

The Search for the Standard Model Production of Four Top Quarks

Lana Beck

A dissertation submitted to the University of Bristol and Vrije Universiteit Brussel
in accordance with the requirements of the degree of PhD in the Faculty of
Science, School of Physics, Bristol and Faculteit Wetenschappen en
Bio-ingenieurswetenschappen, Vakgroep Fysica, Brussel.

Dec 2016

Abstract

The Standard Model (SM) of Particle Physics has been incredibly successful and accurate in describing the fundamental particles that make up the world around us and the way they behave. However, we know it is not the ultimate theory of nature as some phenomena remain unexplained. Outstanding questions include what is dark matter, this mysterious material which can be inferred from observations of the universe but has not been directly detected; and where does gravity fit into the picture? These questions motivate our search for physics Beyond the Standard Model (BSM) at the Large Hadron Collider at the CERN research facility. Here we accelerate protons around a 27 km ring and collide them at four experiments located underground. When we collide protons we effectively collide the more fundamental particles within the protons, quarks and gluons.

This thesis focuses on research undertaken at the CMS experiment studying the heaviest quarks, top quarks, which are not found in nature but instead are produced in high-energy experiments. Top quarks are most often produced in pairs, however this thesis focuses on the search for the simultaneous production of four top quarks, which is an incredibly rare process in comparison. A precision measurement of this rare process would be a stringent test on the SM and may give hints of physics beyond the standard model. Untangling the signal of four-top-quark production from the overwhelming background of top-quark-pair production in the output of the detector is incredibly difficult. Algorithms, which are often used in developing artificial intelligence, are therefore employed to exploit subtle differences in signatures, greatly increasing the sensitivity. Results are presented which place tight limits on the rate of four-top-quark production, and projections of the future sensitivity are made including an estimate of when CMS will have sufficient data to definitively observe this process at SM rates. The results also allow us to place constraints on properties of hypothesised BSM particles. Here we interpret the results to place constraints on the mass and top quark-coupling of one such particle, the sgluon.

Samenvatting

Het Standaard Model (SM) van deeltjesfysica is ongelofelijk succesvol en nauwkeurig in het beschrijven van de fundamentele deeltjes in de wereld om ons heen en de manier waarop ze zich gedragen. Maar we weten dat het niet de ultieme natuurkundige theorie is, omdat een aantal verschijnselen onverklaard blijven in het SM. Openstaande vragen zijn: wat is donkere materie, het bestaan van deze mysterieuze substantie kan worden afgeleid uit waarnemingen van het heelal, maar is nog niet rechtstreeks waargenomen; en waar past de zwaartekracht in het plaatje? Deze vragen motiveren onze zoektocht naar de fysica voorbij het Standaard Model (BSM) bij de Large Hadron Collider op het CERN. Hier versnellen we protonen rond een ring van 27 km omtrek. Op vier plaatsen aan de ring liggen ondergrondse experimenten waar de protonen worden gebotst. Wanneer protonen botsen, bestuderen we effectief de werkelijk fundamentele deeltjes in het proton, quarks en gluonen. Dit proefschrift richt zich op onderzoek gebruikmakend van de proton-proton botsingen waargenomen door het CMS-experiment. Het zwaarste bekende elementaire deeltje, de top-quark, is niet te vinden in de natuur, maar in plaats daarvan kan worden geproduceerd in de botsingen bij de LHC. Top quarks worden meestal geproduceerd in paren, maar dit proefschrift richt zich op de zoektocht naar de productie van vier top-quark tegelijkertijd, dat is een ongelofelijk zeldzaam proces in vergelijking met paarproductie. Een nauwkeurige meting van deze zeldzame proces zou een strenge test zijn van het SM en kan hints geven of er nieuwe deeltjes worden gemaakt samen die in vier top quarks uiteenvallen. Het identificeren van het signaal van vier top-quark productie in de overweldigende achtergrond van top-quark-paarproductie in de data is een ongelofelijk moeilijke wetenschappelijke uitdaging. Hiervoor worden machine-learning algoritmen toegepast. Op deze wijze kunnen subtile verschillen tussen de productie van top quark paren en vier top quarks in de botsing worden benut, en dit leidt tot een aanzienlijke verhoging van de gevoeligheid van de data-analyse. De resultaten van dit onderzoek plaatsen de meest strakke grenzen aan de werkzame doorsnede voor productie van vier top quarks. De resultaten zijn ook in staat om beperkingen te geven op de eigenschappen van eventuele hypothetische BSM deeltjes. In dit onderzoek worden daarom de resultaten ook geïnterpreteerd als een functie van de massa en top quark-koppeling van zo'n hypothetisch nieuw deeltje, het zogenaamde sgluon.

Acknowledgements

First of all I would like to thank my supervisors, Freya Blekman and Joel Goldstein, for being incredibly supportive, helpful and great fun to work with over the past four years. I would also like to thank James Keaveney for teaching me everything possible about being a four-top expert in the first two years of my PhD, for continuing to discuss physics problems with me since then and for being great craic. It was also a pleasure to work with Jesse and Steve from UCR on combining our four-top searches together and I wish Denys and Long the best in taking the analysis forward from here. I would like to thank my collaborators on our phenomenology project, Didar, Benjamin and Kentarou, for a fruitful and interesting collaboration. I would like to thank both the Bristol and Brussels research group for being great colleagues and fun to work with. I would particularly like to thank Robin for all the advice, memes and the exchange of swearing at pieces of code, and Maarten for housing and feeding me at the very start of my PhD while I found somewhere to live. Many thanks go to Martisse, you are sunshine on a cloudy day. Thanks to Leonor for being a supportive flatmate over this past two years and helping me out in these last few weeks of writing. Thanks to Mark for being my rock, for encouraging me and helping me through this writing period and for always being there to laugh with me. I am incredibly appreciative of the never-ending support from my parents who brought me up to believe that I could be whatever I wanted to be.

Lastly, of course, this thesis would not be possible without the work from everyone in the CMS collaboration who work together for the common goal of pushing the boundaries of physics.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol and Vrije Universiteit Brussel. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol or Vrije Universiteit Brussel. The dissertation has only been presented to the University of Bristol and Vrije Universiteit Brussel for attainment of a joint degree and has not been presented to any other university for examination either in the United Kingdom or overseas.

Signed

Date

Contents

1	The CMS detector and the Large Hadron Collider	1
1.1	LHC	1
1.2	CMS detector	4
1.2.1	Magnetic Solenoid	5
1.2.2	Tracker	6
1.2.3	Electromagnetic Calorimeter	7
1.2.4	Hadron Calorimeter	9
1.2.5	Muon Chambers	10
1.2.6	Trigger	12
1.2.7	Upgrades for Run 2	13
1.2.8	Data collections	14
A	Cross checks on Run 2 $t\bar{t}t\bar{t}$ analysis at $\sqrt{s} = 13$ TeV	17
A.1	Comparison of alternative $t\bar{t}$ generators	17
A.2	TTZ, TTW, TTH MC backgrounds	17
A.3	Comparison of the Gradient Boost and AdaBoost boosting algorithms within the BDT	18
A.4	Event-level BDT templates	19
A.5	Systematic shape studies	23

A.5.1 Studies of impact of systematic uncertainties	23
A.6 Correlation matrices for fit nuisance parameters	27
References	29

List of Figures

1.1	The LHC accelerator complex at CERN. Protons are accelerated from LINAC 2 into the BOOSTER synchrotron. From there they are further accelerated in the proton synchrotron (PS) and super proton synchrotron (SPS) before finally being injected in two counter-rotating beams in the large hadron collider (LHC). The beams are crossed at the four experiments: CMS, LHCb, ATLAS and ALICE [2].	2
1.2	The integrated luminosity in fb^{-1} for proton-proton collision at the CMS experiment from 2010 to 2016 [3].	3
1.3	The CMS detector [6].	5
1.4	The tracking system [8].	7
1.5	The ECAL system where modules contain submodules and the 36 supermodules are made up of four modules each. The supercrystals are made up of groups of 5 x 5 crystals in the endcap [10].	9
1.6	The HCAL system [10].	10
1.7	The muon chamber system [12].	11
1.8	The muon transverse momentum resolution as a function of transverse momentum (p_T) using the muon system only (black), the inner tracking only (blue), and both (red), in regions of $ \eta < 0.8$ (left) and $1.2 < \eta < 2.4$ (right) [8].	13
A.1	Inclusive BDT distribution for $t\bar{t}$ generators POWHEG +PYTHIA, MADGRAPH_MLM and MADGRAPH_AMC@NLO FxFx	17

A.2	BDT discriminator shapes for all categories, as indicated along the x axis. The ratio plot shows the difference between each distribution and the nominal $t\bar{t}$ distribution divided by the $t\bar{t}$ distribution.	18
A.3	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the 6 N_{jets} and $3N_{\text{tags}}^M$ category.	19
A.4	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right) are shown for the 6 N_{jets} and $\geq 4 N_{\text{tags}}^M$ category.	20
A.5	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the 7 N_{jets} and $2 N_{\text{tags}}^M$ category.	20
A.6	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the 7 N_{jets} and $3 N_{\text{tags}}^M$ category.	20
A.7	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right) are shown for the 7 N_{jets} and $\geq 4 N_{\text{tags}}^M$ category.	21
A.8	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the 8 N_{jets} and $2 N_{\text{tags}}^M$ category.	21
A.9	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the 8 N_{jets} category and $3 N_{\text{tags}}^M$ category.	21
A.10	The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right) are shown for the 8 N_{jets} and $\geq 4 N_{\text{tags}}^M$ category.	22

A.11 The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (left) are shown for the $\geq 9 N_{\text{jets}}$ and $2 N_{\text{tags}}^M$ category.	22
A.12 The BDT shapes for JER systematic in $t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	23
A.13 The BDT shapes for JES systematic in $t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	23
A.14 The BDT shapes for JES systematic in $t\bar{t}t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	24
A.15 The BDT shapes for ME scale systematic in $t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	24
A.16 The BDT shapes for ME scale systematic in $t\bar{t}t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	24
A.17 The BDT shapes for PU systematic in $t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	25
A.18 The BDT shapes for ttGenerator choice systematic in $t\bar{t}$ for the $\mu +$ jets channel (left) and $e +$ jets channel (right).	25
A.19 The correlation matrices for background only for the fit parameters. .	27

List of Tables

1.1	Run 1 datasets at 8 TeV, when they were recorded and how much data was recorded.	15
1.2	Run 2 datasets at 13 TeV, when they were recorded and how much data was recorded.	15
A.1	Expected limits using jet categories of 6, 7, 8, 9+ jets for different BDT boosting algorithms.	19
A.2	Expected limits on $t\bar{t}t\bar{t}$ production which each systematic removed in turn	26

1 | The CMS detector and the Large Hadron Collider

This chapter discusses the Large Hadron Collider (LHC) which is located on the Franco-Swiss border near Geneva, approximately 100 m underground at the site of the Conseil Européen pour la Recherche Nucléaire (CERN)¹. It is a 26.7 km long synchrotron particle accelerator with four interaction points where four experiments are located. This thesis focuses on results from the Compact Muon Solenoid detector described in Section 1.2. The other experiments include ATLAS which is a multi-purpose experiment like CMS, the LHCb detector which focuses on the study of the physics of B hadrons, and the ALICE detector which is used to study the quark-gluon plasma. Details of the function and purpose of each of the CMS sub-detectors are given in this chapter. The algorithms which are used to reconstruct particles from the information given from the sub-detectors are given in Chapter ??.

