 TeV, the SingleElectron dataset was used for the electron channel which is based on a single, isolated electron. In the muon channel, the SingleMu dataset was used for both centre-of-mass energies, requiring a single, isolated muon. See Section 5.3 for more details about triggers. The datasets used are shown in Tables A.1 and tab:datasets8TeV in Appendix A. Only data that is certified as golden, which is defined as data taken with the detector working without any major faults, is used. The masks used to filter out data taken in other periods are specified in Table A.3 in Appendix A.

5.2.2 Simulated Samples

The Monte Carlo generators used in this analysis are MadGraph [?], PYTHIA [?], POWHEG [?, ?, ?], HERWIG [?] and MC@NLO [?, ?].

The signal for this analysis is the production of a $t\bar{t}$ pair which decays semi-leptonically, i.e. each of the t quarks decays to a W boson and a b jet, with one of the W bosons decaying hadronically to two jets and the other decaying leptonically to a lepton (electron or muon) and an associated neutrino.

Additional, lower energy jets, can be produced from other scatterings in the same proton-proton interaction and from gluon radiation from quarks in the decay. Events with W bosons or Z bosons and additional jets are a significant background to semi-leptonic $t\bar{t}$ analyses. In the case of W + jets, leptonically decaying W bosons and additional jets would provide a similar event signature to that of a semi-leptonic $t\bar{t}$ decay. However, in general, these processes can be removed from the signal selection because the final decay products in W + jets events have lower energies than those from semi-leptonic $t\bar{t}$ decays, since the top quark has a high mass. Another characteristic of W + jets events is that the jets are less likely to be the necessary b jets in $t\bar{t}$ events. Similarly, Z + jets events can mimic $t\bar{t}$ events where the leptonic decay of Z bosons to a lepton and an anti-lepton. In these cases, a veto on a second lepton together with jet multiplicity requirements removes much of this background, although misidentification and misreconstruction of leptons as jets could still mean some

of these events still pass the signal selection, although this contamination would be small.
TODO:FEYNMANN DIAGRAM OF W Z PRODUCTION.

The QCD events that can form a background to $t\bar{t}$ analyses, with the multi-jet final states resulting from gluon-gluon fusion and quark-antiquark annihilation from proton proton collisions. At high orders, sufficient numbers of jets can be produced to produce an event signature similar to that of $t\bar{t}$ events. The leptons required for this to happen can come from jets that are misreconstructed and misidentified as leptons, or from the b and c quark decays. In the muon channel, it is possible to use isolation requirements to identify and remove a jet of high enough energy that it leaves signals in the muon chambers, because it will also deposit much energy in the calorimeters. Fake leptons caused by misreconstruction and misidentification, although not a regular occurrence, form a significant background because the QCD cross section is large at the centre of mass energies in this analysis. Fake leptons contained in b and c jets will not begin from the primary vertex and so can be mostly excluded from the signal selection using their distinct track signature compared to prompt leptons.

On the other hand, the electron channel poses a more problematic QCD background, due mainly to the conversion of photons, whether produced at the interaction point or through subsequent decays and radiation, into electrons and positrons. The identification and removal of such events has been described in Section 3.7.1. However, the large uncertainty in the cross section of QCD events, large contamination from higher order processes with additional jets in the signal region of this analysis and the difficulty in Monte Carlo modelling of such contributions lead to significant disagreements in the numbers of events passing the signal selection in data and in simulation. Therefore, the QCD background is modelled using a data driven method, described in Section ??.

Table ?? shows the Monte Carlo signal and background samples used in this analysis. Top pair production is modelled using MadGraph, single top events are modelled with POWHEG, W and Z boson plus jets production is modelled using MadGraph and QCD multi-jet and matrix element

A simultaneous fit is done with three templates in bins of each variable. The binning choice is made based on two variables: purity (equation) and stability (equation). The purity of a bin is sensitive to events moving into/out of a bin and the stability is sensitive to events moving inout/out of a bin. The bins for each global variable are selected such that each bin has purity and stability values of 0.5 or greater, meaning that at least half of the events created in a bin remain in that bin.

Electrons can be faked by jets, so the electron scale factors could be different in ttbar events than in events from which the scale factors were derived because our events contain a lot

more activity than events from which the scale factors were derived. For muons it is not so bad because muons leave a much cleaner signal inside CMS and there is no risk of jets faking muons.

