

BRUNEL UNIVERSITY



DOCTORAL THESIS

**Measurement of the inclusive top-quark pair
plus a radiated photon production cross
section in the dilepton channel in pp
collisions at 8TeV**

Author:

Nik Berry

Supervisor:

Prof. Akram Khan

Prof. Peter Hobson

*A dissertation submitted to Brunel University
in accordance with the requirements
for award of the degree of Doctor of Philosophy*

in

*the Faculty of Particle Physics
Centre for Sensors & Instrumentation*

December 1, 2015

Declaration of Authorship

I, Nik Berry, 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:

(Signature of student)

DPA

"Real courage is when you know you're licked before you begin, but you begin anyway and see it through no matter what."

Harper Lee, To Kill a Mockingbird

Abstract

We present the top-quark pair plus photon production cross section measured in pp collisions at a centre-of-mass energy of 8 TeV with the CMS detector at the Large Hadron Collider, using data recorded in 2012 corresponding to an integrated luminosity of $L_{\text{int}} = 19.6 \text{ fb}^{-1}$. The measurement is performed in the dilepton decay channel. The signal region is defined by the final state of the process $\text{pp} \rightarrow W + W \rightarrow b\bar{b}$, with a minimum photon transverse energy of $E_T(\gamma) > 20 \text{ GeV}$ and minimum distance of $\Delta R(\gamma, b/\bar{b}) > 0.1$ between the photon and the b-quark in $\eta - \phi$ space. Signal events are simulated using the WHIZARD event generator. The normalized cross-section, $R = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}}$, is exploited in order to cancel various sources of systematic uncertainties. The largest contribution to the systematic uncertainty of 17.3 arises due to the modelling of the $t\bar{t}$ background process. We measure the fiducial normalised cross-section, requiring $E_T(\gamma) > 25 \text{ GeV}$ and $|\eta(\gamma)| < 1.4442$ on the final state photon, to be $R^{fid.} = (0.72 \pm 0.04(\text{stat.}) \pm 0.15(\text{syst.})) \times 10^2$. Extrapolated into the signal region, we obtain $R = (0.89 \pm 0.05(\text{stat.}) \pm 0.18(\text{syst.})) \times 10^2$. Using a recent CMS $t\bar{t}$ cross-section measurement at 8 TeV, we calculate the top pair plus photon production cross-section to be $\sigma_{t\bar{t}+\gamma}^{CMS} = 2.0 \pm 0.1(\text{stat.}) \pm 0.4(\text{syst.})$. Being in agreement with the $t\bar{t} + \gamma$ SM expectation of $\sigma_{t\bar{t}+\gamma}^{SM} = 1.8 \pm 0.5 \text{ pb}$, this is the most accurate measurement of the $t\bar{t} + \gamma$ process to date, and the first at a center-of-mass energy of 8 TeV.

The acknowledgements and the people to thank go here, don't forget to include your project advisor...



Contents

Introduction	1
1 Theoretical Motivation	2
2 The LHC and the CMS Detector	4
2.1 The Large Hadron Collider	4
2.1.1 Pre-LHC accelerator complex	4
2.1.2 Design of the LHC	6
2.1.3 Physics goals	6
2.1.4 Luminosity at the LHC	7
2.1.5 Performance throughout run I	8
2.2 The CMS Detector	9
2.3 Inner Tracking System	11
2.3.1 Tracker performance in Run I	13
2.4 Electromagnetic Calorimeter	15
2.4.1 Overview	15
2.4.2 Composition of the ECAL	16
2.4.3 Photodetectors	18
2.4.4 Performance of the ECAL throughout Run I	18
2.5 Hadron Calorimeter	19
2.5.1 Overview	19
2.5.2 Performance of the HCAL in Run I	20
2.6 Superconducting Solenoid	20
2.7 Muon System	22
2.7.1 Performance of the muon system in Run I	24
2.8 Trigger and Data Acquisition	24

2.8.1	Performance of the Trigger in Run 1	26
3	Event Reconstruction & Simulation	27
3.1	Event Reconstruction	29
3.2	Computing	29
3.2.1	Event Data Model	29
3.2.2	Analysis Software	30
3.3	Simulation of the CMS detector	30
3.4	Monte Carlo Simulation	31
3.4.1	Monte Carlo event generators	31
3.5	Simulated samples for the $t\bar{t} + \gamma$ analysis	33
3.6	Simulation of the $t\bar{t} + \gamma$ Signal Sample	34
3.6.1	Official $t\bar{t} + \gamma$ $2 \rightarrow 7$ process MADGRAPH Sample Production	37
3.7	Physics Object Reconstruction	37
3.7.1	Charged Particle Tracking	38
3.7.2	Primary Vertex Reconstruction	39
3.7.3	Calorimeter Clustering Algorithm	40
3.7.4	The Particle Flow Algorithm	40
3.7.5	Electron reconstruction	40
3.7.6	Electron identification	42
3.7.7	Muon reconstruction	44
3.8	Higher-level Object Reconstruction	45
3.8.1	Jet reconstruction	45
3.8.2	Jet Identification	47
3.8.3	Jet energy corrections	47
3.8.4	B-tagged Jets	48
3.8.5	Particle flow jet identification	51
3.8.6	Reconstruction of Missing Transverse Energy	51
4	Event Selection & Background Estimation	53
4.1	Event Selection	53
4.2	Pre-selection: Selection of $t\bar{t}$ Events	54
4.3	Trigger and Event Cleaning	54
4.3.1	Trigger selection	54
4.3.2	Filtering	56

4.4	Di-lepton Selection and Veto	57
4.4.1	Electrons	57
4.4.2	Muons	59
4.4.3	Corrections to simulated events	60
4.5	Jet Selection and b-tag Requirements	60
4.6	Selection of $t\bar{t} + \gamma$ events	61
4.6.1	Cut based photon ID	62
4.6.2	Super-cluster footprint-removal for photon isolation	63
4.7	Background Estimation	63
5	Measurement of the inclusive $t\bar{t} + \gamma$ cross-section	66
5.1	Signal Definition and Background Processes	66
5.1.1	Signal definition	66
5.1.2	Background processes	66
5.1.3	Backgrounds of photon signature	66
5.2	$t\bar{t} + \gamma$ Signal Simulation	66
5.3	Phase Space Overlap Removal	66
5.4	Event Selection	71
5.5	Photon Purity Estimation	71
5.6	Signal Acceptance Calculation	71
6	Systematic Uncertainties	74
6.1	Flat Rate Uncertainties	75
6.1.1	Luminosity	75
6.1.2	Lepton Efficiencies	76
6.2	Shape Uncertainties	76
6.2.1	Parton Density Function	76
6.2.2	Pile-up Re-weighting	76
6.2.3	Jet Energy Corrections	77
6.2.4	Missing Transverse Energy	77
6.2.5	B-tagging Efficiency	78
6.2.6	Data-driven Reweighting	78
6.3	Modelling Uncertainties	78
6.3.1	QCD Renormalisation and factorisation scales	78
6.3.2	Parton-level matching thresholds	78

6.3.3	Analysis-dependent modelling uncertainties	78
6.4	Impact of uncertainties	78
7	Results	
8	Measurement of the anomalous couplings of the photon to the top quark	83
9	Electron Conversion Veto	84
Conclusions		85
Appendices		87
A.1	Electron & Muon Efficiencies	87
Bibliography		94

List of Figures

1.1	Top-photon vertex. Left: Leading Order (LO). Right: One-Loop correction (NLO)	2
1.2	The Standard Model of particle physics.	3
2.1	A full schematic of the full CERN accelerator complex.	5
2.2	(Left) Peak and (right) integrated luminosity recorded by the LHC between 2010 and 2012 for proton operation. The 2010 luminosity values have been multiplied by a factor 20 [1].	9
2.3	Left: The accumulation of the integrated luminosity produced at the LHC vs time for runs in 2010, 2011, and 2012. The 2010 integrated luminosity is multiplied by 100 in order for it to be visible on the plot. Right: Total integrated luminosity vs time for the 2012 run in CMS and the LHC.	10
2.4	A cross-sectional view of the CMS detector.	10
2.5	The sub-detectors of the CMS silicon tracker system: TOB=outer barrel, TIB=inner barrel, TID=inner disc, TEC=endcaps, PIXEL=pixeldetector. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits. [2].	12
2.6	Muon reconstruction efficiency in thee tracker as functions of pseudorapidity (left) and the number of proton-proton interaction vertices (right) [3].	14
2.7	Primary vertex resolution in the transverse plane (left) and along the beam-line (right) as functions of the number of tracks attached to the vertex [3].	14
2.8	Geometric view of one quarter of the ECAL (top). Layout of the CMS electromagnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (bottom) [2].	16

2.9	Geometric view of one quarter of the ECAL (top). Layout of the CMS electromagnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (bottom) [2].	17
2.10	Simulation of fractional response from EE as a function of η , for different integrated luminosities. Right: Deterioration of the energy resolution in EE as a function of η , for different integrated luminosities [4].	18
2.11	Longitudinal view of one quarter of the detector in the $r - \eta$ - plane, showing the positions of the HCAL parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF) [2].	20
2.12	General artistic view of the 5 modules composing the cold mass inside the cryostat, with details of the supporting system (vertical, radial and longitudinal tie rods) [2].	22
2.13	Layout of one quadrant of CMS. The figure shows the four DT stations in the barrel (MB1-MB4, yellow), the four CSC stations in the endcap (ME1-ME4, green), and the RPC stations (RB1-RB4 and RE1-RE3) [2].	23
2.14	Architecture of the Level 1 trigger [2].	25
3.1	Schematic overview of a hadron collision process [5].	28
3.2	A radiography of a quarter of the simulated tracker geometry in the (a) fast and (b) full simulation [6].	31
3.3	Process generation. The red, blue, and green boxes depict the matrix element calculation. Background processes with the same final state are included as well [7].	36
3.4	Schematic showing a displaced secondary vertex due to a b-quark decay from the primary vertex forming a b-jet. The impact parameter, d_0 , measures the displacement with respect to the primary vertex along the z-axis, and L_{xy} measures the displacement from the primary vertex in the transverse plane.	49
3.5	Logarithmic distribution of the partons as a function of the CSV b-tagging discriminant [8].	50
4.1	63
4.2	65
5.1	Graphic representation of the signal and background definitions [?].	67

5.2 Examples of Feynman graphs of the considered $t\bar{t}$ backgrounds. QCD showering is necessary in all cases to obtain the required jet multiplicity [9].	68
5.3 Flow chart showing each stage of the analysis. The box numbers represent the outlined analysis steps.	69
5.4 Background of photon identification: Initial state radiation (ISR). Left: Quark fusion. Right: Gluon fusion does not give rise to ISR (based on [10]).	70
5.5 Background of photon identification: Final state radiation (FSR). All charged particles in the top-pair decay tree contribute to FSR [10].	70
5.6	71

List of Tables

2.1	LHC design parameters [11].	7
2.2	Tower segmentation in azimuthal and polar angle for the hadronic barrel, end-cap and forward calorimeter [12].	19
2.3	Parameters of the LHC superconducting solenoid [13].	21
3.1	Dataset information for signal and background MC samples.	35
3.2	Variables used in electron identification algorithms [14].	43
4.1	Triggers for each di-lepton channel.	55
4.2	Dataset information for each run in each respective decay channel.	64
6.1	Systematic uncertainties and their contribution to the cross-section ratio.	79
6.2	80
7.1	Values used for calculating the cross section ratios in the $\mu^+\mu^-$, e^+e^- , and $e\mu$ channels.	81
7.2	82
A.1	Electron ID & isolation efficiencies [15].	87
A.1	Muon ID & isolation efficiencies [16].	87

Introduction

DRAFT

Chapter 1

Theoretical Motivation

DGLAP Higgs related to top interaction lagrangian with effective operators electric dipole moment and a top quark with 4/3 charge anomolous couplings talk about top quarks - different chapter maybe?

$$\mathcal{L}_{t\bar{t}\gamma} = -eQ_t\bar{t}\gamma^\mu tA_\mu - e\bar{t}\frac{i\sigma^{\mu\nu}q_\nu}{m_t}(d_V^\gamma + id_A^\gamma\gamma^5)tA_\mu \quad (1.1)$$

$$\delta d_V^\gamma = \frac{\sqrt{2}}{e}\text{Re}[c_W C_{uB\phi}^{33} + s_W C_{uW}^{33}] \frac{vm_t}{2} \quad (1.2)$$

$$\delta d_A^\gamma = \frac{\sqrt{2}}{e}\text{Im}[c_W C_{uB\phi}^{33} + s_W C_{uW}^{33}] \frac{vm_t}{2} \quad (1.3)$$

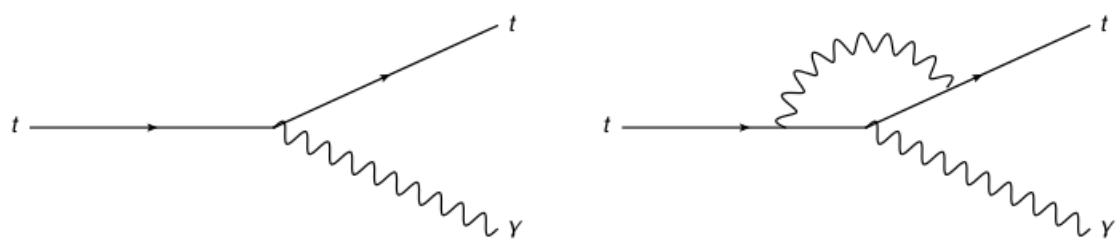


Figure 1.1: Top-photon vertex. Left: Leading Order (LO). Right: One-Loop correction (NLO).