1.1 LHC

The LHC accelerates two beams of protons which circulate in opposite directions. This is achieved using a system of superconducting dipole magnets which have an aperture for each beam direction. Quadrupoles are used to squeeze the beam. The protons are sourced from a bottle of hydrogen where a strong electric field is used to excite the electrons from the hydrogen atoms leaving protons behind. The protons are then accelerated through the linear accelerator LINAC2, where they reach an energy of 50 MeV. The LINAC2 works by using radio frequency cavities to charge cylindrical conductors and the protons are attracted by negative conductors and repelled by positive conductors, which boost them forward. From LINAC2, the protons are accelerated by the Proton Synchrotron Booster (PSB) to 1.4 GeV, then they are accelerated by the Proton Synchrotron (PS) to 25 GeV. This is followed by

¹This chapter is largely adapted from Ref. [1]

a boost to 450 GeV in the Super Proton Synchrotron (SPS), which is the final accelerator before the protons are injected into the LHC ring where their final collision energy can be achieved.

The protons are accelerated in bunches which are collided at the interaction points of each experiment. This results in many collisions per bunch crossing despite the fact that many protons will miss each other and continue to be accelerated around the LHC. The background of particles coming from the frequent less interesting collisions is called *pileup* (PU).

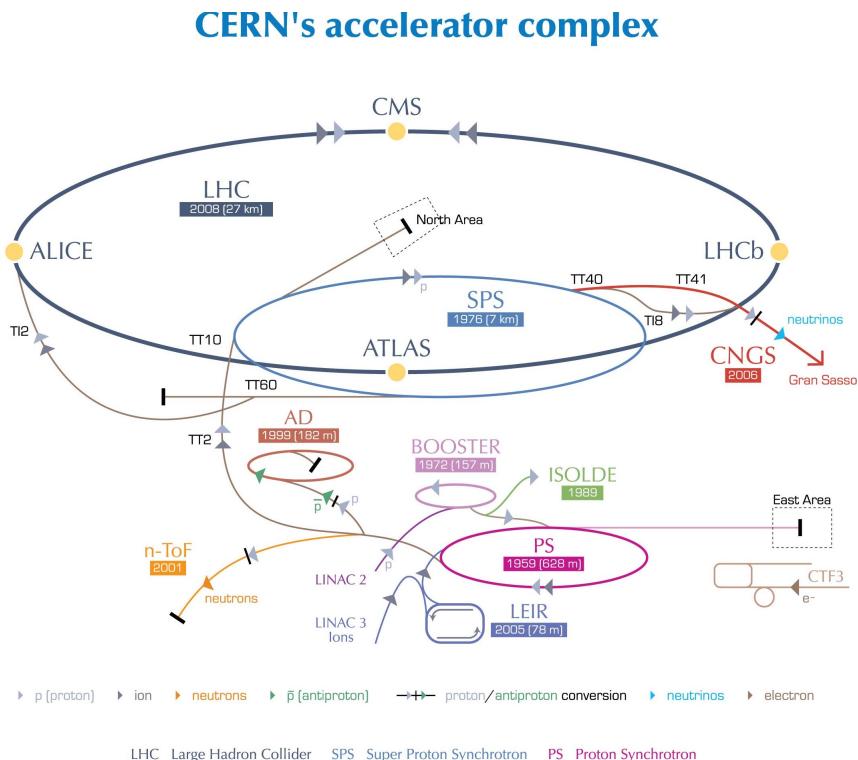


Figure 1.1: The LHC accelerator complex at CERN. Protons are accelerated from LINAC 2 into the BOOSTER synchrotron. From there they are further accelerated in the proton synchrotron (PS) and super proton synchrotron (SPS) before finally being injected in two counter-rotating beams in the large hadron collider (LHC). The beams are crossed at the four experiments: CMS, LHCb, ATLAS and ALICE [2].

The *Luminosity* (\mathcal{L}) is a measure of the instantaneous collision rate and can be calculated using Eq. 1.1, where f is the bunch frequency, and N_1 and N_2 are the numbers of particles in each bunch. The effective collision area is $4\pi\sigma_x\sigma_y$ where σ_x and σ_y are the x and y components of the beam, transverse to the beam direction.

It is assumed that each beam has the same cross section.

$$\mathcal{L} = \frac{f N_1 N_2}{4\pi \sigma_x \sigma_y} \quad (1.1)$$

The *integrated luminosity* (\mathcal{L}_{int}) which is shown in Fig 1.2 for the CMS experiment from 2010 until 2016 for proton-proton collisions is the delivered luminosity integrated over time. The number of events, N_{events} , produced for a particular particle physics process can be calculated from the cross section, σ , using Eq. 1.2.

$$N_{\text{events}} = \mathcal{L}_{\text{int}} \times \sigma \quad (1.2)$$

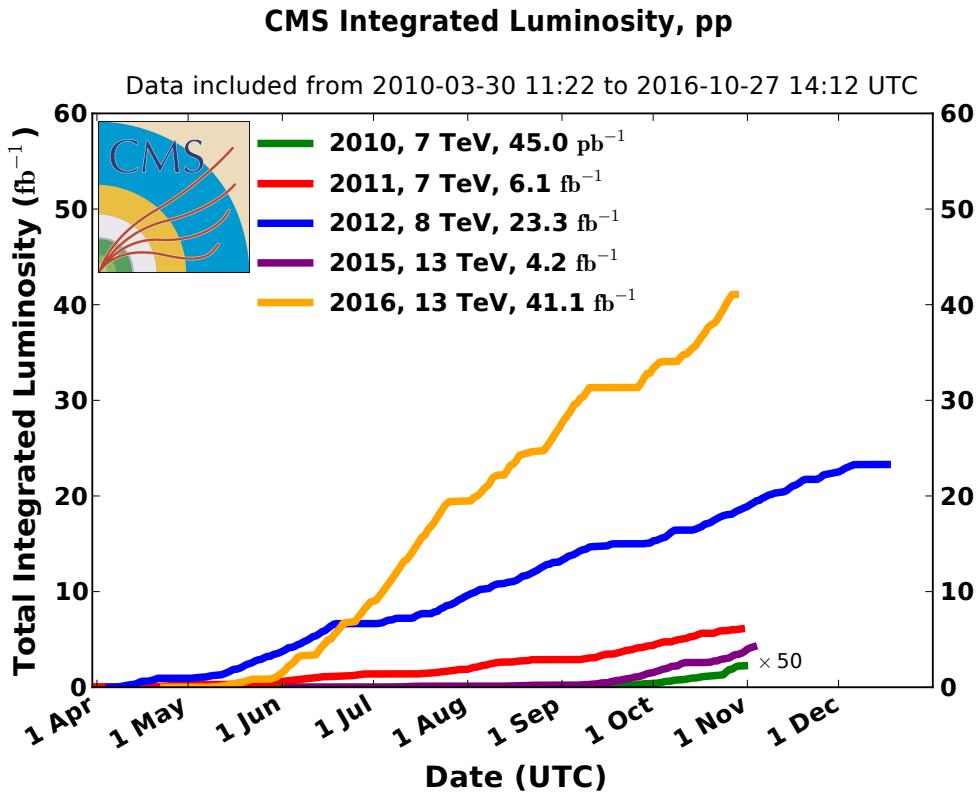


Figure 1.2: The integrated luminosity in fb^{-1} for proton-proton collision at the CMS experiment from 2010 to 2016 [3].

The LHC was designed to have a centre of mass collision energy of $\sqrt{s} = 14 \text{ TeV}$. The intention was to start the machine with a lower energy in September 2008 and to obtain $\sqrt{s} = 10 \text{ TeV}$ by the end of 2008. However, an electrical fault 10 days into operation in October 2008 caused damage to over 50 of the superconducting magnets. The LHC was shut down for repairs until November 2009 when LHC

achieved the record breaking 1.18 TeV per beam. The centre of mass energy was then ramped up to $\sqrt{s} = 7$ TeV in 2011. In 2012 this was increased to $\sqrt{s} = 8$ TeV and a dataset with an integrated luminosity of $\approx 20 \text{ fb}^{-1}$ was recorded. This dataset from the run phase known as ‘Run 1’ was used for the analysis in Chapter ???. In 2013 after Run 1, Long Shutdown 1 (LS1) commenced to make upgrades to the LHC and the detectors to allow the machine to run at $\sqrt{s} = 13$ TeV for ‘Run 2’. Run 2 began in March 2015 and results from Run 2 are the focus of the analysis in Chapter ???.

Run 1 saw the great success of the discovery of the Higgs boson, one of the main objectives of the LHC [4, 5]. In Run 2, the search for new physics continues where precision measurements will test the predictions of the SM. The CMS and ATLAS detectors are considered to be general-purpose detectors which can be used to test many areas of physics within and beyond the standard model. This includes searches for dark matter, supersymmetric particles, vector-like-quarks, lepton-flavour-violating processes, light Higgs and charged Higgs bosons, and studies of the properties of the Higgs boson such as the couplings and rare decay modes.

1.2 CMS detector

The CMS detector is a hermetic detector with a large magnetic solenoid which causes charged particles to follow a curved trajectory as they traverse the detector. Closest to the beam line is the silicon tracker which makes the most accurate position measurements. Next are the calorimetry systems for electromagnetic and separately for hadronic particles. All of these detectors are contained within the magnetic solenoid. The muon chambers are outside the solenoid where they detect muons, which are much more penetrating than other particles. The cylindrical coordinate system of the detector is defined as follows: the x -axis points towards the centre of the LHC ring, the y -axis points upwards and the z -axis points along the beamline in the anti-clockwise direction. The azimuthal angle ϕ is measured in the (x, y) plane

1.2. CMS detector

clockwise from the x axis. The polar angle θ is measured clockwise from the z -axis. More commonly the pseudorapidity, defined in Eq. 1.3, is used instead of θ .