We use the lepton trigger, identification and isolation scale factors provided by the Muon POG for both 7 TeV and 8 TeV. However, for electrons, although we have 8 TeV scale factors provided by the EGamma POG, we have had to derive scale factors ourselves for 7 TeV because not provided (and the trigger we use is not included in the 7 TeV Summer 11 Legacy datasets). The hadronic leg of the electronHad trigger needs to be measured in relation to the 4 jets. ASK ABOUT THIS

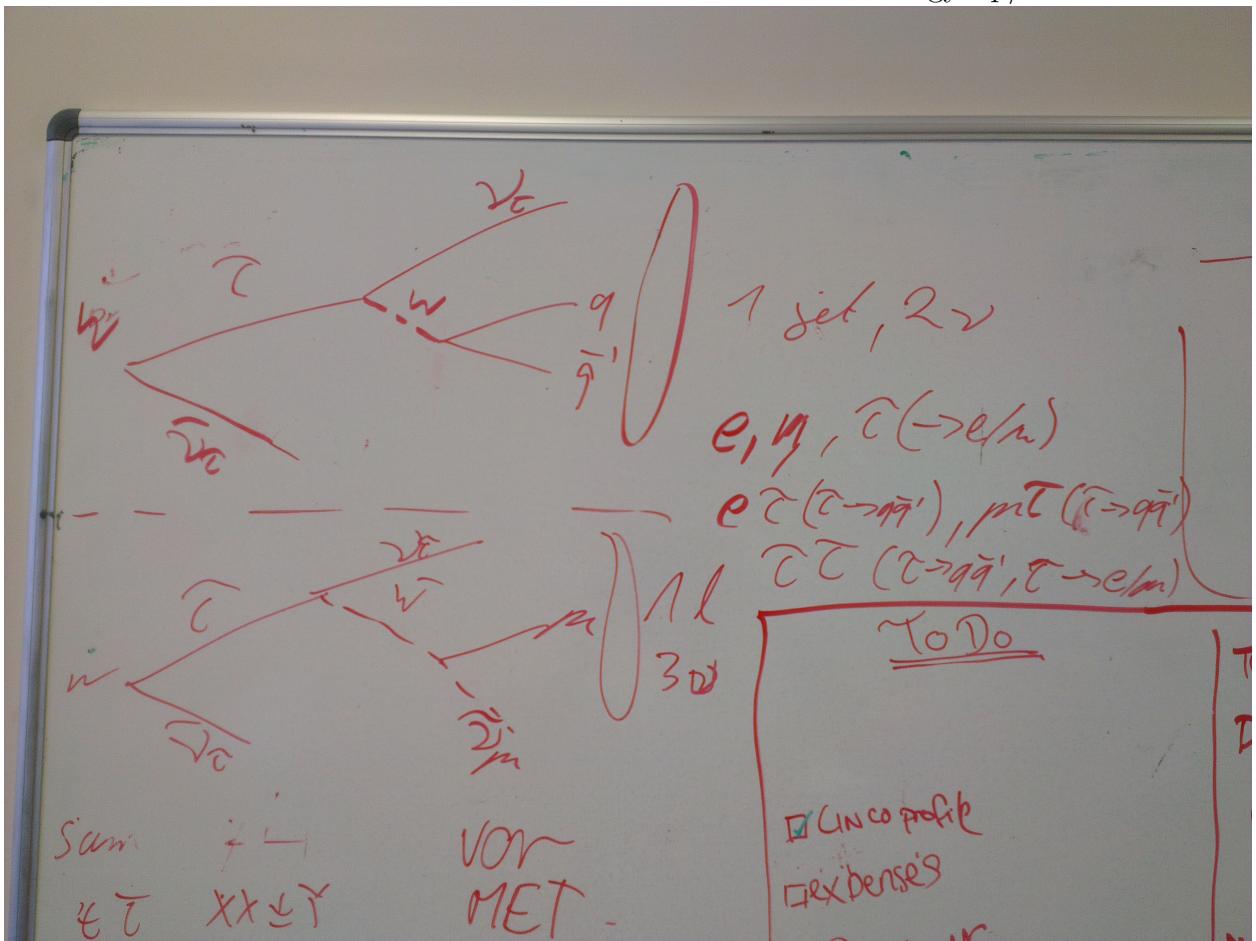
Remember: - V_Jets template combined over all global variable bins. - QCD template also inclusive over all global variable bins

Within the framework of this analysis, prompt leptons are considered as those which come from W or Z boson or from tau-lepton decays. These leptons are usually well isolated, whereas misidentified leptons originate either from semi-leptonic heavy flavor decays within jets or are simply misreconstructed genuine jets. In both cases these leptons are generally not isolated.