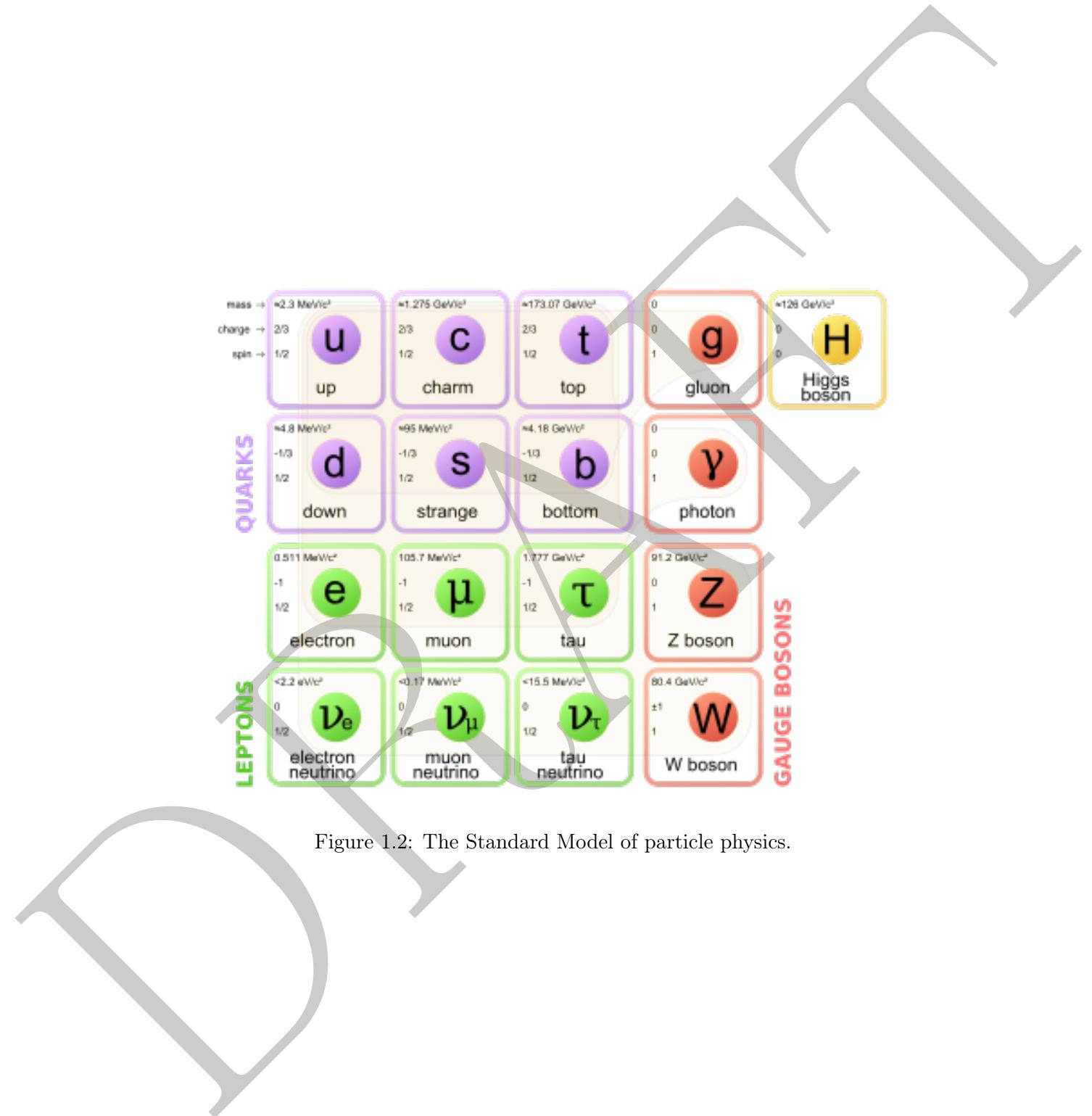


Figure 1.2: The Standard Model of particle physics.

Chapter 2

The LHC and the CMS Detector

2.1 The Large Hadron Collider

The Large HAdron Collider (LHC) is currently the largest, and highest energy, particle accelerator ever created. Located, on average, one hundred metres under the Franco-Swiss border at Geneva, the LHC is installed in the 26.7 km tunnel that once contained the Large Electron-Positron Collider (LEP) which ran from 1989 until the end of 2000. The project was approved by the CERN council in December of 1994. Originally, the accelerator was designed as a two-stage project: constructed to run at a centre-of-mass energy of $\sqrt{s} = 7$ TeV, and later an upgrade to $\sqrt{s} = 14$ TeV. This was due to budget constraints which did not include contributions from non member states.

After many setbacks, the first run began in 2010 and continued until the end of 2011 when the beam energy was then increased to $\sqrt{s} = 8$ TeV for the whole of 2012 before shifting to Long Shutdown 1 (LS1) from 2013 to 2015. During LS1 the CERN accelerator complex, shown in Figure 2.1, was completely upgraded in order to run at a new unprecedented centre-of-mass energy of $\sqrt{s} = 13$ TeV before ramping up to the original design energy of $\sqrt{s} = 14$ TeV.

2.1.1 Pre-LHC accelerator complex

The proton acceleration process begins by injecting Hydrogen (H_2) gas into a Duoplasmatron surrounded by an electric field, whereby the electrons become ionised through interactions with the free electrons from the cathode forming a plasma. This strips the electrons from the Hydrogen leaving just the protons. The remaining protons are then linearly accelerated by the LINAC 2 accelerator, which uses radio frequency (RF) cavities to accelerate bunches

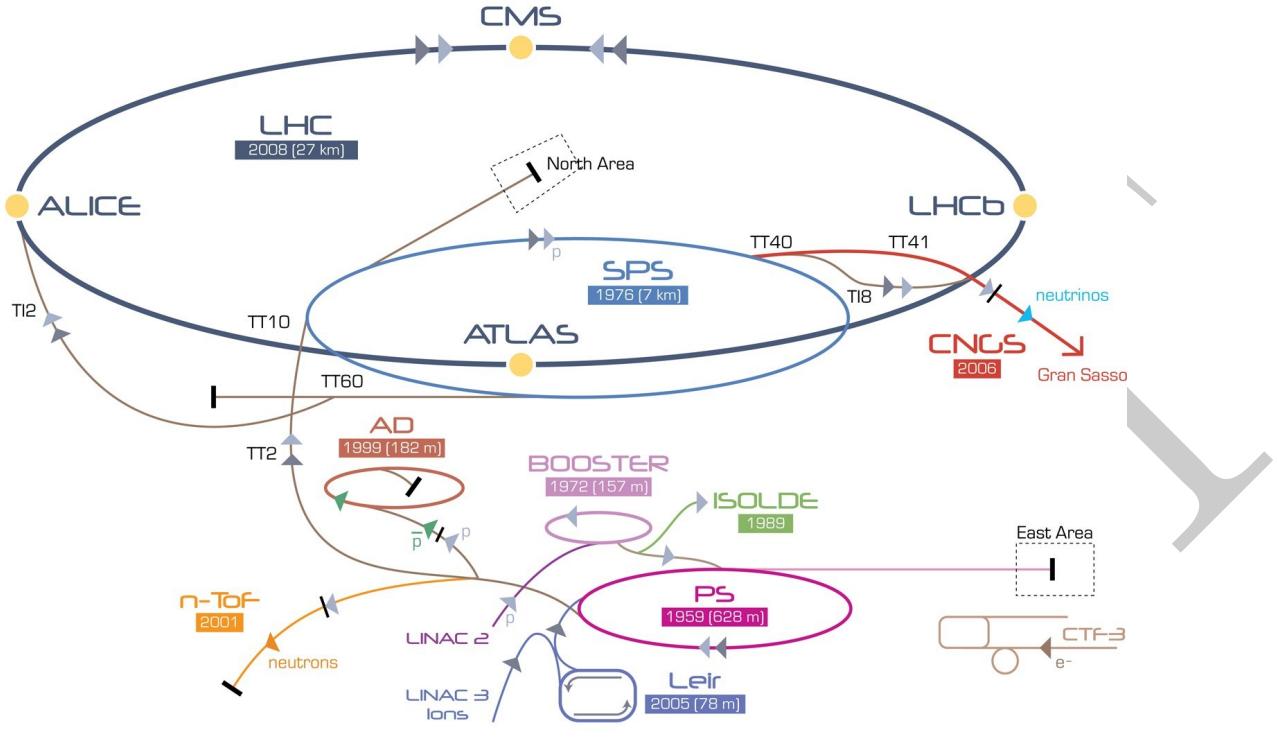


Figure 2.1: A full schematic of the full CERN accelerator complex.

of protons. By the end of this step the protons have reached an energy of up to 50 MeV and gained 5% in mass. The next stage in the sequence sees the protons enter the Proton Synchrotron Booster (PSB) which is composed of superimposed synchrotron rings which accelerate the received protons up to 1.4 GeV in 1.2 seconds before injection into the Proton Synchrotron (PS). The advantage of the Booster is that it allows the PS to accept over 100 times more protons by squeezing the proton bunches such that they have a much smaller cross-section.

The PS is an essential component in the accelerator complex at CERN, where it accelerates either protons received from the PSB, or heavy ions from the Low Energy Ion Ring (LEIR). The apparatus first ran on the 24th of November 1959, and was, at that time, the world's highest energy particle accelerator. Having a circumference of 628 metres, the PS comprises 277 conventional (room temperature) electromagnets, as well as 100 dipole magnets that serve to bend the beam around the ring. The PS accelerates protons, as well as other particles, up to 25 GeV in 3.2 s. The final stage of acceleration, before injection into the LHC, lies in the Super Proton Synchrotron (SPS). The SPS is the second largest of the CERN accelerators with a circumference of 7 kilometers, and provides beams for various experiments other than LHC: such as the NA61/SHINE and NA62 experiments, the COMPASS experiment, and the CNGS neutrino experiment. Protons are accelerated to 450 GeV in 20 s within the SPS before

injection into the LHC. Before the creation of LEP or the LHC the SPS was the primary collider at CERN, and in 1983 the collaboration won the Nobel prize for the discovery of the W and Z bosons in proton-antiproton collisions. The SPS comprises 1317 conventional electromagnets and 744 dipoles.

2.1.2 Design of the LHC

Two beams of protons are injected into the LHC and accelerated using RF cavities, one clockwise and the other counter-clockwise, taking roughly 20 minutes for each beam to reach the design energy of 7 TeV per beam. The two beams come into collision at four points around the ~ 27 km ring where the collisions are recorded by the four detectors placed on the beam line. There are two all-purpose discovery detectors, namely CMS and ATLAS, studies of mesons by LHCb, and the study of heavy ions by the ALICE experiment. Because the tunnel in which the LHC is placed was designed for LEP it has an internal diameter of only 3.7 m, which is not large enough to install two separate beam pipes, and thus a design for a twin-bore magnet [17] was created which would save space and cut costs substantially. Each beam is designed to hold 2808 bunches of protons with a bunch spacing of 50 ns. The protons are guided around the ring in a vacuum by superconducting electromagnets which are cooled to 1.9 K (-271.3°) by using liquid helium. It consists of 1232 dipole magnets that are each 15 metres in length, and 392 quadrupole magnets that are 5-7 metres in length each which focus the beams. Before collisions can begin, a final shaping and cleaning of the beam takes place. Parameters for the LHC can be seen in Table 2.1.2.

2.1.3 Physics goals

There are many physics goals aimed to be achieved during the running of the LHC, but there are certain aims that are of a higher priority than others. One of the main focuses was the discovery of the Higgs boson and electroweak symmetry breaking, which was announced on the fourth of July 2012. This discovery was a triumph for the physics community in that it shed light on a fundamental building block of the universe which was theorised to exist some sixty years before its discovery. The theoretical physicist Peter Higgs subsequently won the Nobel prize for his work predicting the existence of a massive gauge boson in 1964. The Higgs a since been measured in various decay channels by both they ATLAS and CMS experiments with on-going studies aiming to measure properties of the boson, such as a the spin. Other physics goals include the search for supersymmetry, CP violation measurements, and studies

LHC Parameters		
Parameter	2012 Run	Design Value
Beam Energy (TeV)	4	7
Maximum number of bunches	1380	2808
Number of particles per bunch	1.7×10^{11}	1.15×10^{11}
Bunch spacing (ns)	50	25
Revolution frequency (kHz)	11.245	11.245
Transverse beam size (μm)	18.8	16.6
Peak luminosity ($\text{cm}^{-2}\text{s}^{-1}$)	7.7×10^{33}	10^{34}
Stored beam energy (MJ)	140	362
Normalised emittance at start of fill (mm mrad)	2.5	3.75
β^* in IP 1 and 5 (m)	0.6	0.55

Table 2.1: LHC design parameters [11].

of quark-gluon plasma using the ALICE experiment.

2.1.4 Luminosity at the LHC

Due to the nature of individual detectors, not all require the same levels of delivered luminosity. For example, with CMS being an all-purpose discovery machine, the detector needs as much luminosity as possible, however an experiment like LHCb that measure mesons that are produced frequently and in a certain portion of the solid angle that the others use, less luminosity is required. The peak design luminosities for Run I and Run II are listed in Table 2.1.2. The instantaneous luminosity of a collider is calculated as

$$\mathcal{L} = f \frac{N_1 N_2}{4\pi \sigma_x \sigma_y} \quad (2.1)$$

where f is the collision frequency given by $f = u \times N_b$, the repetition frequency multiplied by the number of bunches in the beam, $N_{1,2}$ are the number of protons per bunch per beam, and $\sigma_{x,y}$ are the horizontal and vertical beam sizes at the interaction point (IP), respectively, and are defined as the product of the beams beta function and the proton beam emittance as shown in Equation 2.2.

$$\sigma_{x,y} = \epsilon_{x,y} \beta_{x,y} \quad (2.2)$$

The emittance of a beam describes the volume of the 6-dimensional phase space occupied by the proton bunch.

2.1.5 Performance throughout run I

Throughout Run I (2010 - 2013) the LHC operated with protons at beam energies of 3.5 and 4 TeV, where the beams consisted of single bunches and trains with different bunch spacing of 150 ns (2010), 75 ns (2011), and 50 ns (2011 and 2012). The performance of the LHC was much greater than initially expected at 50 ns, and culminated in the discovery of a $125\text{GeV}/c^2$ Higgs boson in both the ATLAS [18] and CMS [19] experiments. The use of 25 ns bunch spacing was only implemented in regards to electron-cloud scrubbing runs at the injection stage, and also for tests of future collisions with an upgraded LHC energy. One of the main focuses was to reduce the β^* - the measure of how precisely the beam is focused at the interaction point. For ATLAS and CMS β^* was lowered in steps from 3.5 mm in 2010 to 0.6 mm in 2012 by using tighter collimator settings. Other runs with mixed particle beams were also performed: such as proton-Pb, Pb-Pb, intermediate proton energy (1.38 TeV), and high beta.