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (1.3)$$

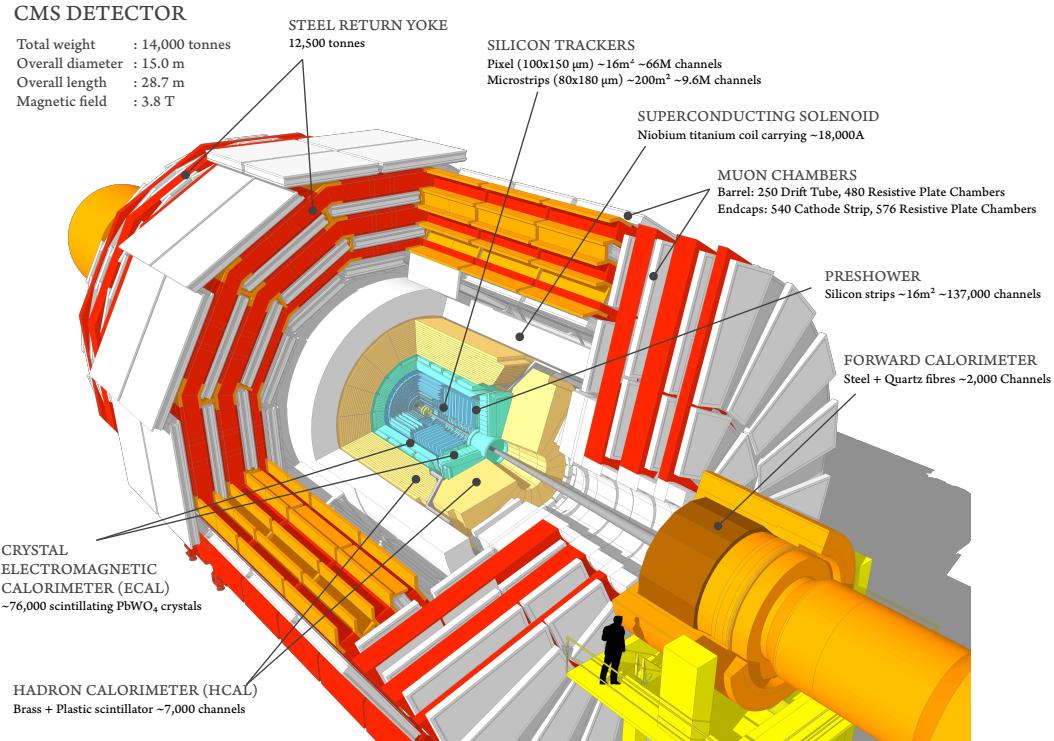


Figure 1.3: The CMS detector [6].

1.2.1 Magnetic Solenoid

The superconducting magnetic solenoid at the core of CMS was designed to have a magnetic field of 4 T. The free bore magnet has a diameter of 6.3 m and length of 12.5 m. It uses a 4-layer winding of NbTi superconductor which is required to generate this high magnetic field. The magnet is cooled within a cryostat using liquid helium to a temperature of 4.5 K. The magnetic field is returned via an iron yoke consisting of five barrel wheels and two endcaps which are made of three discs each. The outer dimension of the iron flats is 14 m. The tracker, electromagnetic calorimeter and hadronic calorimeter are constrained to be within the inner dimensions of the solenoid as can be seen in Fig. 1.3. The solenoid provides a homogeneous

magnetic field, which bends particle trajectories transverse to the beam direction, over the silicon tracker region shown in Fig. 1.4.

1.2.2 Tracker

The tracking system of CMS lies inside the superconducting solenoid and surrounds the interaction point. It is 5.8 m long with a diameter of 2.5 m. Silicon detectors are used as they can provide the high granularity and fast response required to reliably reconstruct the trajectories of particles coming from the collision vertex before the next collision occurs. Reconstructing secondary vertices is also particularly important for identifying jets originating from heavy flavour quarks such as bottom quarks. This is integral for distinguishing final states involving top quarks.

It is estimated that at a PU of 20 there are 1000 particles traversing the tracker at every bunch crossing. The silicon detectors have been designed to have the radiation hardness to last for the design time of ten years. The minimum material possible was used in order to reduce the amount of multiple scattering, photon conversion and bremsstrahlung. Cooling the tracker helps to prevent thermal runaway from leakage current, hence it was cooled to 0°C in Run 1 and -20°C in Run 2. The tracking system has a nominal momentum resolution of 0.7 (5.0)% for a particle with a momentum of 1 (1000) GeV in the central region. The impact parameter resolution is around 10 μm for high momentum tracks [7].

The tracking system consists of two main sections: the pixel tracker makes up the innermost section and the strip tracker surrounds the pixel tracker as seen in Fig. 1.4.

Pixel tracker

As the pixel detector is closest to the interaction vertex, it experiences the highest flux of particles at $\approx 1 \text{ MHz per mm}^2$. The fine granularity of the pixel detector ($100 \times 150 \mu\text{m}$ in $(r - \phi)x(x - z)$) is required in order to keep the occupancy below 1%. It consists of three barrel layers which range between 4.0 cm and 10.2 cm from the interaction point and 2 disks which are transverse to the beamline as seen in Fig. 1.4.

1.2. CMS detector

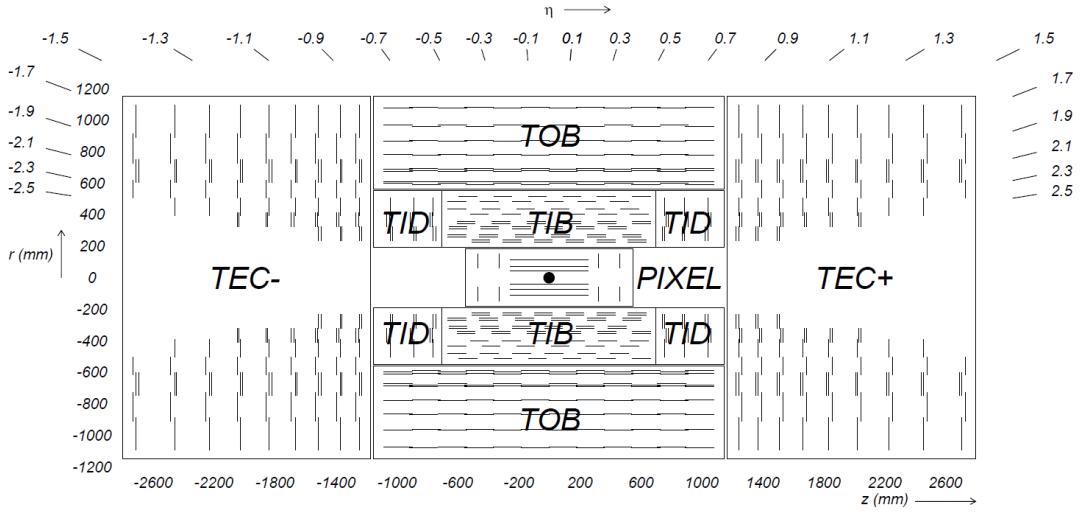


Figure 1.4: The tracking system [8].

Strip tracker

Two types of silicon strip tracker are used. Closest to the interaction point (20 - 50 cm) are the tracker inner barrel detectors (TIB) which contain silicon micro-strips ($10 \text{ cm} \times 80 \mu\text{m}$). An occupancy of $\approx 2\text{-}3\%$ is achieved for a fluence of $\approx 60 \text{ kHz}$ per mm^2 . An increased strip pitch of $180 \mu\text{m}$ can be used in the tracker outer barrel (TOB) due to the lower fluence of 3 kHz per mm^2 . To cover the larger surface area, the effective strip length is increased to 25 cm. However, increased strip length increases the noise. To combat this the strips are made thicker to $500 \mu\text{m}$ compared to $320 \mu\text{m}$ in the TIB. The tracker inner disk (TID) and tracker end cap (TEC) have strips which are aligned radially to the beamline with a strip pitch of $80 \mu\text{m}$ and $200 \mu\text{m}$, respectively. The TID and TEC extend the tracker acceptance to $|\eta| < 2.5$.

1.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous, hermetic detector made up of 61200 (7324) lead tungstate, PbWO_4 , crystals in the barrel (endcap) region. In the barrel region the light from these scintillating crystals is collected in avalanche photodiodes and in the endcap region by vacuum phototriodes. The barrel region covers $|\eta| < 1.479$ whilst the two semi-circular ‘Dees’ which make up the endcaps extend the range to $|\eta| < 3.0$.

Good resolution can be obtained from PbW₀₄ crystals and they are fast and radiation resistant. Equation 1.4 shows the dependence of the resolution on the energy of the particle, E. The stochastic term for the statistical fluctuations on the number of secondary particles produced is represented as S . The noise from the electronics and digitisation is given by N and the constant term C arises from calibration errors and leakage of the shower outside of the calorimeter.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C \quad (1.4)$$

For an electron test beam with no magnetic field and no material between the beam and the ECAL, the parameters in Eq. 1.4 were measured to be $S = 0.028, N = 0.12$ GeV and $C = 0.0003$ [9]. This gives an energy resolution of $\approx 0.5\%$ for a 100 GeV particle.

The crystals have a high density and small radiation length ($X_0 = 0.89$ cm) so that the ECAL can be compact. Another importantly quality is that they are optically clear allowing all fo the light to be collected. The scintillation decay time is comparable to the time between bunch crossings with 80% of light being emitted within 25 ns. The crystals have a tapered shape in the barrel region and lie parallel in the endcap region. The crystals are 1.29 m from the beam line in the barrel region and 3.15 m from the the interaction point in the longitudinal direction in the endcap region. The crystals are contained in thin-walled aluminium structures which make up submodules.

In the endcap region there is a preshower detector, shown in Fig. 1.5, which is a two-layer sampling calorimeter. There are two layers of lead used as a radiator material to initiate the electromagnetic shower and silicon strip detectors after each layer (orthogonal in each plane) to measure the energy deposited. The preshower detector helps to identify neutral pions and to distinguish electrons from minimum ionising particles.