Tau energy: The met energy systematics are electron energy up/down, muon energy up/down and tau energy up/down. These systematics (and JES) are applied to both monte carlo and data. We found that the tau energy is surprisingly high, considerably higher than the electron and muon energy systematics. We may select taus in our signal region unintentionally although we don't select on taus. The signature of ttbar events with taus can mimic the signature of electron + jets or muon + jets ttbar events. The image shows how taus decay. The W from a top could decay to a tau, which would then decay either to a tau neutrino and a virtual W which then decays to two jets which would be very close together (and therefore reconstructed as a single jet). Or, the virtual W could decay to a lepton (electron or muon) and an associated neutrino). These signatures could fake our signal, and therefore end up in our ttbar signal. In our unfolding, we remove fakes. However, we use the shape of the fakes in the central to subtract from the signal distribution in the tau energy up/down variations. The effect is not so pronounced in the electron up/down and muon up/down variations because there are not many electrons or muons in the fakes. Luke found that $\tilde{1}4\%$ of the TTJet events we select on are NOT from the decay we are looking for. This was by dividing the number of events in the fake ST distribution in the unfolding histogram file by the number of events in the measured distribution (i.e. signal) in the same file. The electron channel was actually 13.5% and the muon channel was actually 13.9%. The shape comparison between the fakes and the signal shows that there are more fakes in the higher MET bins, which is why the error in those bins is larger. These are likely to come from

semi-leptonic tau events (where the tau decays to a lepton and a neutrino); tau tau, e tau or mu tau where the tau decays hadronically; or ee, emu or mumu events where one of the leptons is lost (probably low fraction as well). So most of our fakes will include at least one tau. Since semi-leptonic tau would need to decay leptonically, these taus are not identified and, I assume, not included in the MET uncertainty.

Looking at the shape difference between the central and tau up/down MET distributions, there is a significant different (though not by much). Since this affects data, the difference goes through the fit, and the unfolding, and is what we see in the final result. The difference in normalisation in the most affected bin is $\tilde{6}\%$, suggesting that we have more than 6% tau contamination overall (not sure if that is a correct assumption). The normalisation across the bins does not change, of course, i.e. the sum of events in the bins of, in this case MET, remain the same in the central measurement and in the tau energy up/down variations.



5.3 Event Selection

Bibliography

- [1] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.
- [2] Electron reconstruction and identification at $\sqrt{s} = 7$ TeV. Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, 2010.
- [3] Measurement of missing transverse energy in top pair events. Technical Report CMS-PAS-TOP-12-019, CERN, Geneva, 2012.
- [4] Measurement of MET and other global distributions in top pair events. Technical Report CMS-PAS-TOP-12-042, CERN, Geneva, 2013.
- [5] Wolfgang Adam, R Frhwirth, Are Strandlie, and T Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. Technical Report CMS-NOTE-2005-001, CERN, Geneva, Jan 2005.
- [6] Roberta Arcidiacono, Angela Brett, Francesca Cavallari, Andr David, Nicholas Scott Eggert, Giovanni Franzoni, Matteo Marone, Pasquale Musella, Giovanni Organtini, Paolo Rumerio, Alessandro Thea, and Evgueni Vlassov. ECAL Front-End Monitoring in the CMS experiment. Technical Report CMS-CR-2009-181, CERN, Geneva, Jul 2009.
- [7] Erik Manuel Butz. Operation and Performance of the CMS Silicon Tracker. Technical Report CMS-CR-2012-357, CERN, Geneva, Nov 2012.
- [8] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti- $k(t)$ jet clustering algorithm. *JHEP*, 0804:063, 2008.
- [9] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [10] S Chatrchyan et al. Commissioning and Performance of the CMS Silicon Strip Tracker with Cosmic Ray Muons. *JINST*, 5:T03008, 2010.
- [11] Serguei Chatrchyan et al. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6:P11002, 2011.