For the 2012 run the default filling scheme introduced 1374 proton bunches per beam with 50 ns bunch spacing, giving ATLAS and CMS 1368 colliding bunches, 1262 in LHCb, and no colliding bunches in ALICE. The bunch intensity per beam peaked at 1.7×10^{11} protons per bunch, which was then translated into a bunch intensity of 1.6×10^{11} protons per bunch upon stabilisation of the beams. The transverse emittance remained constant throughout the year, despite moving to a different optical configuration with a lower transition energy. At the end of the runs the LHC had delivered an integrated luminosity of 23.3 fb^{-1} to ATLAS and CMS, and over 2.1 fb^{-1} to LHCb. The integrated and peak luminosity can be seen in Figure 2.1.5, and the integrated luminosity recorded by CMS between 2011 and 2013, and also compared to the total integrated luminosity delivered by the LHC, can be seen in Figure 2.1.5.

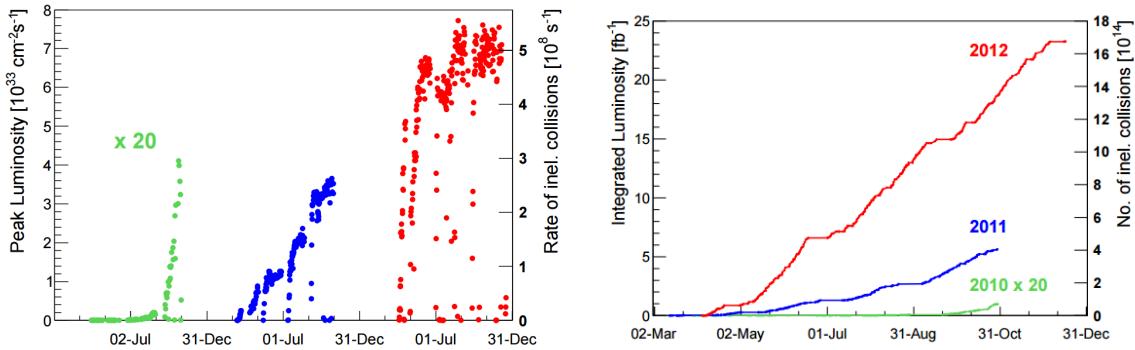


Figure 2.2: (Left) Peak and (right) integrated luminosity recorded by the LHC between 2010 and 2012 for proton operation. The 2010 luminosity values have been multiplied by a factor 20 [1].

2.2 The CMS Detector

The Compact Muon Solenoid (CMS) [2] is one of the two all-purpose discovery machines located approximately 100 metres underground at point 5 (Cessy, France) on the LHC ring. Designed to cover the full solid angle, the hermetic detector is composed of multiple sub-detectors, described in detail in the following sections, designed to perform precision particle detection and withstand extremely high doses of radiation. Unlike the other detectors that lie on the LHC ring CMS is designed with the purpose of precision measurements of Standard Model measurements and the discovery of physics beyond that of the Standard Model. The primary physics motivation for the construction of such a detector was to elucidate the nature of electroweak symmetry breaking of which the Higgs field was theorised to be responsible, which was proved correct in 2012 with the discovery of the quanta that propagates the Higgs field - the Higgs boson. Many theories predict to observe new physics at the TeV scale, and so CMS was designed with the intention to be able to withstand high energy and fluence of particles. Discovering physics beyond the Standard Model would pave the way for a potential unified theory. The detector weighs around 14,000 tonnes and has an overall length of 28.7 metres and diameter of 15 metres. A sectional view of the CMS detector labeling each sub-detector within is shown in Figure 2.2. CMS uses a right-handed coordinate system whereby the x-axis points towards the center of the LHC ring, the y-axis lies perpendicular to the beam, and the z-axis follows the direction of the beam anti-clockwise. The azimuthal angle, ϕ , is measured from the x-axis in the xy plane where the radial component in this plane is

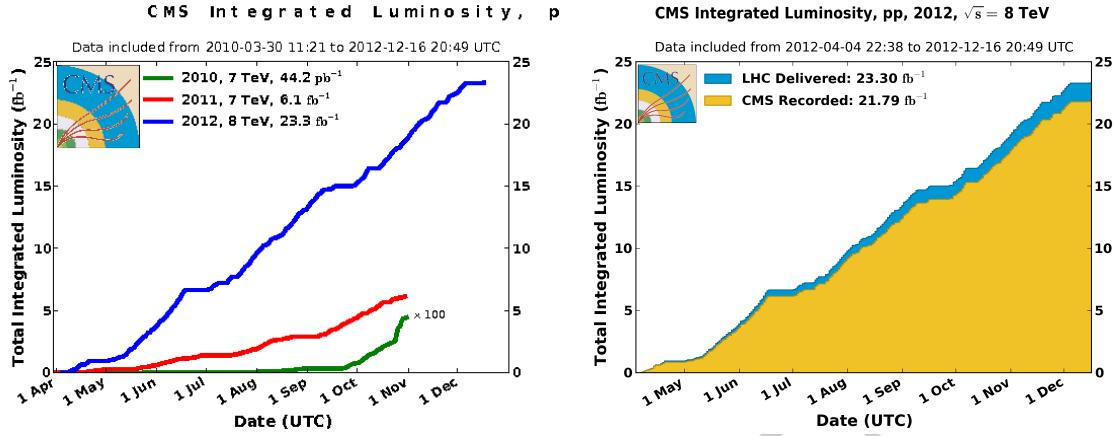


Figure 2.3: Left: The accumulation of the integrated luminosity produced at the LHC vs time for runs in 2010, 2011, and 2012. The 2010 integrated luminosity is multiplied by 100 in order for it to be visible on the plot. Right: Total integrated luminosity vs time for the 2012 run in CMS and the LHC.

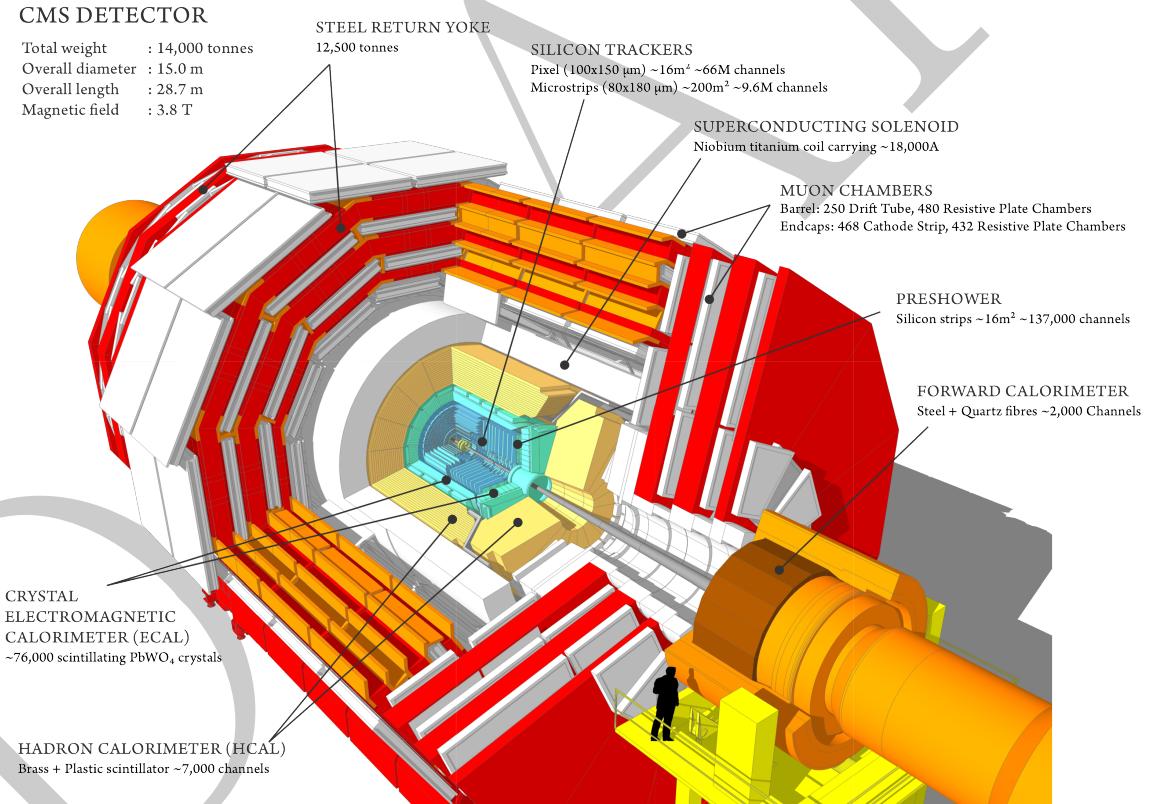


Figure 2.4: A cross-sectional view of the CMS detector.

define by r , and the polar angle θ in the rz plane. The pseudorapidity is thus defined as

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right) \quad (2.3)$$

and the momentum transverse to the beam is defined as p_T , and calculated using the x- and y-components. The transverse energy is defined as $E_T = E \sin \theta$.

2.3 Inner Tracking System

The first sub-detector system located closest to the beam is the Inner Tracking System. The Inner Tracking System is composed of several modules that work in conjunction to provide precise and efficient measurements of the trajectories of charged particles resulting from the beam collisions, as well as a precision reconstruction of secondary vertices, whereby the product from the LHC beam collision decays. The tracker is completely hermetic around the interaction point (IP) of the beam-line, is 5.8 m in length, and has a diameter of 2.5 m. In order to reconstruct particle tracks momentum measurements must be made. To do this the tracker works in combination with the CMS Superconducting Solenoid (Section 2.6) with a magnetic field at 4 T.

Due to the high flux of the LHC at design luminosity the inner tracker will receive around 1000 particles per bunch crossing with around 20 primary vertices per collision, therefore the tracker was designed to operate with a high granularity and fast response time such that trajectories can be precisely identified and associated with the correct bunch crossing. Several challenges arise upon implementation of such technology: the requirement of high power density to the on-detector electronics means that sufficient cooling must be used throughout, which then conflicts with the ideology of keeping material to a minimum to prevent effects such as multiple scattering, bremsstrahlung, photon conversion, and nuclear scattering. Another challenge presents itself in the form of radiation damage to the tracking system due to the large flux of high energy particles over time. The requirements for a high granularity detector using minimum material that can run over a period of roughly 10 years whilst remaining radiation hard lead to a final design entirely based on silicon detector technology.

Shown in Figure 2.3, the Inner Tracking System is composed of a pixel detector with a radii of between 4.4 cm and 10.2 cm, and a silicon strip tracker which is composed of 10 barrel detection layers reaching a radius of 1.1 m. In order to make tracking system completely hermetic the barrel detectors are surrounded by endcaps composed of 2 disks in the pixel detector and 3 plus 9 disks of silicon strip tracker, thus extending the acceptance, A , of the tracker up to a pseudorapidity of $|\eta| < 2.5$. Each individual pixel station covers a region of $100 \times 150 \mu\text{m}^2$ in the $r - \phi$ and z coordinate system, respectively, and is driven by the desired impact parameter resolution. In total the pixel detector contains 66 million

pixels, corresponding to an active area of 1 m^2 .

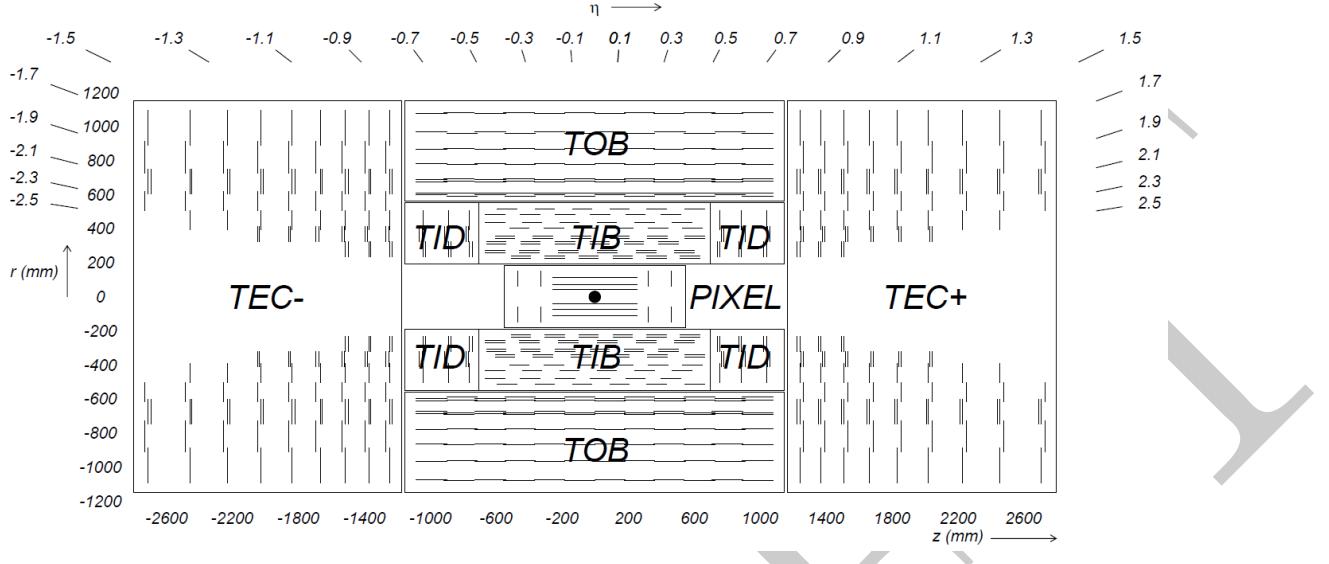


Figure 2.5: The sub-detectors of the CMS silicon tracker system: TOB=outer barrel, TIB=inner barrel, TID=inner disc, TEC=endcaps, PIXEL=pixeldetector. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits. [2].