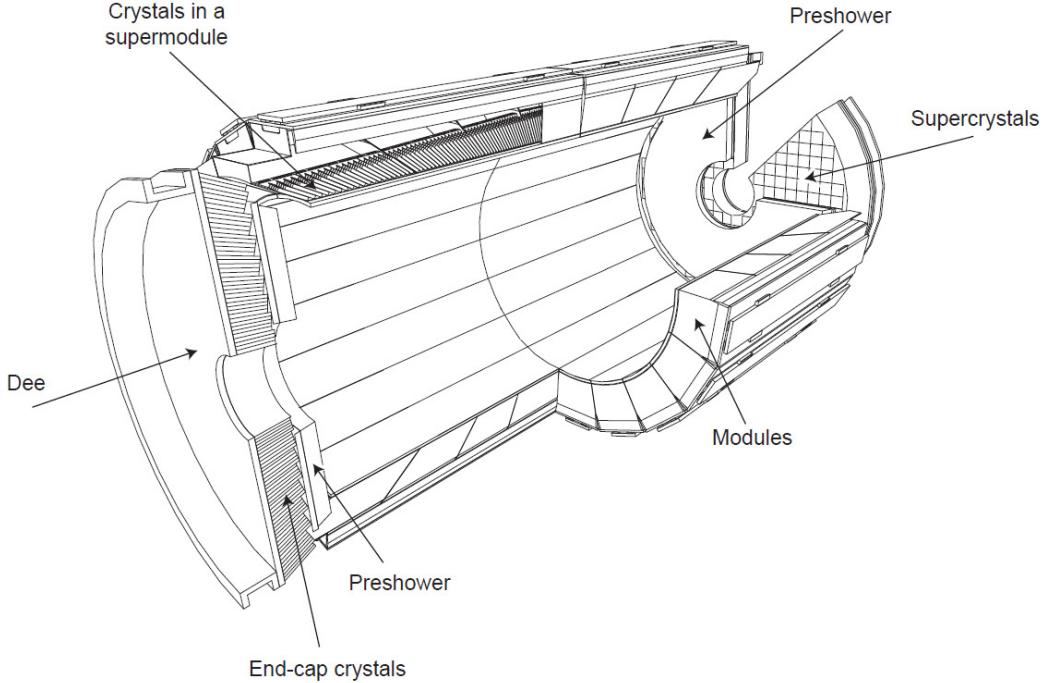


Figure 1.5: The ECAL system where modules contain submodules and the 36 supermodules are made up of four modules each. The supercrystals are made up of groups of 5×5 crystals in the endcap [10].

1.2.4 Hadron Calorimeter

The hadron calorimeter (HCAL) is used for identifying hadron jets. It has barrel (HB) and endcap (HE) regions made up of sampling calorimeters which have coverage of $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$, respectively. The HB and HE are placed between the ECAL and solenoid magnet, and therefore are restricted to the radial dimensions, R, of $1.77 \text{ m} < R < 2.95 \text{ m}$. The scintillators in both the HB and HE have a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ (0.17×0.17) for $|\eta| < 1.6$ (≥ 1.6). The barrel consists of two halves, HB+ and HB-, which are composed of 36 wedges made up of 14 flat brass absorber plates parallel to the beam axis, alternated with plastic scintillator. Stainless steel plates are used for the innermost and outermost plates to provide structural support. The HE has a similar system of alternating absorber and plastic scintillator. Due to their restricted dimensions the HCAL and ECAL do not always contain all of the energy from the particle showers. Between five and ten interaction lengths are contained within the HB, depending on the pseudorapidity. Therefore another detector known as the outer hadronic calorimeter (HO)

is embedded in the muon system, outside of the solenoid magnet, to measure the energy leakage from the HCAL and ECAL. This extends the combined thickness to around twelve interaction lengths. The HE contains around ten interaction lengths. The forward HCAL extends the range from $|\eta| < 2.3$ to $|\eta| < 5.2$ such that very forward jets can be detected. The hermeticity of the detector ensures good coverage on detecting the total hadronic energy and hence good resolution can be obtained on missing transverse energy which could come from neutrinos or BSM particles.

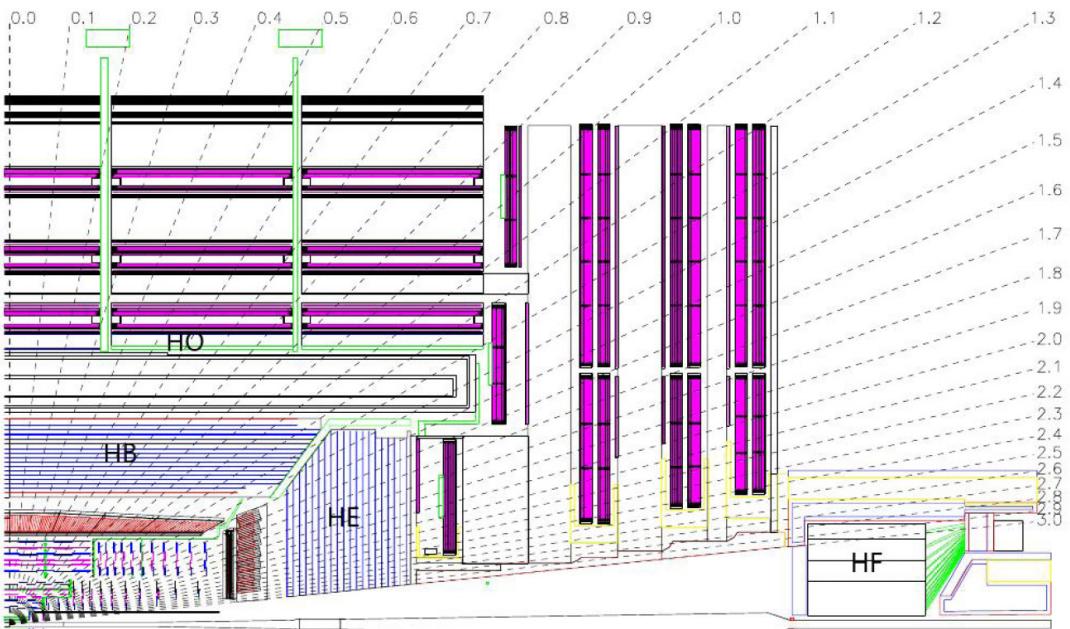


Figure 1.6: The HCAL system [10].

The energy resolution of the HCAL for neutral hadronic particles is $120\%/\sqrt{E(\text{GeV})}$ [11]. Neutral hadronic interactions are the most important consideration for the energy resolution of the HCAL as, unlike charged particles, no additional information can be obtained from the tracker to combine with the calorimetry measurement to reduce the resolution.

1.2.5 Muon Chambers

As the name CMS suggests, muon identification and measurement was a central focus in the design of the detector. The muon chambers are interspersed within the iron flux return yoke. These thick layers of iron act as a hadron absorber. Muons are much less affected by radiative losses through the detector material than

electrons, therefore they are able to penetrate through to the outermost layers of the detector. Figure 1.7 shows the layout of the muon chambers around the detector. The muon chambers consist of a cylindrical barrel section and 2 planar endcaps. The

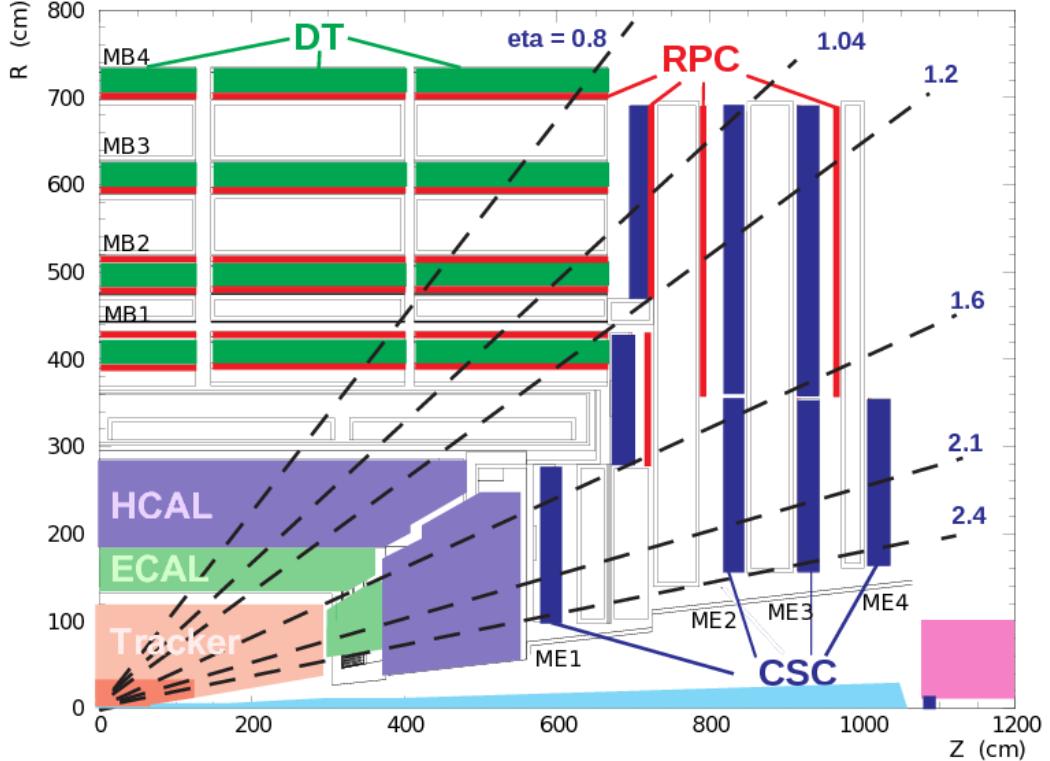


Figure 1.7: The muon chamber system [12].

magnetic field is uniform in the barrel region where drift tubes (DT) are arranged into chambers, some of which make measurements in $r - \phi$ and some of which make measurements in z . This provides a high efficiency for matching individual hits in different stations to one single muon track. In the endcaps, where the muon flux is high and the magnetic field is non-uniform, cathode strip chambers (CSC) are used due to their fine segmentation, fast response and radiation resistance. The CSC stations are aligned perpendicular to the beam line and are positioned between the flux return plates. The cathode strips are positioned as radial lines and the anode wires run perpendicular to the cathode strips. Together they can measure the position in $r - \phi$ and η . The CSCs provide robust pattern recognition for matching hits to other stations and to the tracker as well as for rejecting non-muon backgrounds. Together the barrel and end-cap region provide uninterrupted

coverage up to $|\eta| < 2.4$.