- [12] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.
- [13] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013.
- [14] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
- [15] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [16] CMS Collaboration. Technical proposal for the upgrade of the CMS detector through 2020. Technical Report CERN-LHCC-2011-006. LHCC-P-004, CERN, Geneva, Jun 2011.
- [17] CMS Collaboration. Jet transverse momentum resolution measurement using dijet events. *AN-13-416, work in progress*, 2013.
- [18] The ATLAS collaboration. Combination of ATLAS and CMS results on the mass of the top-quark using up to 4.9 fb^{-1} of $\sqrt{s} = 7$ TeV LHC data. 2013.
- [19] E P Corrin. *Development of Digital Readout Electronics for the CMS Tracker*. PhD thesis, Imperial College, London, 2002.
- [20] D. Delikaris, J.P. Dauvergne, G. Passardi, J.C. Lottin, J.P. Lottin, et al. The Cryogenic system for the superconducting solenoid magnet of the CMS experiment. 1998.
- [21] Cristina Ferro. B-tagging in CMS. *EPJ Web Conf.*, 28:12055, 2012.
- [22] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl.Instrum.Meth.*, A262:444–450, 1987.
- [23] Vardan Khachatryan et al. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. 2015.
- [24] Min Suk Kim et al. CMS reconstruction improvement for the muon tracking by the RPC chambers. *PoS*, RPC2012:045, 2012.
- [25] Christos Leonidopoulos and The CMS Collaboration. Cms tracking performance results from early lhc operation. *The European Physical Journal C - Particles and Fields*, C70:1165–1192, 2010.
- [26] K.A. Olive et al. Review of Particle Physics. *Chin.Phys.*, C38:090001, 2014.

- [27] F M Palmonari. CMS Silicon Strip Tracker Commissioning and Preliminary Performances. *Nucl. Phys. B, Proc. Suppl.*, 197(1):121--124, 2009.
- [28] Tai Sakuma and Thomas McCauley. Detector and Event Visualization with SketchUp at the CMS Experiment. *J.Phys.Conf.Ser.*, 513:022032, 2014.
- [29] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, Switzerland, 2006.
- [30] Christian Weiser. A Combined Secondary Vertex Based B-Tagging Algorithm in CMS. Technical Report CMS-NOTE-2006-014, CERN, Geneva, Jan 2006.

Appendix A

Datasets in cross section analysis

Data set name	Run period	L_{int} / pb $^{-1}$	Runs
ElectronHad 12 Oct 2013 ReReco	Run2011A	2,333	160404--173692
ElectronHad 12 Oct 2013 ReReco	Run2011B	2,738	175833--180252
SingleMu 12 Oct 2013 ReReco	Run2011A	2,331	160404--173692
SingleMu 12 Oct 2013 ReReco	Run2011B	2,766	175833--180252

Table A.1: 7 TeV data sets by run period with the corresponding integrated luminosities (L_{int}) and run numbers.

Data set name	Run period	L_{int} / pb $^{-1}$	Runs
SingleElectron 22 Jan 2013 ReReco	Run2012A	883.3	190456--193621
SingleElectron 22 Jan 2013 ReReco	Run2012B	4,389.0	193834--196531
SingleElectron 22 Jan 2013 ReReco	Run2012C	7,137.0	198022--203742
SingleElectron 22 Jan 2013 ReReco	Run2012D	7,318.0	203777--208686
SingleMu 22 Jan 2013 ReReco	Run2012A	889.4	190456--193621
SingleMu 22 Jan 2013 ReReco	Run2012B	4,424.0	193834--196531
SingleMu 22 Jan 2013 ReReco	Run2012C	7,152.0	198022--203742
SingleMu 22 Jan 2013 ReReco	Run2012D	7,280.0	203777--208686

Table A.2: 8 TeV data sets by run period with the corresponding integrated luminosities (L_{int}) and run numbers.

Data Period	Mask
2011	Cert_160404-180252_7TeV_ReRecoNov08_Collisions11_JSON_v2.txt
2012	Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt

Table A.3: 8 TeV data sets by run period with the corresponding integrated luminosities (L_{int}) and run numbers.