The sensor elements in the silicon strip tracker system are single sided p-on-n type silicon micro-strip sensors [20, 21]. The Tracker Inner Barrel (TIB) and Disks (TID), where the particle flux is smaller, extends to a radius of between $20\text{ cm} < r < 55\text{ cm}$, and has a typical cell size of $10\text{ cm} \times 80\mu\text{m}^2$, strip thickness of $320\mu\text{m}$, and an occupancy of $\sim 2 - 3\%$ per strip per bunch crossing. The outer layer of the silicon strip tracker ranges from $55\text{ cm} < r < 110\text{ cm}$ and $\sim 500\mu\text{m}$ thick, but with a cell size of $25\text{ cm} \times 180\mu\text{m}$ due to lower levels of radiation in the outer region. The TIB and TID are surrounded by the Tracker Outer Barrel (TOB) which has an outer radius of 116 cm and comprises 6 barrel layers of $500\mu\text{m}$ thickness micro-strip sensors and strip patches of $183\mu\text{m}$ on the first 4 layers and $122\mu\text{m}$ on the 5th and 6th layers. Beyond the range of the TOB lies the Tracker EndCaps (TEC+ and TEC-, where the sign represents the location of the endcap along the z -axis) to provide complete coverage. The TECs cover the region $124\text{ cm} < |z| < 282\text{ cm}$ and $22.5\text{ cm} < |r| < 113.5\text{ cm}$ and is composed of 9 disks each consisting of 7 rings of silicon micro-strip detectors, $320\mu\text{m}$ thick on the inner 4 rings, $500\mu\text{m}$ thick on rings 5-7) with radial strips of $97\mu\text{m}$ to $184\mu\text{m}$ average pitch. Therefore, they provide up to 9 ϕ measurements per trajectory.

2.3.1 Tracker performance in Run I

Over the Run I period, from 2010 to 2013, the LHC delivered around 6 fb^{-1} at 7 TeV and 23.3 fb^{-1} at 8 TeV (Figure 2.1.5), out of this approximately 93% was recorded by CMS. The CMS tracker was responsible for roughly one third of the lost data due to the high voltage only being ramped up once stable beams are reached. By the end of Run I approximately 2.3% of the tracker barrel and 7.2% of the endcap modules were inactive associated with faulty wire-bonds or poor connections. During this period around 2.5% of the strip detector became inactive because of short-circuits in the control rings and HV lines, or due to faulty optical communications. Maintenance and repairs began upon shutdown of the LHC, and CMS was able to salvage up to 1.5% of the pixel barrel, up to 0.5% of the pixel endcap modules, and up to 1% of the strip detectors [22].

In order to process the data prior to track reconstruction the hit efficiency must be measured, the points at which a charged particle traversed each layer of the inner tracker. After track reconstruction the efficiency is calculated as the fraction of particles that are expected to pass through the fiducial regions of the sensors in a layer of the detector in which matching hits are found. For the strip detectors a hit is considered to be a hit if the energy deposit is found in the module in which it was expected to be observed. For efficient reconstruction of tracks knowledge of the position of each module in three-dimensional space is required. Distortions and movements of the inner tracker modules were monitored using cosmic ray data and collision tracks by measuring the distance between expected and observed track trajectories. Distortions in tracking lead to biases in the reconstructed track curvature, and were studied using the reconstructed mass of $Z \rightarrow \mu\mu$ events as a function of the positive muon's azimuthal angle. The muon reconstruction efficiency can be seen in Figure 2.3.1.

The CMS tracking software relies on an iterative procedure to measure hits in a high particle occupancy environment. Earlier steps of the tracking process search for tracks with higher p_T due to the more obvious nature of the tracks, which include a smaller impact parameter, and greater number of measured hits in each layer of the tracker. By selecting more obvious processes first, the reconstruction becomes easier as it has fewer events to deal with. Track reconstruction efficiency is measured by using the tag and probe method in $Z \rightarrow \mu\mu$ events [3]. The tracking efficiency is then defined as the number of probes observed to have matching tracks within the tracker and is a function of the number of primary vertices and the pseudorapidity of the tracks and can be seen in Figure 2.3.1. LHC proton-proton events are reconstructed by firstly identifying the tracks, then grouping in accordance with

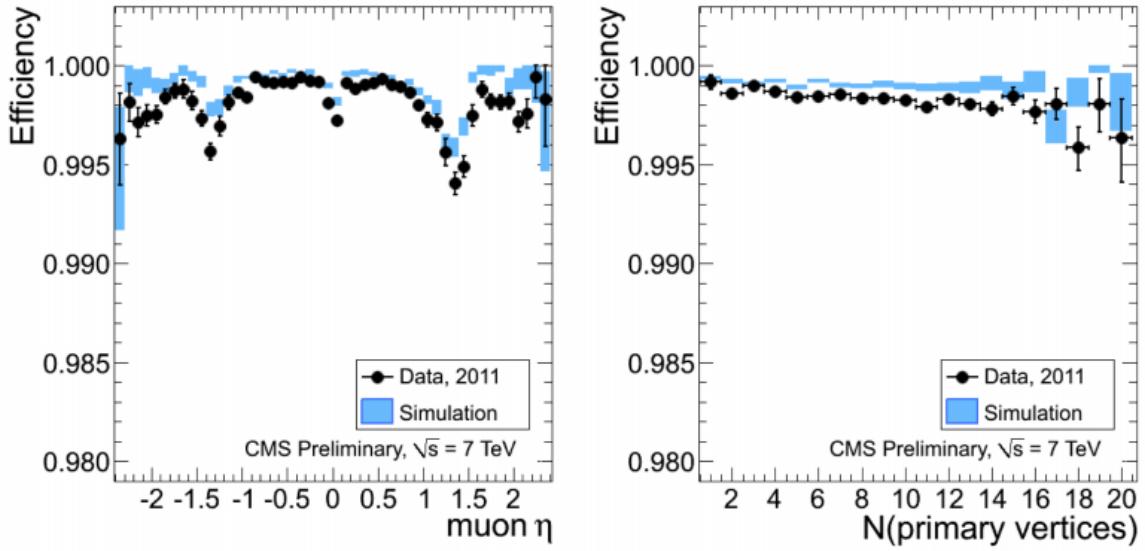


Figure 2.6: Muon reconstruction efficiency in the tracker as functions of pseudorapidity (left) and the number of proton-proton interaction vertices (right) [3].

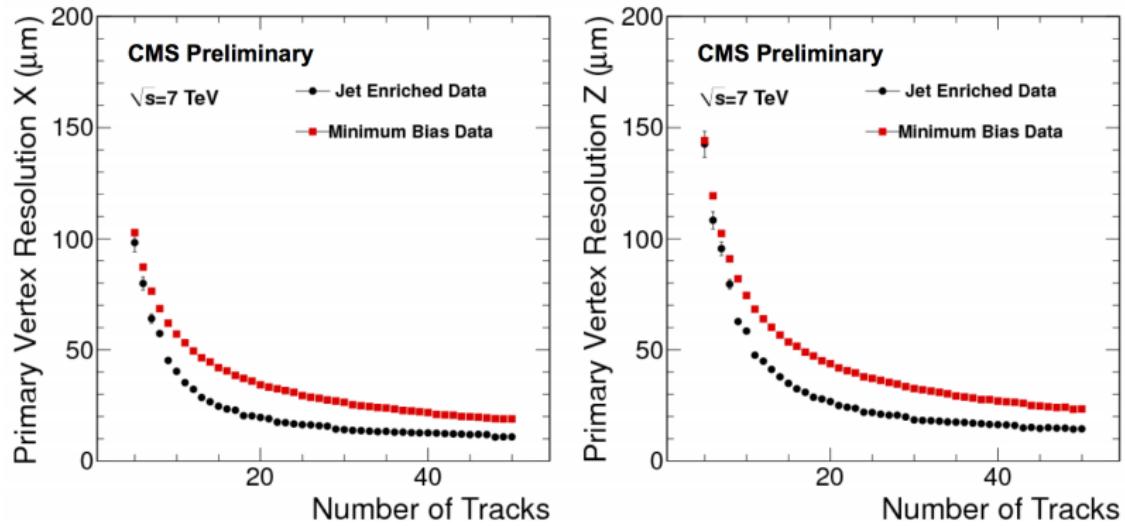


Figure 2.7: Primary vertex resolution in the transverse plane (left) and along the beam-line (right) as functions of the number of tracks attached to the vertex [3].

their primary vertex, and finally fitting to the position of each vertex.

One of the long term damaging effects of high luminosity collisions is radiation damage. Radiation damage in the silicon was monitored throughout Run I and tested by performing special runs where the bias voltage was increased in steps from 0 to the operational voltages. Results showed that the hit efficiency decreased with irradiation at first, then increased with changes in the effective doping [23]. Due to collisions not being completely aligned at the centre of the detector, even irradiation of the modules is seen in the azimuthal direction.

Overall, the CMS tracker has performed exceptionally throughout the Run I three-year period with regards to detector reliability and tracking. The tracker was able to overcome a major problem of high pile-up and reconstruct tracks with excellent efficiency. Less than 3% of the tracker became inactive throughout the entire run, and less than 5% of the delivered luminosity was lost through the tracker.

2.4 Electromagnetic Calorimeter

2.4.1 Overview

Directly after the Inner Tracking System, the second stage of particle identification and reconstruction comes in the form of the Electromagnetic Calorimeter (ECAL). The ECAL serves to stop electromagnetic particles, namely electrons and photons, and measure the energy deposited in the detector. These particles are identified and reconstructed using signatures such as charge, shower shape, and isolation. When an electron passes through the ECAL it showers via bremsstrahlung. Radiation losses due to bremsstrahlung scale with mass as m^{-4} (m^{-6}) when a charged particle travels perpendicular (parallel) to an electric field, and thus heavier electromagnetic particles are less likely to produce a shower. It is possible to differentiate between electrons and positrons by the curvature produced from the Superconducting Solenoid. Photons are neutrally charged and thus do not bend via the magnet, however they produce a shower of electron-positron pairs which can then be measured. The photon shower shape, known as $\sigma_{inj\eta}$, is a prominent variable in this analysis and will be described in detail in Section ??.

A key component that drove the design of the ECAL is the decay channel $H \rightarrow \gamma\gamma$. At the time of design, the Higgs had not been discovered and thus the mass was not known, however it was known that the aforementioned decay mode was sensitive to a low mass Higgs, $m_H < 150$ GeV. Although the branching ratio of the decay is small ($\simeq 0.002$), the signature is clean and is a narrow resonance of two high E_T photons over a non resonant background

[24]. In order to discover the Higgs the detector needed to have a powerful invariant mass resolution and background rejection, translating into a need for extremely efficient photon and electron identification, along with a high position and energy resolution.

2.4.2 Composition of the ECAL

The CMS ECAL is a hermetic, homogeneous fine-grained lead tungstate (PbWO_4) crystal calorimeter [25], shown in Figure 2.4.2. The PbWO_4 crystals are extremely dense ($\delta = 8.28 \text{ g/cm}^3$), thus providing excellent performance and compactness, and thus fit within the Superconducting Solenoid magnet volume. The crystals were designed with an extremely small radiation length, $X_0 = 0.85 \text{ cm}$, and small Molière radius, $R_M = 2.19 \text{ cm}$. The decision to use a homogeneous medium was chosen because of the ability to obtain a greater energy resolution by minimizing sampling fluctuations [25].

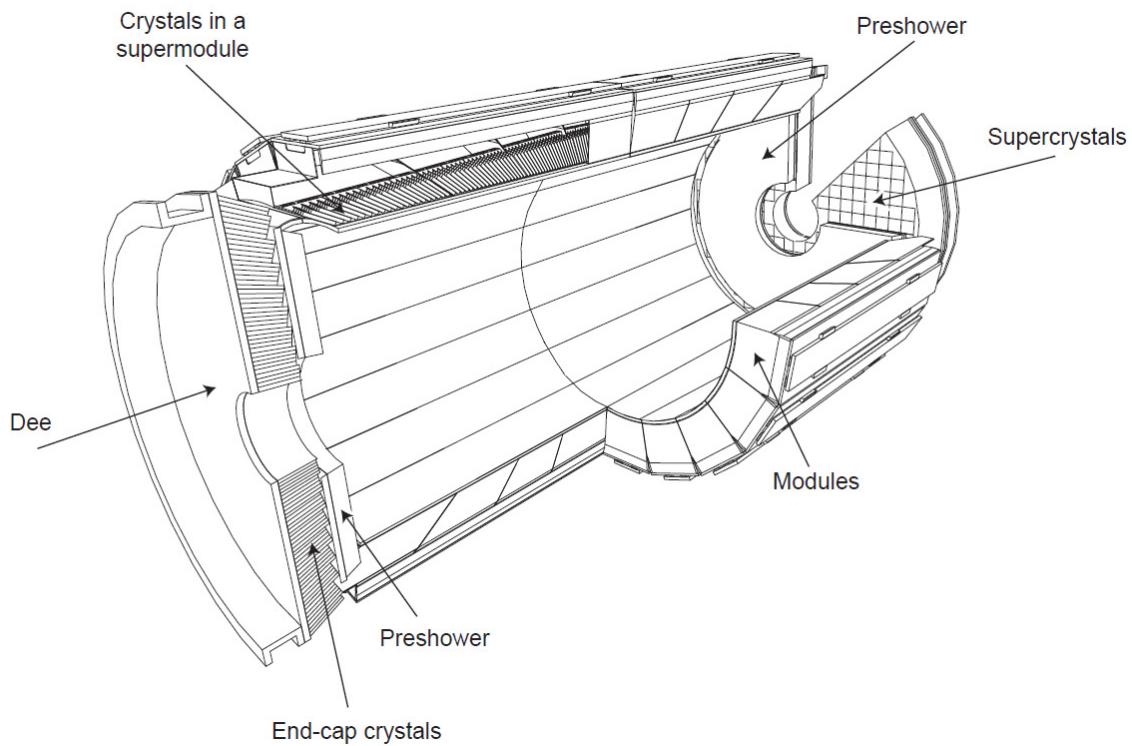


Figure 2.8: Geometric view of one quarter of the ECAL (top). Layout of the CMS electro-magnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (bottom) [2].