Resistive plate chambers (RPC) are added in the barrel and endcap sections of the muon chambers. They consist of two parallel plate chambers which sandwich readout strips, known together as a double-gap module. The sum of the two signals from each gap creates one total induced signal. There are six RPCs in the barrel section and there are three layers of RPCs in the endcap. Where ambiguous tracks exist due to multiple hits in the muon chambers, the RPCs can help to distinguish the correct track. They have a fast response and can time-tag an ionising event in much less than the 25 ns bunch spacing time and hence they can assign candidate tracks to the relevant bunch crossing. The resolution of the RPCs is courser than the DTs and CSCs. This courser information from the RPCs is used in the trigger which is described in the following section.

The muon chamber measurements provide the dominant contribution to the energy resolution for high-momentum muons. For muons with low momentum the tracker provides the dominant contribution to the energy resolution when the tracks are more curved. At around 1 TeV both systems provide a momentum resolution of $\approx 5\%$. Figure 1.8 shows the muon transverse momentum resolution gained from using only the muon system, only the tracking system, and the combination of the two. It can be seen that above 200 GeV the combination of the information from the muon system with the tracking system improves the overall resolution compared to using the tracking system only.

1.2.6 Trigger

As each of the subdetectors have now been described, the remaining hardware and software which collects the information for analysis can now be discussed. The amount of information which can be stored is much less than the amount produced within the subdetectors. Collisions occur at a rate of 40 MHz (beam crossing interval of 25 ns). The rate reduction capability of the trigger system was designed to reduce the rate by at least a factor of 10^5 . This occurs in two main stages, the Level-1 (L1) trigger and High-Level Trigger (HLT). The L1 trigger is composed of

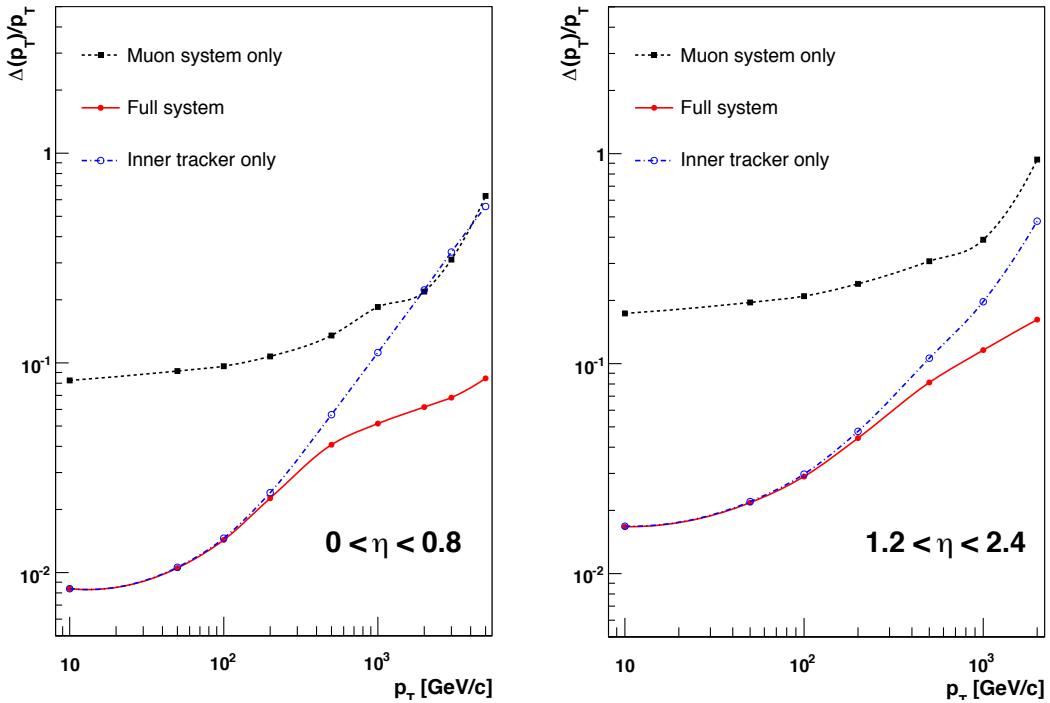


Figure 1.8: The muon transverse momentum resolution as a function of transverse momentum (p_T) using the muon system only (black), the inner tracking only (blue), and both (red), in regions of $|\eta| < 0.8$ (left) and $1.2 < |\eta| < 2.4$ (right) [8].

highly programmable custom electronics, mainly FPGAs. ASIC and programmable memory lookup tables are used where higher speed and radiation hardness is required closer to the beam spot. Full high resolution data is held in the front-end electronics whilst the L1 trigger decides whether or not to keep the event based on courser information from the calorimeters and muon chambers. The HLT is a software based filter farm (≈ 1000 processors) which has access to all of the readout data from the subdetectors. The HLT can assesses more complex information in order to filter out less interesting events and categorise the most interesting ones.

1.2.7 Upgrades for Run 2

In LS1, repairs were made to the LHC magnet splices to allow safe operation at the design energy of $\sqrt{s} = 14$ TeV. All of the detectors on the LHC ring were able to make essential repairs and upgrades. For CMS this included repairing damaged silicon pixels and strips in the tracker and inserting the tracker back into CMS with better centering around the beam line. The temperature of the tracker was

lowered to -20°C to mitigate against radiation damage. In the ECAL the EE and ES subsystems underwent minor repairs. New photodetectors were added to the HO in the HCAL to improve the signal to noise ratio. The main change to the muon systems has been to the RPCS with the addition of the fourth disk (RE4). For the muon system CSCs, 72 chambers have been added to the existing 468 chambers. In addition to the detector upgrades, improvements have been made to the trigger and DAQ.

These upgrades have allowed CMS to operate at $\sqrt{s} = 13$ TeV in 2015 with the smaller bunch spacing of 25 ns compared to the previous 50 ns bunch spacing in Run 1.

1.2.8 Data collections

The data quality monitoring (DQM) group monitor incoming data online by reconstructing a small subset of the data immediately to ascertain whether all sub-detectors in CMS are operational. Offline they go through several stages of DQM as the data are reprocessed and reconstructed. Data which has been collected while all sub-detectors were working, the magnet is at full field and the beam conditions are stable is certified by the DQM group for use in physics analyses. Hence the amount of data used for analyses is less than the total collected, as in Fig. 1.2. Different run ranges are usually separated by short shutdowns for maintenance or they may have different run conditions such as different instantaneous luminosity [13].

The proton-proton collision datasets recorded by CMS and used in this thesis are given in Table 1.1 for the Run 1 data that is used in Chapter ?? and in Table 1.2 for the Run 2 data that is used in Chapter ??.

Dataset	Recorded	$\mathcal{L} \text{ pb}^{-1}$
Single Muon Run A	2012	888
Single Muon Run B	2012	4436
Single Muon Run C	2012	7125
Single Muon Run D	2012	7426
Total		19695
Single Electron Run A	2012	876
Single Electron Run B	2012	4420
Single Electron Run C	2012	7132
Single Electron Run D	2012	7294
Total		19721

Table 1.1: Run 1 datasets at 8 TeV, when they were recorded and how much data was recorded.

Dataset	Recorded	$\mathcal{L} \text{ pb}^{-1}$
Single Muon Run C	2015	17.2
Single Muon Run D	2015	2611.5
Total		2628.7
Single Electron Run C	2015	17.2
Single Electron Run D	2015	2611.5
Total		2628.7

Table 1.2: Run 2 datasets at 13 TeV, when they were recorded and how much data was recorded.

Appendix A | Cross checks on

Run 2 $t\bar{t}t\bar{t}$ analysis at

$\sqrt{s} = 13 \text{ TeV}$

A.1 Comparison of alternative $t\bar{t}$ generators

Figure A.1 shows that the uncertainty from the MADGRAPH AMC@NLO generator is contained within the uncertainty from the MADGRAPH MLM generator, therefore it is conservative to use the MADGRAPH MLM generator as the systematic shape for differences in the BDT distribution due to generator choice.

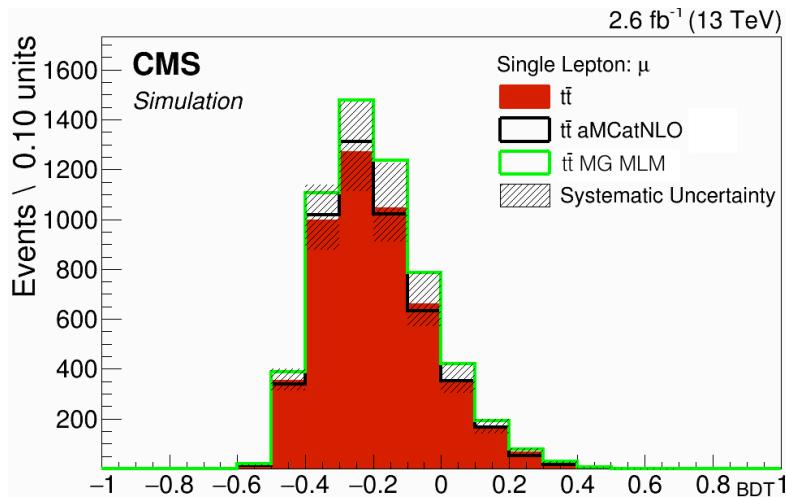


Figure A.1: Inclusive BDT distribution for $t\bar{t}$ generators POWHEG +PYTHIA, MADGRAPH MLM and MADGRAPH AMC@NLO FxFx

A.2 TTZ, TTW, TTH MC backgrounds

The contributions from $t\bar{t} + B$, where $B = W, Z$ or H , were added to the predicted $t\bar{t}$ yields to give a prediction for the net $t\bar{t} + B$ background. The event-level BDT discriminant shapes for these contributions closely follow those of the $t\bar{t}$ contribution

and are very different from those predicted for the $t\bar{t}t\bar{t}$ signal as a function of both the number of jets and the number of b-tagged jets. The differences are small and they are covered by the $t\bar{t}$ scale uncertainties. Therefore, no additional systematic uncertainties were considered necessary to cover these backgrounds.

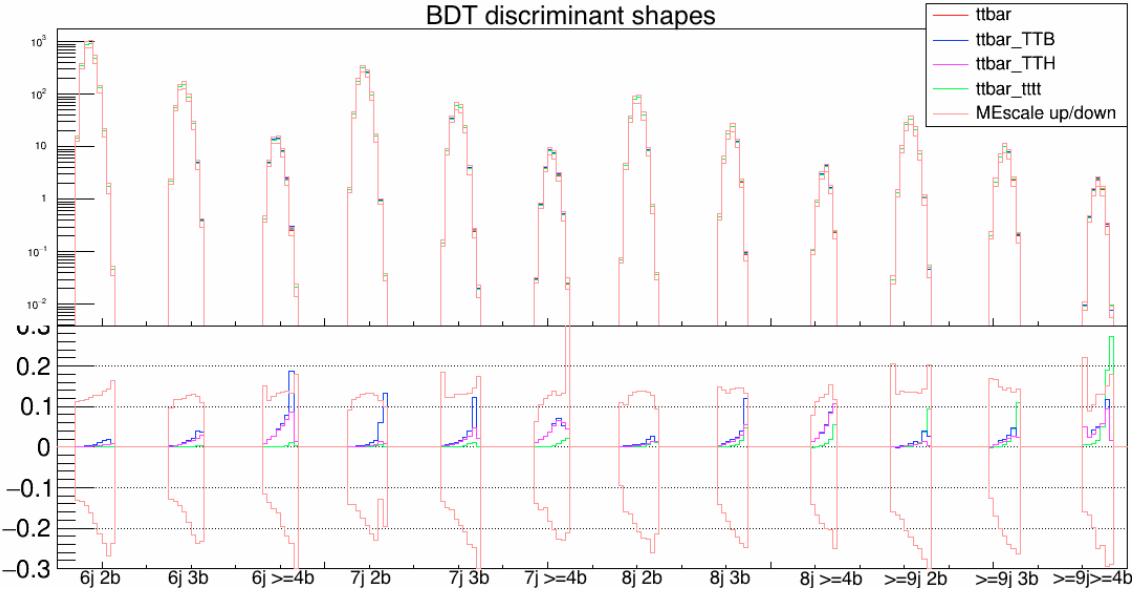


Figure A.2: BDT discriminator shapes for all categories, as indicated along the x axis. The ratio plot shows the difference between each distribution and the nominal $t\bar{t}$ distribution divided by the $t\bar{t}$ distribution.

A.3 Comparison of the Gradient Boost and AdaBoost boosting algorithms within the BDT

For this study, the following three BDTs were trained:

1. *GradNeg* - Gradient boosting taking into account negative weighting information in training and testing
2. *GradBoost* - Gradient boosting ignoring negative weighting information in training and testing
3. *AdaBoost* - AdaBoost boosting ignoring negative weighting information in training and testing

A.4. Event-level BDT templates

Each BDT was trained with the same set of input features and using the same sample of events to train and test the BDTs. The expected limits and uncertainties are shown for each strategy in Table A.1 for the $\mu +$ jets and $e +$ jets final states. Note that this study was performed at an earlier stage in the analysis so the results do not correspond exactly to the final expected limit given in Section ???. The BDT output discriminator distribution was only split into N_{jets} categories of 6, 7, 8, 9+ jets at this stage rather than N_{jets} and N_{tags}^M categories.

Table A.1: Expected limits using jet categories of 6, 7, 8, 9+ jets for different BDT boosting algorithms.

Algorithm	$\mu +$ jets	uncertainty	$e +$ jets	uncertainty
GradNeg	18.1	+8.0, -5.3	27.6	+12.9, -8.3
GradBoost	18.7	+8.3, -5.5	28.8	+12.9, -8.3
AdaBoost	10.7	+6.4, -4.0	21.6	+10.9, -7.0

It can be seen from Table A.1 that the difference between including negative weight information in the GradNeg strategy and not including it in the GradBoost strategy have a negligible effect on the expected limit within the uncertainties. Using negative weights may slightly optimise the modelling for training but not significantly hence the AdaBoost strategy can be used without negative weights as it has a significant benefit in lowering the expected limit.

A.4 Event-level BDT templates

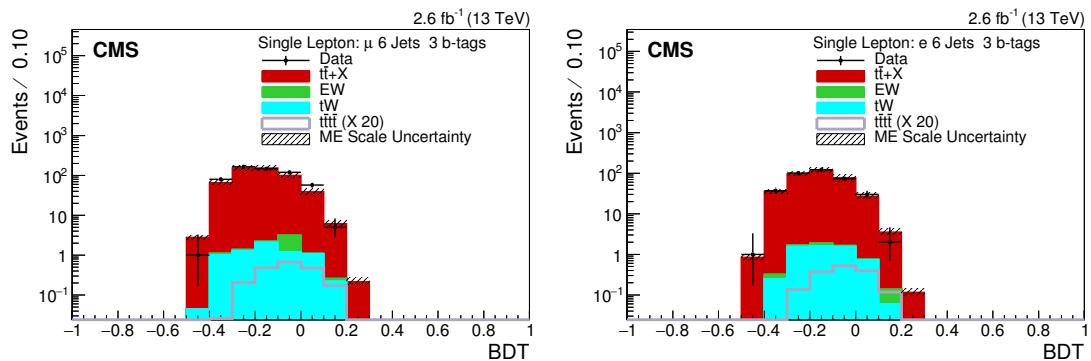


Figure A.3: The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (left) are shown for the 6 N_{jets} and $3N_{\text{tags}}^M$ category.

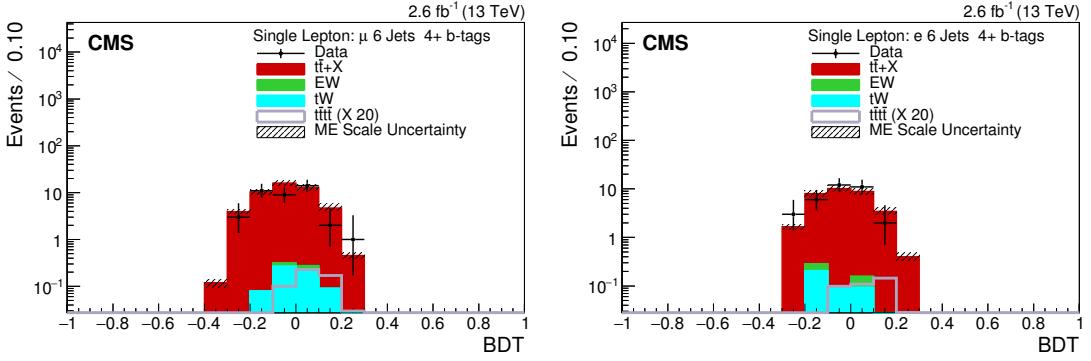


Figure A.4: The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (right) are shown for the $6 N_{\text{jets}}$ and $\geq 4 N_{\text{tags}}^{\text{M}}$ category.

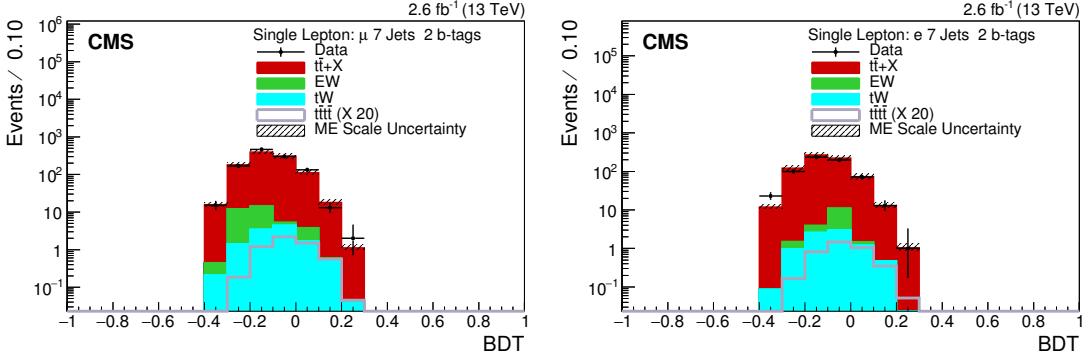


Figure A.5: The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (left) are shown for the $7 N_{\text{jets}}$ and $2 N_{\text{tags}}^{\text{M}}$ category.

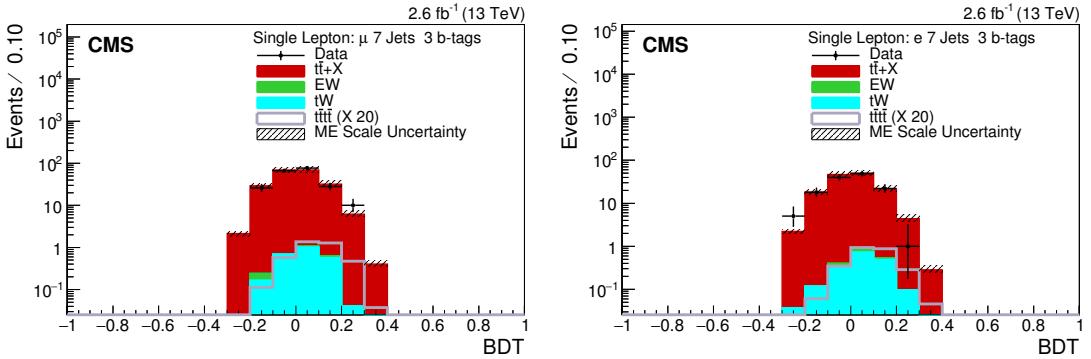


Figure A.6: The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (left) are shown for the $7 N_{\text{jets}}$ and $3 N_{\text{tags}}^{\text{M}}$ category.

A.4. Event-level BDT templates

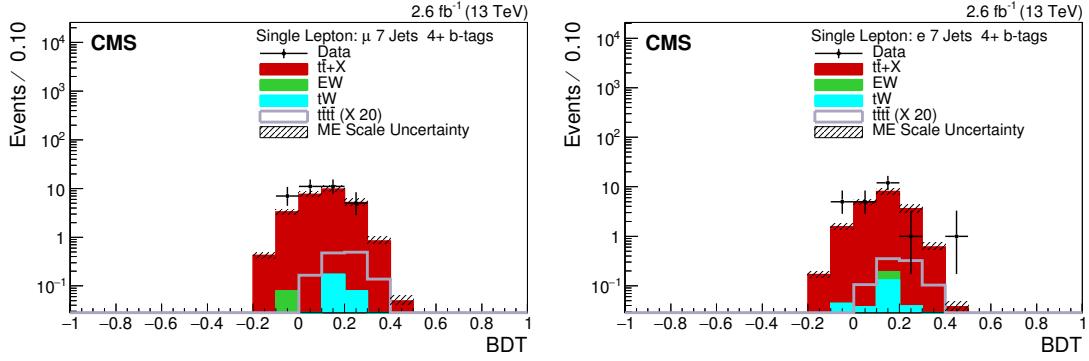


Figure A.7: The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right) are shown for the $7 N_{\text{jets}}$ and $\geq 4 N_{\text{tags}}^{\text{M}}$ category.