There are 75,848 within the ECAL, and are arranged into a barrel section (EB), covering a pseudorapidity rang of $|\eta| < 1.4442$, which is then surrounded by endcaps and thus extending

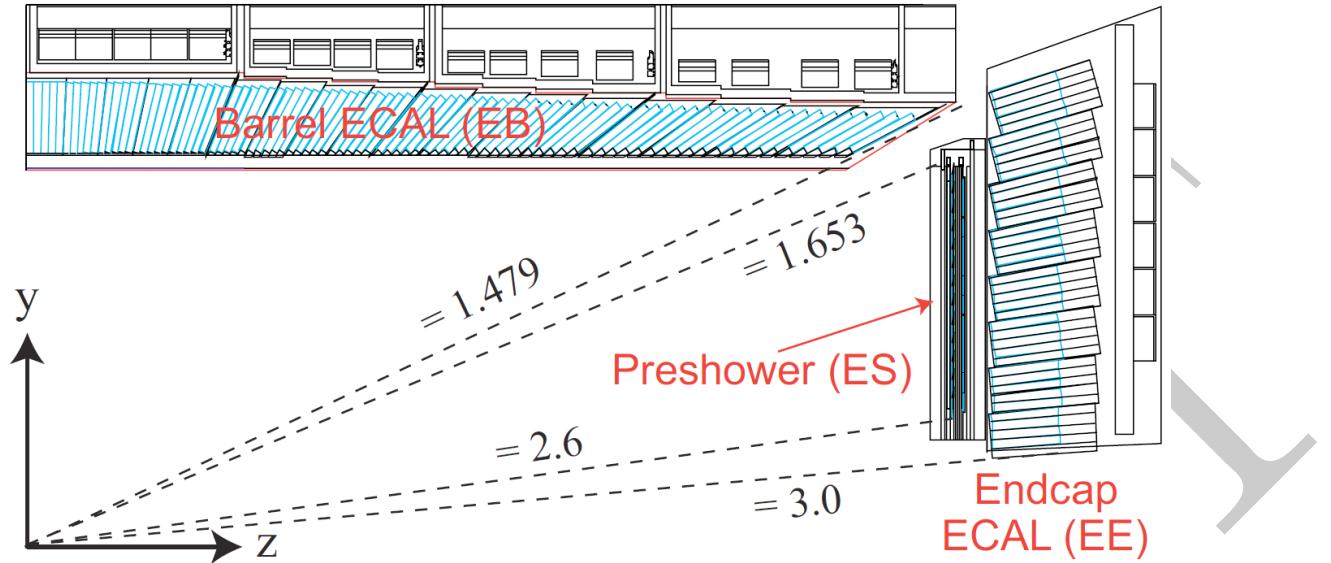


Figure 2.9: Geometric view of one quarter of the ECAL (top). Layout of the CMS electromagnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (bottom) [2].

the pseudorapidity range to $|\eta| < 3.0$. The length of the crystals within the barrel are 230 mm and 220 mm in the endcap regions, which corresponds to ~ 26 (EB) and ~ 25 (EE) radiation lengths. The crystals are projective and also slightly off-pointing in position, $\sim 3^\circ$ with respect to the IP. This configuration provides a full coverage and ensures that there are no cracks in the calorimetry that are aligned with particle trajectories. Within the barrel there is no longitudinal segmentation, and therefore the angle at which a photon is measured relies on the reconstructed PV from the silicon tracker. EB crystals are $2.2 \times 2.2 \text{ cm}^2$ on the front face, and $2.86 \times 2.86 \text{ cm}^2$ in the endcaps, giving rise to a total crystal volume of 11 m^3 and a weight of 92 t.

The barrel crystals are arranged into 36 supermodules (or superclusters), each containing 1,700 crystals, whereas the endcaps are arranged into two D-shaped segments comprising 3,662 crystals each. The final section of the ECAL is the pre-shower detector system (ES) placed directly in front of the endcaps at $1.65 < |\eta| < 2.6$ and can be visualised in Figure 2.4.2. The ES is composed of 4,288 sensors, 137,216 silicon strip sensors, each $1.90 \times 61 \text{ mm}^2$ with x-y view, and has a total of ~ 3 radiation lengths. The purpose of the ES is provide improved separation of photons to π^0 s.

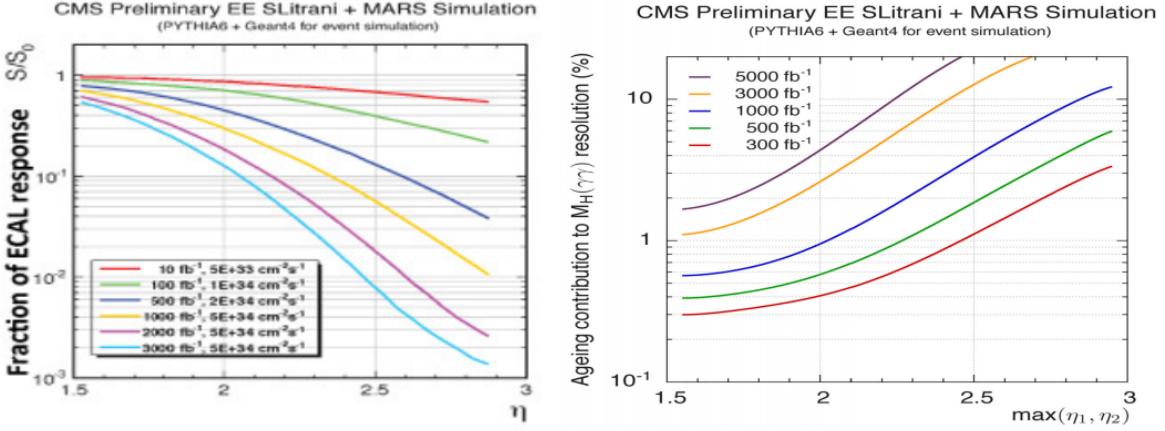


Figure 2.10: Simulation of fractional response from EE as a function of η , for different integrated luminosities. Right: Deterioration of the energy resolution in EE as a function of η , for different integrated luminosities [4].

2.4.3 Photodetectors

The light read-out system for the barrel crystals comes in the form of Hamamatsu avalanche photodiodes (APD). There are two APDs for each crystal which are read in parallel, each measuring $5 \times 5 \text{ mm}^2$ with a quantum efficiency (QE) of 75%. The gain is set at ~ 50 and they are insensitive to the 4 T magnetic field from the Superconducting Solenoid. The endcap crystals scintillation light is read out by vacuum photo-triodes (VPT), each with an area of 280 mm^2 with a 20% QE and gain of ~ 10 . The barrel APDs are temperature sensitive ($\frac{1}{E} \frac{dE}{dT} \sim -2.3\% C^{-1}$) whereas the VPT sensitivity to temperature is assumed to be negligible relative to that of the crystals.

2.4.4 Performance of the ECAL throughout Run I

The energy resolution, $\frac{\sigma_E}{E}$ of the ECAL crystals can be parameterised by

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{A}{\sqrt{E}}\right)^2 + \left(\frac{B}{E}\right)^2 + C^2 \quad (2.4)$$

where A and B are the stochastic term for scintillation showers and noise term due to read-out electronics and PMTs, respectively. C is a constant term which is a direct measure of the performance of the PbWO₄ crystals.

	HB/HO	HE $ \eta \leq 2.5$	HE $ \eta > 2.5$	HF $ \eta \leq 2.5$	HF $ \eta > 2.5$
$\Delta\eta \times \Delta\phi$	0.087×0.087	0.087×0.087	0.175×0.175	$0.175 \times 0.175s$	0.175×0.35

Table 2.2: Tower segmentation in azimuthal and polar angle for the hadronic barrel, endcap and forward calorimeter [12].

2.5 Hadron Calorimeter

2.5.1 Overview

The Hadron Calorimeter (HCAL), shown in Figure 2.5.1, lies directly after the ECAL within the volume of the superconducting solenoid, and is designed to measure energy deposits from hadron showers and clusters of collinear high-energy hadronic particles known as jets. A key role of the HCAL is to measure the missing transverse energy (MET) produced by events containing neutrinos, and possible events that could be associated with new physics that may be seen at higher energies. In order to produce such a precision measurement, the HCAL must have an extremely high jet energy resolution and be completely hermetic in ϕ .

The HCAL is a sampling calorimeter, composed of alternating layers of absorbing large brass plates and plastic scintillation tiles. Because the precision of the energy measurement depends on the total amount of the hadronic shower detected, the material must be thick enough to absorb the majority of the event. The size restriction for the HCAL is limited to the distance between the end of the ECAL ($r = 1.77$ m) to the inner superconducting solenoid ($r = 2.95$ m). In order to efficiently measure the full shower, and therefore energy deposit, an outer section of the HCAL (HO) has been placed just after the solenoid and between the muon system which can then also be used as an extra layer of absorbing material with a thickness of 11.4 interaction lengths (λ_l). There are four segments of the HCAL in total: the HCAL barrel (HB), endcaps (HE), outer HCAL (HO), and forward calorimeter (HF). The barrel and endcaps combined covers an area with pseudorapidity up to $|\eta| < 3.0$, and the forward segment covers the region up to $|\eta| < 5.0$.

There are approximately 90,000 scintillators installed within the HB and HE combined. Light that is collected by the plastic scintillating tiles is read about by wavelength-shifting fibres (WSF) that are embedded within the units, which are then transported through transparent fibres to hybrid photo-detectors (HPD) encompassing 19 independent pixels. Each scintillator has a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta| \geq 1.6$, whereas the forward segment changes with respect to pseudorapid-

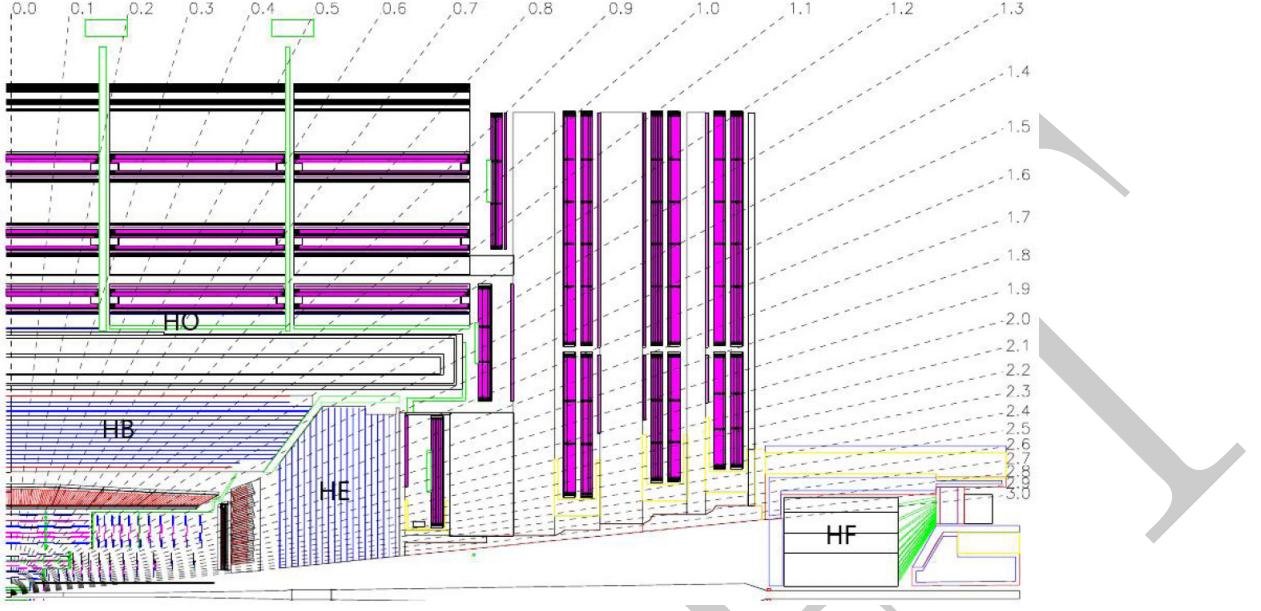


Figure 2.11: Longitudinal view of one quarter of the detector in the $r - \eta$ - plane, showing the positions of the HCAL parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF) [2].

ity as $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ at $|\eta| = 3.0$ to $\Delta\eta \times \Delta\phi = 0.175 \times 0.35$ at $|\eta| = 5.0$ [12], as seen in Table 2.5.1. The forward section of the HCAL is placed 11.2 m from the IP in order to reconstruct particles boosted in the forward direction, and is expected to experience a much higher flux of particles at higher energy (760 GeV) than the rest of the HCAL (100 GeV) at $\sqrt{s} = 14$ TeV [26]. It is composed of 5 mm thick steel absorber plates, each with quartz fibres implemented as an active medium. The quartz fibres detect Čerenkov light produced from the electromagnetic component of particle showers.

2.5.2 Performance of the HCAL in Run I

2.6 Superconducting Solenoid

The CMS Superconducting Solenoid, shown in Figure 2.6, is the most powerful magnet in the world, 100,000 times stronger than the Earth’s magnetic field and stores enough energy to melt 16 tonnes of gold, and the most essential feature of the detector. In order to achieve a good momentum resolution in such a detector, without making tight cuts on muon chamber resolution and alignment, a powerful magnetic field was chosen. A large bending power can be achieved by a modestly sized solenoid, as long as it is a high-field superconducting one,

Superconducting Solenoid Parameters	
Parameter	Value
Field (T)	4
Length (m)	12.9
Weight (t)	250
Inner bore (m)	5.9
Current (kA)	19.5
Number of turns	2168
Stored energy (GJ)	2.7
Hoop stress (atm)	64

Table 2.3: Parameters of the LHC superconducting solenoid [13].

due to the bending beginning at the primary vertex. The requirement for the bending power of the solenoid is dictated by the narrow states decaying into muons, and by the unambiguous determination of the sign for muons with a momentum of around 1 TeV/c. In order to obtain a precision measurement, a momentum resolution of $\Delta p/p \approx 10\%$ at $p = 1$ TeV/c. A suitable length to radius ratio is required to obtain a good momentum resolution in the forward region.