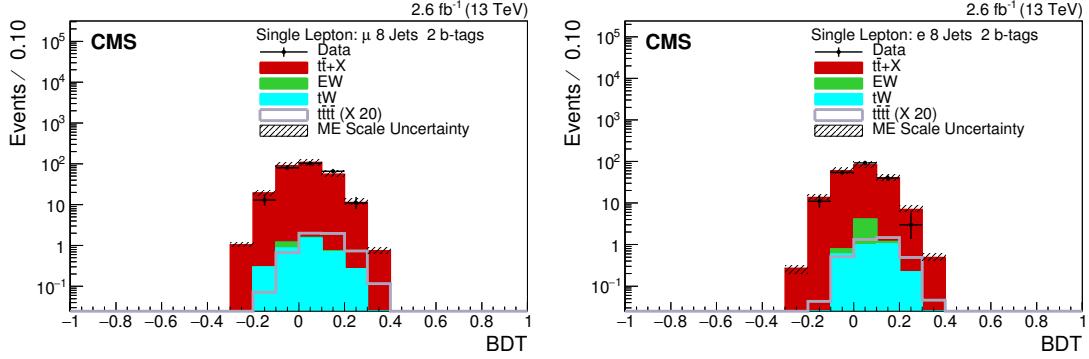


Figure A.8: The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the $8 N_{\text{jets}}$ and $2 N_{\text{tags}}^{\text{M}}$ category.

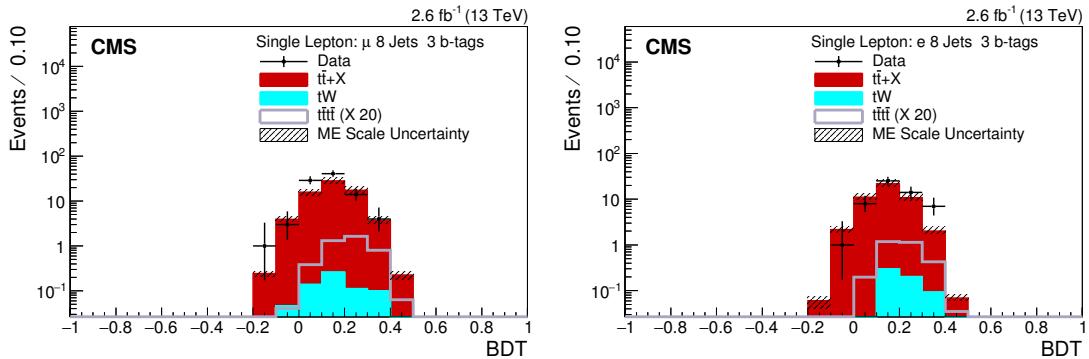


Figure A.9: The BDT output distributions for AdaBoost for data and simulation in the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (left) are shown for the $8 N_{\text{jets}}$ category and $3 N_{\text{tags}}^{\text{M}}$ category.

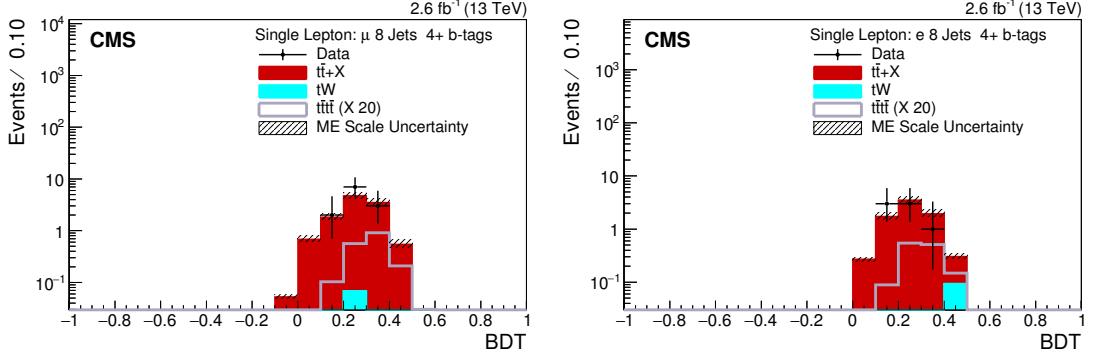


Figure A.10: The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (right) are shown for the 8 N_{jets} and $\geq 4 N_{\text{tags}}^M$ category.

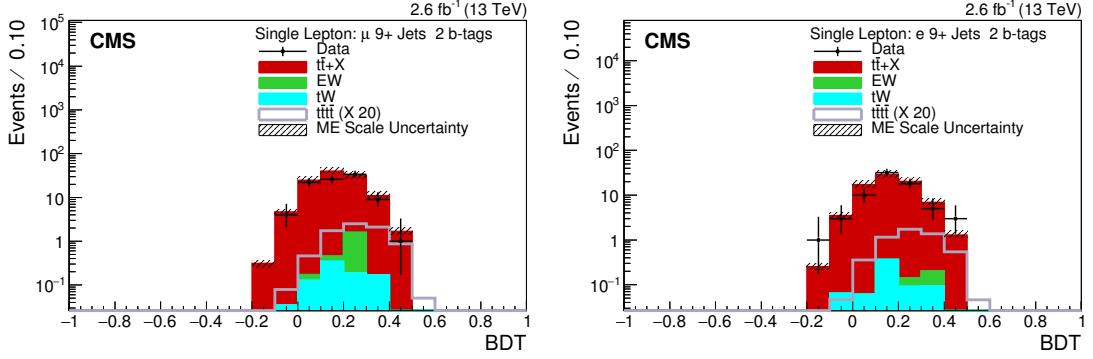


Figure A.11: The BDT output distributions for AdaBoost for data and simulation in the $\mu +$ jets channel (left) and $e +$ jets channel (left) are shown for the $\geq 9 N_{\text{jets}}$ and $2 N_{\text{tags}}^M$ category.

A.5 Systematic shape studies

In this section the alternative BDT distribution shapes are examined for a few of the shape systematic described in Section ???. The largest systematic uncertainties are the JES, ME scale systematics and the $t\bar{t}$ generator choice. The JER and PU up/-down (red/cyan) shapes deviate very little from the nominal distributions (blue). In $t\bar{t}$ there are several distributions including JER and PU systematics show relatively flat behaviour with respect to the nominal distribution and in future analyses could be considered for incorporation into a normalisation systematic uncertainty.

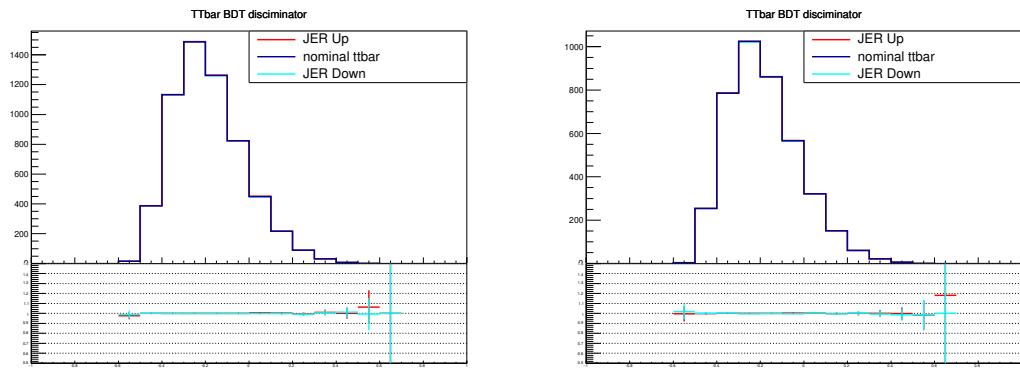


Figure A.12: The BDT shapes for JER systematic in $t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

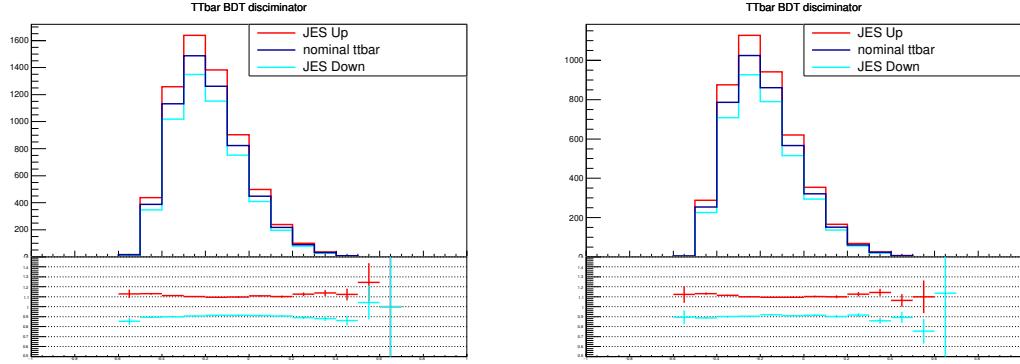


Figure A.13: The BDT shapes for JES systematic in $t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

A.5.1 Studies of impact of systematic uncertainties

The impact of each systematic uncertainty on the expected limit is shown in Table A.2 by removing each systematic from the fit and recalculating the expected limit. The systematic uncertainties which have the largest impact are the $t\bar{t}$ ME

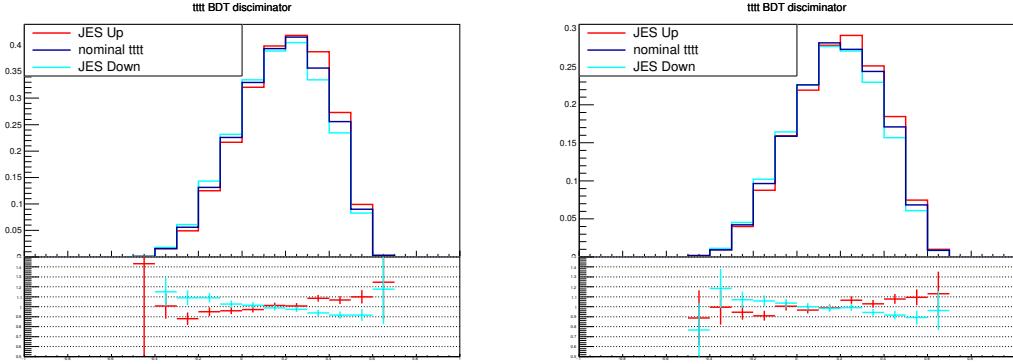


Figure A.14: The BDT shapes for JES systematic in $t\bar{t}t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

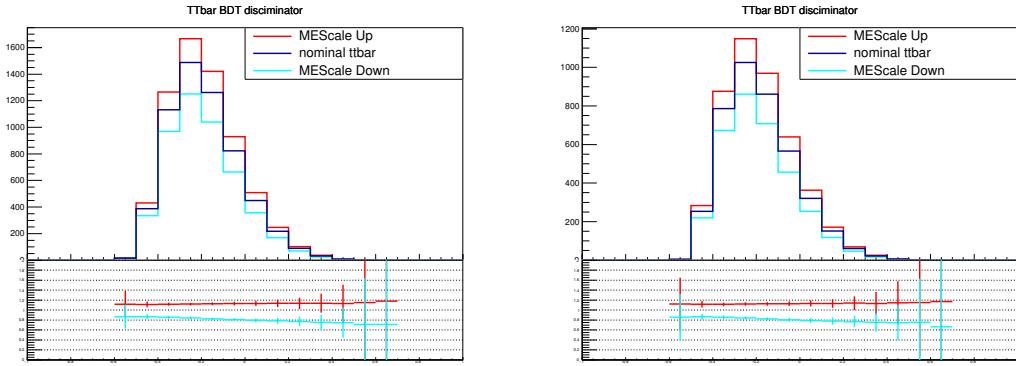


Figure A.15: The BDT shapes for ME scale systematic in $t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

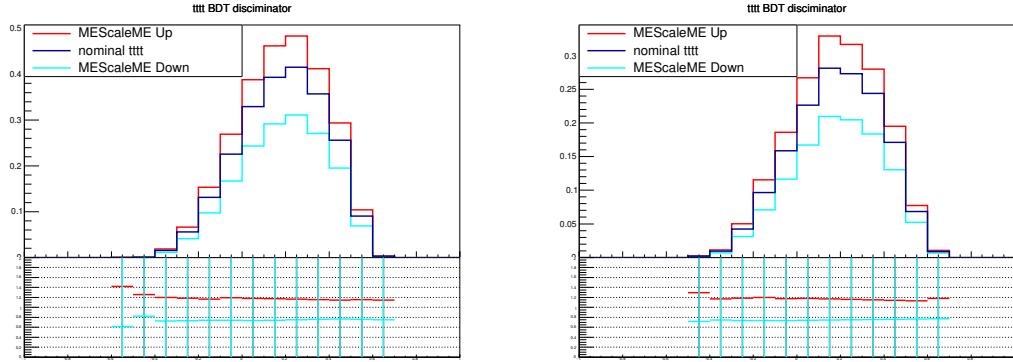


Figure A.16: The BDT shapes for ME scale systematic in $t\bar{t}t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

scale, $t\bar{t}$ PS scale and the JES scale, which is applied to both $t\bar{t}$ and $t\bar{t}t\bar{t}$. The $t\bar{t}$ ME scale has a large impact on the modelling of the signal process whereas the $t\bar{t}$ PS scale has a big effect on the modelling of the additional jets produced in high jet multiplicity $t\bar{t}$ events which pass the baseline event selection.

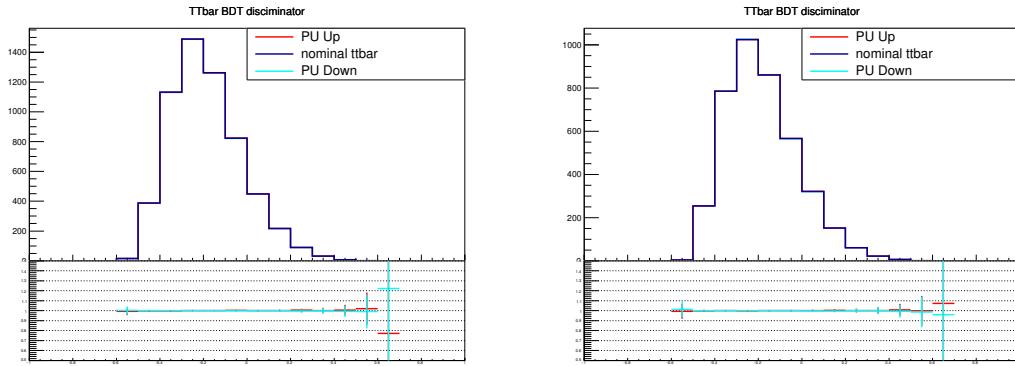


Figure A.17: The BDT shapes for PU systematic in $t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

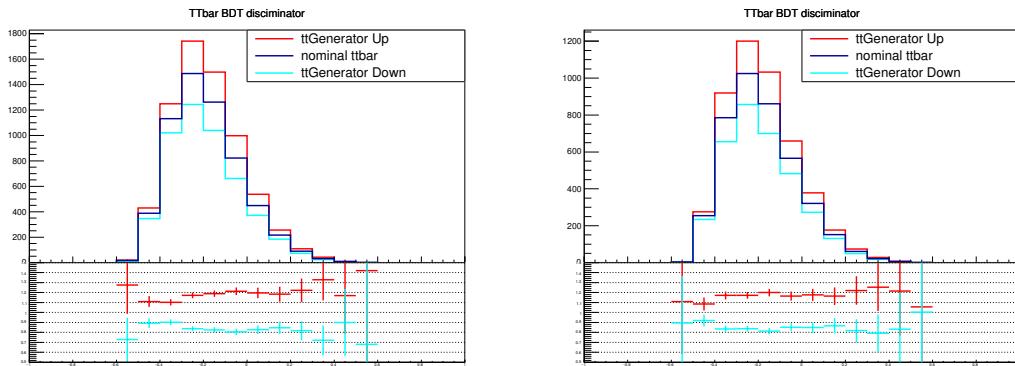


Figure A.18: The BDT shapes for ttGenerator choice systematic in $t\bar{t}$ for the $\mu + \text{jets}$ channel (left) and $e + \text{jets}$ channel (right).

Table A.2: Expected limits on $t\bar{t}t\bar{t}$ production which each systematic removed in turn

Systematic uncertainty removed	Expected limit ($\times \sigma_{t\bar{t}t\bar{t}}^{SM}$)
None	16.0000
$t\bar{t}$ ME scale	16.0625
$t\bar{t}t\bar{t}$ ME scale	14.4375
JER	15.9375
JES	15.3125
PS scale	15.0625
PU	16.0625
Generator uncertainty	15.9531
$t\bar{t}$ heavy flav	15.5625
Luminosity	16.0625
Lepton SF Mu	16.0625
Lepton SF El	15.9062
$t\bar{t}$ norm	16.0000
$t\bar{t}t\bar{t}$ norm	15.9062
EW norm	16.0000
Single top norm	16.0000
btagWeightCSVCFErr1	16.0625
btagWeightCSVCFErr2	16.0625
btagWeightCSVHF	15.5625
btagWeightCSVHFStats1	15.9375
btagWeightCSVHFStats2	15.9531
btagWeightCSVLF	15.9375
btagWeightCSVLFStats1	16.0625
btagWeightCSVLFStats2	15.9062

A.6 Correlation matrices for fit nuisance parameters

The correlation matrix for the fit nuisance parameters in the background only scenario can be seen in Fig. A.19. There is some correlation between the various b-tagging scale factors and also a correlation between the heavy flavour $\sigma_{t\bar{t}b\bar{b}} / \sigma_{t\bar{t}jj}$ modelling and the $t\bar{t}$ ME scale, where the $\sigma_{t\bar{t}b\bar{b}} / \sigma_{t\bar{t}jj}$ is expected to be related to the choice of ME scale.

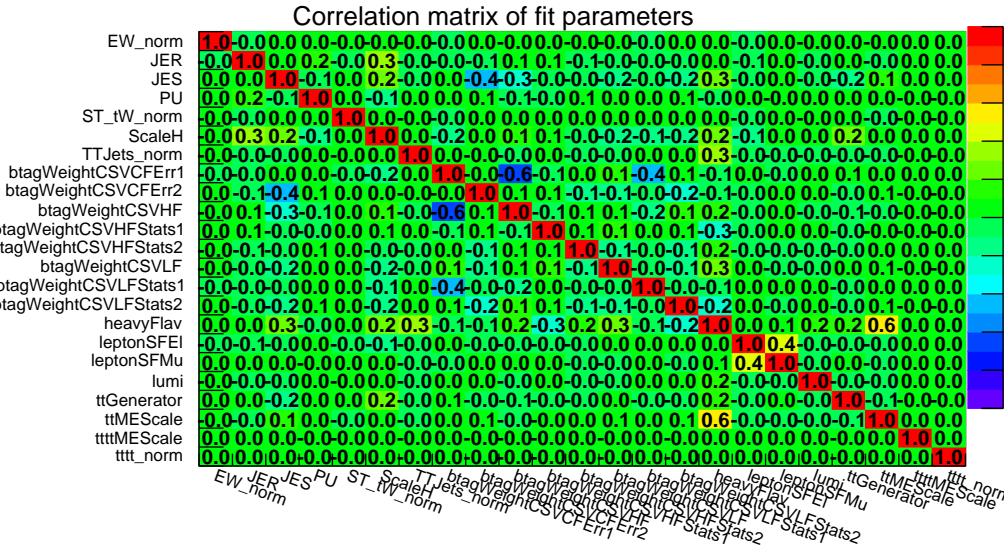


Figure A.19: The correlation matrices for background only for the fit parameters.

References

- [1] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, 2008.
- [2] Fabienne Marcastel. CERN’s Accelerator Complex. La chaÑone des accélérateurs du CERN. Oct 2013. General Photo.
- [3] Cms luminosity - public results.
- [4] G. Aad et al. Observation of a new particle in the search for the standard model higgs boson with the {ATLAS} detector at the {LHC}. *Physics Letters B*, 716(1):1 – 29, 2012.
- [5] S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.
- [6] Tai Sakuma and Thomas McCauley. Detector and event visualization with sketchup at the cms experiment. *Journal of Physics: Conference Series*, 513(2):022032, 2014.
- [7] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, and etal. Cms tracking performance results from early lhc operation. *The European Physical Journal C*, 70(4):1165–1192, 2010.
- [8] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, 2008.
- [9] P Adzic. Energy resolution of the barrel of the cms electromagnetic calorimeter. *Journal of Instrumentation*, 2(04):P04004, 2007.

- [10] Bora Isildak. *Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ tev*. PhD thesis, Bogazici U., 2011.
- [11] Florian Beaudette. The CMS Particle Flow Algorithm. In *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pages 295–304, 2013.
- [12] Min Suk Kim et al. CMS reconstruction improvement for the muon tracking by the RPC chambers. *PoS*, RPC2012:045, 2012. [JINST8,T03001(2013)].
- [13] L Tuura, A Meyer, I Segoni, and G Della Ricca. Cms data quality monitoring: Systems and experiences. *Journal of Physics: Conference Series*, 219(7):072020, 2010.