Approaching 13 m in length, and 6 m in diameter, the solid mass weights approximately 250 t at an operating temperature of -268.5°C – a degree warmer than outer space. Originally designed to run with a uniform magnetic field of 4 T within the 5.9 m bore, the eventual operating level was set to 3.8 T in order to increase the lifetime. Such a magnetic field requires a return yoke, which can be viewed in the CMS schematic in Figure 2.2, of which the return field is large enough to saturate 1.5 m of iron and weighs 12,500 t. This allows four muon stations to be integrated within the return yoke, ensuring robustness and full geometric coverage. The magnet and return yoke use almost twice as much iron as the Eiffel Tower. The solenoid is composed of four layers of high-purity niobium-titanium cable coextruded with a aluminium-stabilised conductor as used in previous experiments, such as ALEPH and DELPHI at LEP, and H1 at HERA. However, the huge increase in certain parameters, such as magnetic field strength, ampere turns, and stored energy, which can be seen in the parameters listed in Table 2.6 [27].

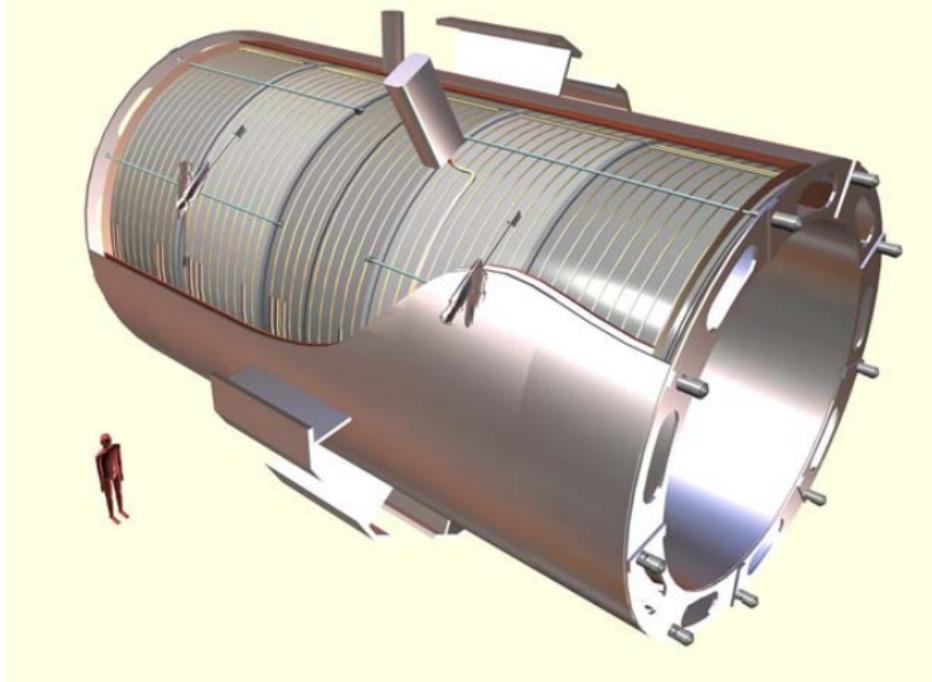


Figure 2.12: General artistic view of the 5 modules composing the cold mass inside the cryostat, with details of the supporting system (vertical, radial and longitudinal tie rods) [2].

2.7 Muon System

The final sub-detector, or rather set of sub-detectors, lying between each wheel of the iron return yoke, is the muon system (Figure ??). The muon system plays a huge role in the detecting signatures of interest of the extremely high background events seen in Run I, and which are expected to increase in Run II. Signatures such as the so called “gold-plated” decay of a standard model Higgs decaying to two Z bosons, both of which decay into two muons ($H \rightarrow ZZ \rightarrow \mu^+\mu^-\mu^+\mu^-$), are ideal candidates because the best mass resolution can be achieved as muons are much less effected than electrons by radiative losses within the tracker material. Due to the relative ease of detecting muons, this decay channel (discovered in Run I, [1]) highlights the discovery potential for muon final state decay modes and the demand for such a wide angular coverage within the muon detection system.

Precise and efficient muon measurements were a central theme in the design of the CMS experiment, as can be seen from the name. There are three functions which the muon system serves: identification of muons, momentum measurements, and triggering. All of these functions are made possible when used in conjunction with the superconducting solenoid and its flux return yoke. The system was designed to measure the charge and momentum of muons

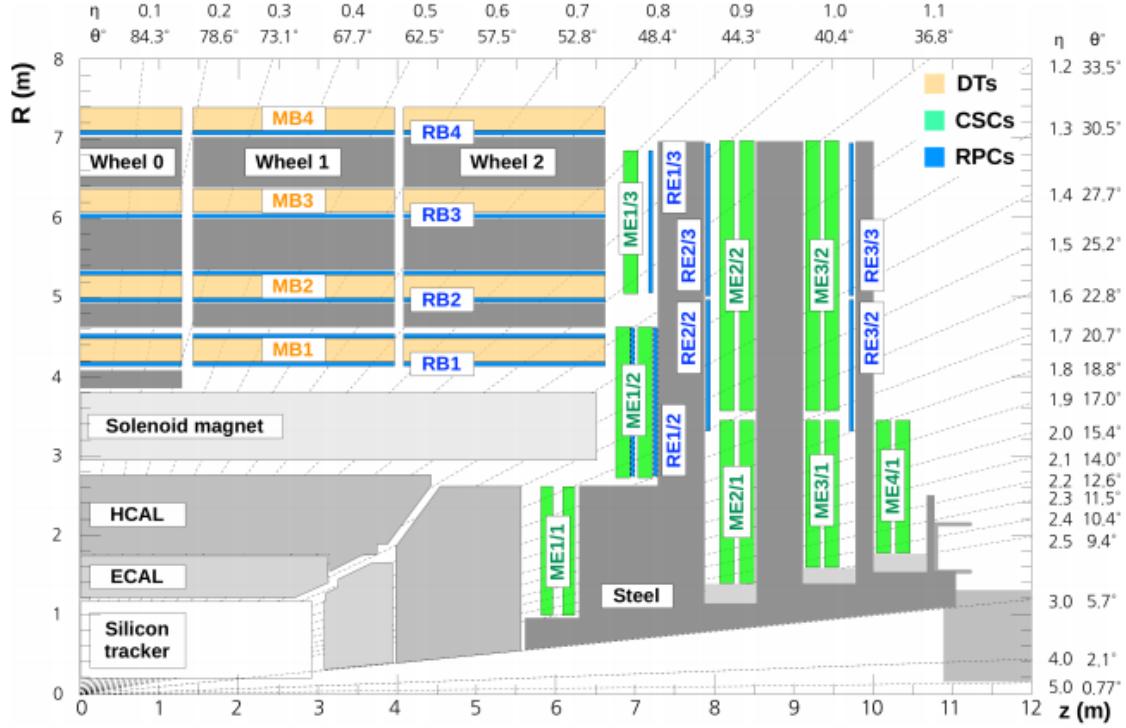


Figure 2.13: Layout of one quadrant of CMS. The figure shows the four DT stations in the barrel (MB1-MB4, yellow), the four CSC stations in the endcap (ME1-ME4, green), and the RPC stations (RB1-RB4 and RE1-RE3) [2].

over the full kinetic range of the LHC. The design of the muon system was based around the nature of the solenoid magnet, thus it is composed of a cylindrical barrel section surrounded by two planar endcaps to provide fully hermetic coverage. The design of the system had corresponds to 25,000 m² of detection planes and thus the muon chambers must be robust, reliable, and also inexpensive. Therefore, three types of gaseous particle detectors were implemented: drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC).

Drift tubes with standard rectangular drift cells are implemented in the muon barrel section where the neutron-induced background is small, the muon rate is low, and the 4 T magnetic field is uniform and mostly contained within the steel return yoke. Covering a pseudorapidity region of $|\eta| < 1.2$, the barrel drift tube chambers are arranged into four layers integrated within the return yoke. The first three layers contain 8 chambers which measure the muon coordinates in the $r - \phi$ bending plane, and 4 that measure z along the beam-line. The fourth station is the same except that it does not measure z . The arrangement and number of chambers in each station were chosen to provide efficient rejection of background

hits and linking muon hits from different stations into one single track.

In contrast to the barrel region, the background and muon rates are high and the magnetic field is non-uniform. For this region cathode strip chambers are used. The CSCs have an extremely fast response time, resistance to radiation, fine segmentation, and can identify muons within a pseudorapidity range of $0.9 < |\eta| < 2.4$. Again there are four layers of chambers interspersed between the return yoke plates, however they are perpendicular to the beam. The strips are positioned radially outward and provide precise measurements in the $r - \phi$ bending plane. Each CSC comprises 6 layers, providing robust pattern recognition and rejection of backgrounds, efficiently matching hits in the CSCs to those in the tracker.

The third sub-detector in the muon system is implemented to serve as a complimentary, dedicated trigger system composed of resistive plate chambers (RPC) in both the barrel and endcaps. The decision to include the RPCs was based on the uncertainty in the eventual background rates and ability to measure the correct beam-crossing time when the LHC reaches design luminosity. The RPCs are able to operate in high rate environments by using fast, highly-segmented, and independent trigger, but provide a more accurate position resolution than the DTs or CSCs. The p_T threshold is sharp over a large segment of the pseudorapidity range, $|\eta| < 1.6$.

2.7.1 Performance of the muon system in Run I

2.8 Trigger and Data Acquisition

At the LHC design energy of $\sqrt{s} = 14$ TeV, the proton-proton collision frequency reaches up to 40 MHz when operating with 25 ns bunch spacing. Depending on the luminosity, a number of collisions will occur at each crossing of the protons, but since every event produces ~ 1 MB of raw data, this corresponds to a total of $\sim 40TB s^{-1}$. Also, at the design luminosity of $\mathcal{L} = 10^{34} \text{ cm}^{-2} s^{-1}$, around 20 inelastic events will be superimposed onto events of interest, known as pile-up (PU). The amount of resulting data is much too large to store and process, and thus a filtering system for interesting events is implemented to reduce the total number of events recorded. This is the trigger system, and begins the process of event selection in the CMS experiment. The rate reduction capability is designed to be at least a factor of 10^6 for the combination of L1 Trigger and HLT.

The trigger is designed as a two-stage system: Level 1 (L1) Trigger, and High-Level Trigger (HLT), respectively. The L1 trigger is a hardware system consisting mostly of custom-designed, highly programmable electronics, located partly on the detector and partly in the

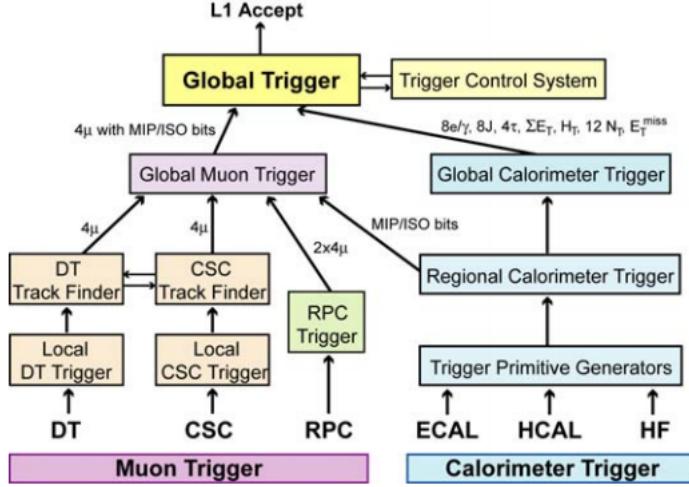


Figure 2.14: Architecture of the Level 1 trigger [2].

underground control room approximately 90 m from the detector itself. The L1 trigger makes a decision based on information from only the muon system and calorimetry, and is shown in Figure 2.8. The tracker information is not used in making trigger decisions due to the length of time needed to reconstruct tracks exceeding that of the L1 trigger. The data used is coarsely segmented, with the high resolution data being held in pipeline memories in the front-end electronics. Constructed with a design output rate limit of 100 kHz, this in actuality translates in practice to a maximal output rate of ~ 30 kHz when taking into account a safety factor of 3. We can divide the L1 trigger into different components: local, regional, and global components. The first components are the Local Triggers, also named Trigger Primitive Generators (TPG), and are collections of information from deposits in calorimeter trigger towers and track segments, or hit patterns in the muon chambers, respectively. Next are the Regional Triggers that combine the information obtained and use pattern logic in order to determine the rank and sort of trigger object, such as an electron or muon candidates. The rank can then be defined as a function of energy or momentum and quality of the deposit, reflecting the confidence level assigned to the L1 parameter measurements. The final components are the Global Calorimeter and Global Muon Triggers, assigning the highest rank calorimeter and muon objects throughout the detector. The candidates are then transferred to the Global Trigger, the highest component of the Level 1 hierarchy, which then makes a decision of whether to accept or reject the event for evaluation by the HLT.

The HLT is a software system residing in a CPU filter farm containing around 1000 commercial processors. The advantage of the HLT is that it has access to the full event infor-

mation from the entire detector simultaneously, including event information unavailable on the timescale of the L1. It can then perform complex calculations, similar to that performed in the analysis off-line software, by constantly evolving complex algorithms, resulting in a highly flexible trigger system. The event rate at the HLT level is reduced to ~ 300 Hz and a final data rate of 300 MBs^{-1} recorded on a large storage disk (the Storage Manager) at the experimental site. The data is later transferred to the base tier (Tier 0) of The Grid computing network for further processing and then physics analysis.

2.8.1 Performance of the Trigger in Run 1

Chapter 3

Event Reconstruction & Simulation

Monte Carlo (MC) simulations are an essential part of current particle physics analyses and are used to mimic physical processes that correspond to those which are observed within the LHC, and other such experiments. Analysts compare findings in data to simulation in order to extract signal processes, and also to perform statistical analysis on results obtained. It is of the utmost importance that the simulated events must be as accurate as physically possible in order to mimic real life processes and perform a scientifically accurate analysis. Will talk about methods for generating events, including the different MC generators and tunes used in the evaluation of theoretical uncertainties, and interpretation in terms of the CMS detector in the first section of this chapter.

Roughly speaking, we can divide the different steps of event reconstruction into three separate processes. The first of which records basic information, such as hits within the pixel detectors of the inner tracking system, and calorimeter energy clusters, for ‘low level’ objects in each sub-detector. The information is then passed to the PF algorithm which uses information from all the sub-detectors in order to reconstruct events much more accurately. Finally, the events are refined by other complex statistic and mathematical techniques and used to reconstruct higher level objects, such as jets and MET. The second part of chapter will focus on the PF process [28, 29] as mentioned above.

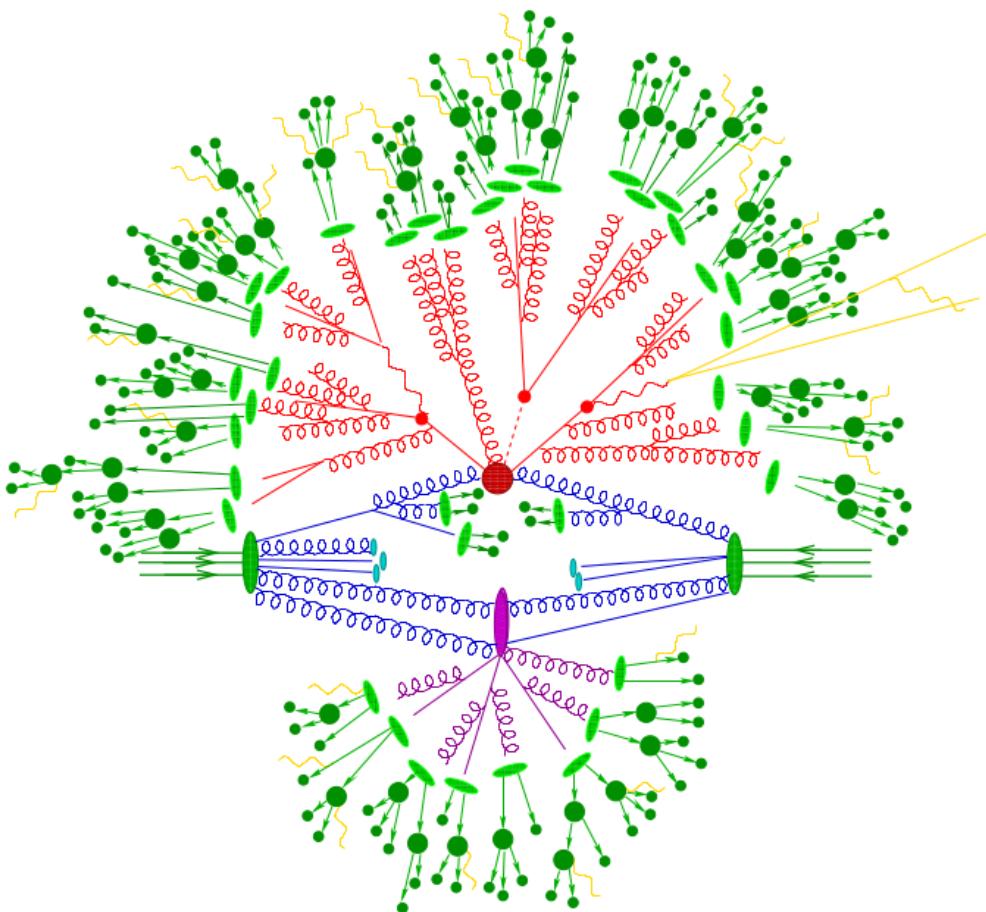


Figure 3.1: Schematic overview of a hadron collision process [5].

3.1 Event Reconstruction

3.2 Computing

3.2.1 Event Data Model

Physically, an event is the result of the hard scattering process created when the LHC collides two bunches of protons together. We cannot know the full properties of an interaction without processing the data, and this is where the Event Data Model (EDM) is derived. The information about the event is measured as energy deposits, clusters, and tracks in each sub-detector of the CMS experiment that is read out by a series of processing electronics. When it comes to reconstructing the data and performing statistical analysis, we read in an event as a C++ object storing the raw data. These objects are presented to a user stored as ROOT files [30]. There are three different data formats that data is stored as, each of which contains a different level of precision for describing events. The three levels are as described below.

RAW file formats contain very primal information about the data, including the L1 and HLT decisions. At this stage events have a size of roughly 1.5 MB.

RECO is the next step in the formatting of data by reconstructing events obtained from the RAW format with pattern recognition and compression algorithms. This includes reconstructed detector hits, calorimeter clusters, and reconstructed physics objects such as electrons, jets etc. The typical event size at this level is around 250 kB.

AOD (Analysis Oriented Data) is created by filtering the RECO data from the reconstructed detector-level objects, where the higher level physics objects are calculated. The size of the events is reduced to ~ 50 kB.

Almost all physics analysis groups use the RECO and AOD data formats. All data used in this analysis uses the AOD data format, where event sizes are reduced by filtering the RECO data, as can be seen in Table 3.5. The objects are delivered in the data files as C++ objects, which are then transformed into vectors or plain basic types. By selecting on the physics objects that are central to our analysis, we can further reduce the event size to ~ 3 kB, thus reducing the run time of an analysis. The processing of RECO and AOD formats for analysis begins with taking the data and using a specifically designed framework to process a set of “nTuples” for each physics process, and categorising them into classes, such as objects.

The benefit of constructing such “nTuples” lies in the reduction of processing time by allowing the analysis to be run locally, rather than re-processing the full dataset in AOD, RECO, or RAW from each time.

3.2.2 Analysis Software

Analysis is usually performed within a computing environment specific to an experiment. CMS provides an extensive software framework, CMSSW [31], that provides users with a large range of algorithms to create, handle, and analyse data. CMSSW is fundamental in regards to MC simulation, detector calibration and alignment, and the for the reconstruction of data and analysis thereof. The framework is a modular structure that combines a single configurable executable (`cmsRun`) with a number of plugin modules containing event processing algorithms etc. CMSSW is continuously being updated to keep up-to-date with new analysis techniques and processes. The versions of CMSSW used in this analysis are listed below:

- CMSSW_5_3_9 for the analysis of $t\bar{t} + \gamma$.
- CMSSW_5_2_10_nTuple for the ntuple processing for the $t\bar{t} + \gamma$ analysis.

The framework used is modular similar to CMSSW. The code is split into different modules designed to model the data at various stages of processing. We essentially design four modules to carry out the reading in of the data, transferring it into a readable C++ format, selecting and implementing cuts on our events, and outputting the information in the form of a histogram in order to statistically analyse the data, as described below.

- Reader files that translate plain data types stored in ROOT files into C++ objects.
- Reconstruction objects process the output of the readers in the format of real objects, such as leptons and quarks.
- Selections are written for each decay channel in an analysis to select on objects that exist in the final state of an event.
- Analysers are used to create histograms of different variables at various stages of selection, implement scale factors, and add weights to samples.

3.3 Simulation of the CMS detector

The simulation of the CMS detector is an incredibly complex task and very time consuming to run such a simulation. In order to perform such a procedure GEANT4 [32] is used to simulate

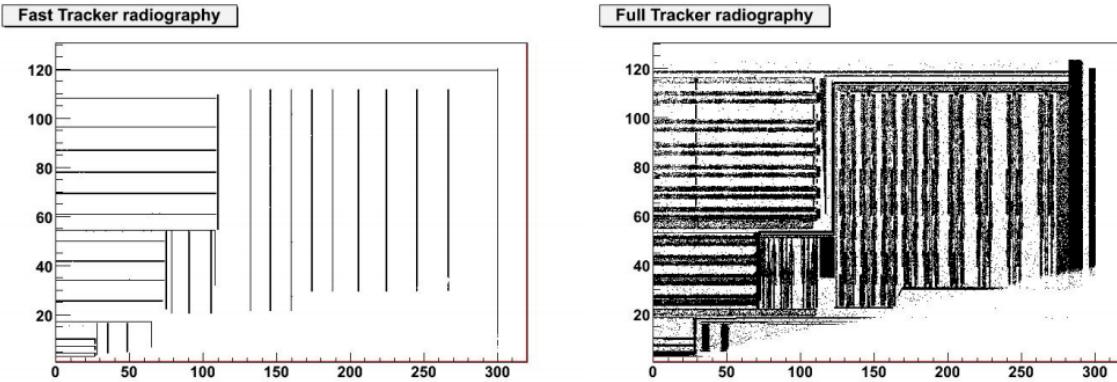


Figure 3.2: A radiography of a quarter of the simulated tracker geometry in the (a) fast and (b) full simulation [6].

the geometry of the whole detector, divided into sub-detectors, and track the particles as they pass through the different materials. Implemented within the CMSSW framework are two packages that perform detector simulation, Full Simulation (FullSim) [33] and Fast Simulation (FastSim, previously named FAMOS) [6]. In the FastSim package physical processes are described in detail, such as the electromagnetic and hadronic interactions, energy deposits, and electronic detector responses. The FastSim package is designed with a much lower level of detail incorporated and reduces the computational time by 3 orders of magnitude. This allows analysts to carry out custom MC productions within a reasonable time. A comparison between the two packages is shown in Figure 3.3.

3.4 Monte Carlo Simulation

3.4.1 Monte Carlo event generators

MC generators are programs that simulate the properties of particles such that results can be compared to those obtained from data in order to perform a statistical analysis and verify findings. CMS uses many event generators in order to simulate specific physics processes that are of interest. Hard scattering processes are simulated via matrix element calculation and then parton showering is subsequently added. In order to convolute the two procedures a matching is implemented. Finally, hadronisation is then modelled as the showered partons form colourless bound states with other partons. MC event generators produce a list of all particles/partons that are created in an event along with their kinematic properties, such as p_T , and includes production of underlying events (UE) and additional primary vertex

interactions from pile-up (PU). Underlying events are classified as any source of interaction produced that does not originate from the initial hard scattering process, such as initial and final state radiation (ISR and FSR), and remnants from the beam. We define PU as any other interaction produced from the same bunch crossing as our hard scattering. Bunch crossings can contain up to 20 different interactions, that is to say 20 different primary vertices are observed. Figure 3.1 shows the hard scattering collision process as produced by MC event generators.

In the analysis presented in this thesis each physical process was simulated by different MC event generators, such as WHIZARD MADGRAPH PYTHIA MC@NLO POWHEG. The MC event generators which are used in this analysis, as previously mentioned, are described in more detail below. Generators are usually combined by interfacing with another generator in order to optimise for each simulation step described in Section 3.1. An example of this can be seen within the main background sample for this analysis, the $t\bar{t}$ sample, which is generated using both MADGRAPH and PYTHIA.

WHIZARD [34] is a LO event generator designed to calculate multi-particle scattering cross-sections efficiently and simulated event samples. Tree-level matrix elements are automatically generated for arbitrary partonic processes, in particular the MSSM is supported including an interface to the SUSY Les Houches Accord input format. It is also possible to interface matrix elements from alternative processes, such as loop corrections. WHIZARD uses a multi-channel method for phase space integration and is able to calculate numerically stable signal and background cross-sections and also generate unweighted event samples with a reasonable efficiency for final state events containing up to 6 or more particles. The polarisation of initial and final states is treated in the same manner, and quark and lepton flavours are automatically summed over where needed. For hadron collider physics studies, the standard LHAPDF library is incorporated. For fragmentation and hadronisation of the final states, PYTHIA and HERWIG interfaces are provided, both of which follow the Les Houches Accord.

MADGRAPH [35] is a leading order multi-purpose matrix-element MC generator. Similar to the WHIZARD generator, MADGRAPH automatically generates matrix elements for scattering processes with up to 6 and above final state particles. The matching of matrix elements to parton showers is performed following the MLM prescription [36] only if a parton-jet pair satisfies a predefined ΔR requirement, and if none, or more than one jets, are found then the event is rejected. There is also a certain p_T threshold

for partons that they must pass in order to be considered for matching.

PYTHIA [37] is a widely used tool in the particle physics community. It is used for the generation of events within high-energy collisions, and it does so by utilising a complex set of physics models to process the evolution of a 2-body (or more) scattering to a complex multi-particle final state. The generator contains a library of hard processes, models for parton showers for initial and final states, matching among hard processes and parton showers, multi-parton interactions, beam remnants, string fragmentation and particle decays.

MC@NLO [38] provides a method for matching next-to-leading order (NLO) calculations of QCD processes with a particle shower from simulation in MC. MC@NLO improves on many aspects with respect to a LO generator, such as PYTHIA. Aspects such as the total exclusive rates are accurate to NLO, hard emissions are treated as in NLO computations with soft/collinear emissions handled by MC, and matching between hard and soft emission regions is smooth. This provides an advantage for heavy flavour physics, such as top quark production. A small amount of events with negative weights are generated, however the process of unweighting is possible with a reasonable efficiency.

POWHEG [39] (Positive Weight Hardest Emission Generator) is a NLO event generator similar to MC@NLO described above. The difference between the two arises in the basic idea behind POWHEG, whereby it generates the hardest radiation first before passing the event to any shower generator to perform subsequent, softer radiation. Thus, as it does not depend on any parton shower program in particular, the output of POWHEG can be easily interfaced with any shower generator capable of handling the given user process. Another feature of POWHEG is that events can be created with positive (constant) weight.

3.5 Simulated samples for the $t\bar{t} + \gamma$ analysis

As previously mentioned, different MC samples are simulated using different event generators for various physical processes. Table 3.5 provides all the different MC sample datasets used in the $t\bar{t} + \gamma$ analysis along with their respective cross-sections and number of processed events in each sample. This section will focus on background samples where generation of the signal $t\bar{t} + \gamma$ sample can be seen in Section 3.6. It can be seen that the main background to this analysis, TTJets ($t\bar{t}$), is was generated using the MADGRAPH5 event generator and

interfaced with the TAUOLA generator. TAUOLA is a MC event generator designed specifically for the modelling of tau lepton decays. The samples are then passed to PYTHIA6 for parton showering and hadronisation as described in Section 3.4.1. Each $t\bar{t}$ decay process, fully-leptonic, semi-leptonic, and fully hadronic is treated individually and generated as three independent samples. The advantages of treating the decay modes separately are that no scale factor has to be implemented in order to account for the different branching ratios of each decay channel and it is also more convenient to observe the $t\bar{t} + \gamma$ content in each channel separately.

Similarly, both Drell-Yan samples, W+Jets, $W + \gamma$, $Z + \gamma$, and diboson background samples are also simulated in the same fashion as the TTJets samples. Single top events are simulated in a slightly different fashion, whereby they are generated by the POWHEG generator, also described in Section 3.4.1. Again, the events are then passed to PYTHIA6 to model parton showering and hadronisation. Single top samples are also split into different decay modes: tW, s-channel, t-channel. Top quarks and anti-top quark processes are treated separately and simulated in different samples.

3.6 Simulation of the $t\bar{t} + \gamma$ Signal Sample

Three different techniques were used to define the $t\bar{t} + \gamma$ signal process. The concepts are illustrated in Figure 3.6 and shows the final state of the process using each technique [7]. The parton distribution function CTEQ6L1 [40] is interfaced to WHIZARD via LHAPDF [41]. The process utilises variable renormalisation and factorisation scales. This is such that, event by event, the two are set to 172.5 GeV (m_t) plus the E_T of the generated photon. Upon varying the scale of each, we arrive at a systematic uncertainty of $^{+7.0}_{-8.3}\%$, as shown in Chapter 6. Initial and final state radiation is taken into account, as well as hadronisation, and is simulated using PYTHIA6 [42], TAUOLA and PHOTOS [43] as preconfigured in CMSSW. We use the same configuration as for the top-pair sample.

Restrictions on the final state particles have been set, named generation cuts, such that a proper integral is retained when calculating matrix elements. As a method to cope with infrared divergences, a minimum energy or momentum is required. We treat collinear divergences by introducing a minimum distance in the $\eta - \phi$ plane. These cuts likely will not affect the measurement due to the cuts within selection being tighter than generator level cuts. The different generation cuts are described in brief below:

2 → 3 At this level only quantum mechanical interferences from initial state radiation are

$t\bar{t}(Hadronic)$	/TTJets_HadronicMGDecays_8TeV-madgraph/Summer12_DR53X-	245.8	31223821
PU_S10_START53_V7A_ext-v1/AODSIM			
PU_S10_SemiLeptMGDecays_8TeV-madgraph/Summer12_DR53X-	245.8	25424818	
PU_S10_START53_V7A_ext-v1/AODSIM			
/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-	245.8	6923652	
tauola/Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM			
Drell-Yann, $10 < m_{ll} < 50$	/DYJetsToLL_M-10To50_TuneZ2Star_8TeV-	11050.0	37835275
Drell-Yann, $m_{ll} > 50$	madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM		
Z + γ	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-	3350.0	30459503
	tarball/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM		
	/ZGToLLG_8TeV-madgraph/Summer12_DR53X-	159.12	6588161
	PU_RD1_START53_V7N-v1/AODSIM		
Single Top tW	/T_tW-channel-LDR_TuneZ2star_8TeV-powheg-	11.1	497658
Single TopBar tW \bar{t}	tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM		
Single Top t	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-	11.1	493460
Single TopBar t	tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM		
Single Top s	/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-	56.4	99876
Single TopBar s	PU_S10_START53_V7A-v3/AODSIM		
W+Jets	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-	30.7	1935072
	/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-		
	PU_S10_START53_V7A-v1/AODSIM	3.79	259961
	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-		
	PU_S10_START53_V7A-v1/AODSIM	1.76	139974
	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-		
	tarball/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM		
	/WGToLNuG_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-		
		36257.2	57709905
		553.92	4802358

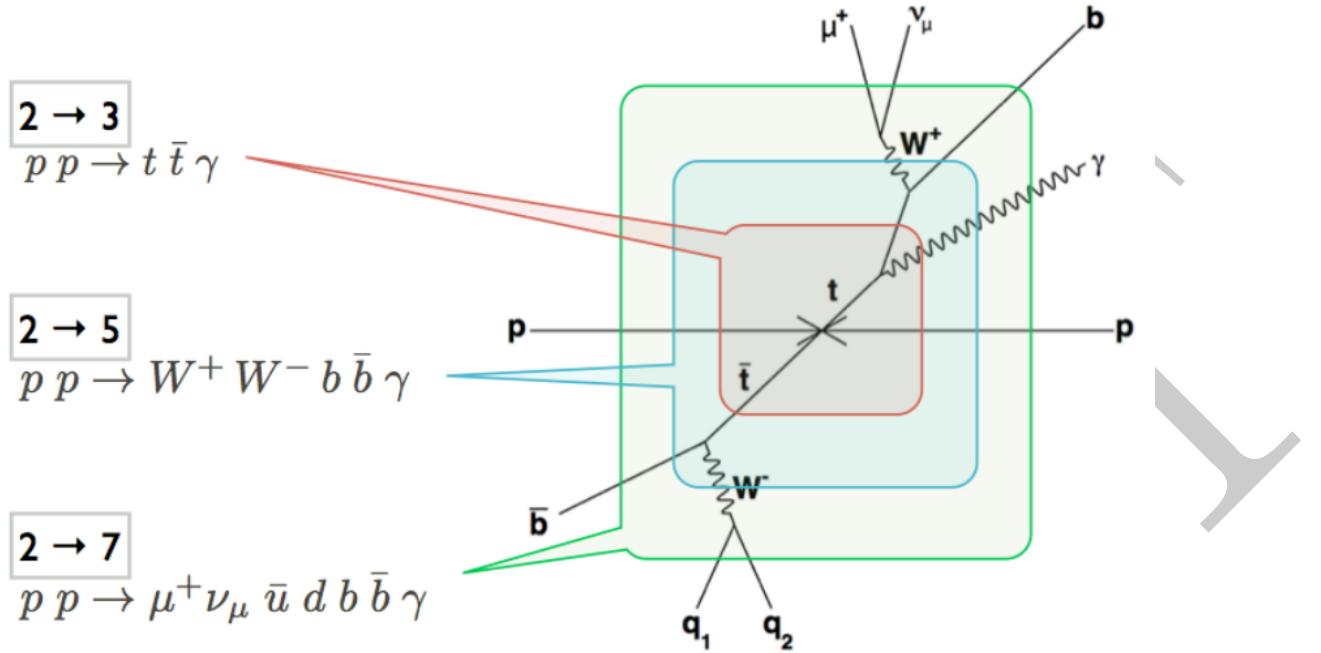


Figure 3.3: Process generation. The red, blue, and green boxes depict the matrix element calculation. Background processes with the same final state are included as well [7].

considered. The CPU time required in for tree level processes is moderate.

2 → 5 In this case, the decay of the top quark is included and thus photons that have radiated from a W boson or a b-quark, as well as interference effects between the two, must be taken into account. This is a significant process, as we must expect photons stemming from a W or b to contribute significantly to our signal. Photons that are radiated from the W are considered negligible, because the W decay products are highly boosted in top-quark events giving rise to, mostly, collinear emissions.

2 → 7 In this scenario we consider photon radiation and interference from all decay products. CPU time is much more intensive in this case due to the many more Feynman diagrams to be computed.

Originally, this analysis used the $2 \rightarrow 5$ technique with initial generator level cuts of $E_T > 20$ GeV and $\Delta R(\gamma, b/\bar{b}) > 0.1$ using WHIZARD [34], which is a leading order (LO) event generator. The variable factorisation and renormalisation scales are set to $m_{top} + E_T(\gamma)$, and a scale variation uncertainty of 8% has been applied to the WHIZARD $t\bar{t} + \gamma$ cross-section result, which gives 1.8 ± 0.5 pb as the SM expectation for the signal process, where the scale variation uncertainty, and uncertainty on the k-factor are added in quadrature.

3.6.1 Official $t\bar{t} + \gamma$ 2 → 7 process MADGRAPH Sample Production

Details of signal sample generation are described in [7]. The phase space for generation was chosen in the following way:

- $p_T(\gamma) > 13 \text{ GeV}$
- $|\eta(\gamma)| < 3.0$
- $\Delta R(\gamma, \text{all}) > 0.3$, where ‘all’ refers to any other generator level particle
- $p_T(\text{jet}) > 15 \text{ GeV}$
- $p_T(b) > 20 \text{ GeV}$
- $|\eta(b)| < 5.0$
- $|\eta(\text{jet})| < 5.0$
- $|\eta(\text{lepton})| < 3.0$
- $\Delta R(\text{jet}, \text{jet}) > 0.5$
- $\Delta R(\text{jet}, \text{lepton}) > 0.5$

There is no cut on lepton transverse momentum, but there are cuts on the momenta of quarks (jets). This makes the ratio of hadronic and leptonic W decays generated with these cuts differ from W branching ratio without any cuts.

3.7 Physics Object Reconstruction

The CMS experiment employs a complex algorithmic technique, Particle Flow (PF) [28], as part of a chain of reconstruction tools to reconstruct the full topology of events produced by collisions using information from each sub-detector. PF uses the information obtained from lower-level object reconstruction including tracking and clustering of energy deposits within each sub-detector. The primary goal of PF is to determine the object type, momentum, and energy for all objects within a singular event. Such objects include electrons, muons, charged hadrons, neutral hadrons, and photons. This information is then used to reconstruct higher-level objects, such as jets and MET.

3.7.1 Charged Particle Tracking

The tracking of charged particles begins with the reconstruction of tracks in the inner detector system. As the charged particles traverse the detector they deposit energy which we call hits, an iterative tracking algorithm then inputs the information from a number of hits in order to reconstruct tracks of individual particles [44]. By reconstructing the particle tracks we have access to information about the particle, such as the momentum before any effects from the magnet, and the impact parameter (IP), the distance of closest approach to this point used in later reconstruction.

Around two thirds of the particles than constitute a jet are charged particles, therefore if we are able to accurately reconstruct the tracks of charged particles in the detector, and the accuracy for correctly reconstructing a jet increases greatly. As all objects are reconstructed using the PF algorithm, which uses tracker information to fit tracks before reconstructing higher-level objects (Section ??) we see that information gleaned from tracker hits are vital for physics analysis. Several properties of tracks from the inner detector system contribute to its importance in reconstructing tracks, such as the p_T resolution which, for charged hadrons with a p_T of $\mathcal{O}(10^2)$ GeV, is significantly better than in the calorimeter systems.

One of the key parameters in object reconstruction is the track efficiency, the number of real tracks found over the number that incorrectly reconstructed or originate from a different source ('fake tracks'). Therefore, a high tracking efficiency is required in order to use the PF algorithm to it's full potential, and thus an iterative algorithm for tracking was designed [44]. The tracking algorithm can be broken down into roughly 5 processes, which are described below:

Local Reconstruction records signals in the silicon strips and pixels, converts them into 'digis' and then groups them into clusters. Both the position and uncertainty associated with an individual digi is calculated during this stage of reconstruction.

Track Seeding defines seeds, the basis of full particle reconstruction, by combining into 'pairs' and 'triplets' hits from the pixel detector of the inner tracker.

Pattern Recognition combines hits from the tracker and combines them in order to reconstruct potential particle trajectories, by starting at the centre of the tracker system and working outwards through each layer. The tracks are then filtered using a Combined Kalman Filter [45], which is a combinatorial variation on a Global Kalman Filter. All hits are taken into account from a single layer of the tracker and used as input informa-

tion for a ‘proto-track’, which is then used to calculate an estimate of the position and the uncertainty of the position in the next layer of the tracker. It is important that we must also take into account the energy loss as the particle traverses each layer of the detector. If we find several hits that are considered as signal, then we define multiple tracks. We must also include the case where a particle is highly energetic, such that it only leaves deposits within some of the tracker layers and not all.

Fitting The Kalman filter then fits each of the reconstructed hits twice from the different layers defined in the pattern recognition process. The first fitting process is performed using hits from central layers outwards, whereas the second fitting uses hits from the outer layers inwards. The two fits are complimentary to one another and remove any bias obtained during the track clustering stage.

Quality Checking removes tracks that are categorised as low quality, such that reconstructed hits have low hit multiplicity and a high χ^2 . Low quality tracks arise from aspects such as inconsistencies and misreconstructed hits in the track finding process; a single seed in the tracker layer can rise to multiple tracks, as well as multiple seeds originating from the same track, can be removed by performing the quality check of tracks.

The algorithm iterates over the process 6 times in order to filter out as many fakes/misreconstructed tracks as possible. For each iteration in the algorithm, cuts are loosened, such as the p_T , such that the tracking efficiency (ϵ) is increased while minimising the number of fakes. At the end of each iteration, after passing the track quality testing, high quality tracks are removed and the iterative process restarts. For the first four iterations only the seeds from the pixel detector are incorporated into the algorithm, where the last two use information of hits from the silicon strip detector. By performing track reconstruction in this manner we include tracks that decay outside the pixel volume, such as long-lived particles, such as heavy flavour hadrons and τ leptons, and photon conversions. At the end of the process, we are able to reconstruct tracks for particles with at least 3 hits in the tracking system with a p_T as low as 150 GeV, a primary vertex originating at least 50 cm from the beam axis, and all with a fake-rate of approximately 1% [44].

3.7.2 Primary Vertex Reconstruction

Our ability to identify and reconstruct vertices of particles in events has a large impact on the reconstruction of whole event topologies and kinematics. If we are to correctly assign

tracks to collisions, we must efficiently reconstruct the coordinates of the primary vertex. The same can also be said for secondary vertex reconstruction, used for the identification of heavy-flavour hadrons and τ leptons, as well as photon conversions. The main challenges encountered in vertex reconstruction stem from multiple overlapping events from a high track density and particle interactions within the tracker volume.

Once we have reconstructed all charged particle tracks to a sufficient degree of efficiency, we begin use this information to reconstruct the primary vertices for interactions within an event. Similar to the reconstruction of particle tracks, tracks must pass a selection criteria in order to be defined as originating from a primary vertex. A primary vertex must be defined as having a small impact parameter with respect to the beam line, a minimum number of hits in each layer of the pixel and silicon strips, and a small χ^2 significance ($\chi^2/N_{d.o.f}$). For the tracks that meet the imposed requirements, they are then clustered along the z-axis at their point of closest approach to the beam. We then define these as our primary vertex candidates and begin the fitting process.

An Adaptive Vertex Fitter (AVF) [46] is then implemented and uses a 3D fit to reconstruct the PV candidates, whereby each track is assigned a weight that is a function of its χ^2 contribution to the vertex. After each iteration of the fitting process, the given weights are translated to the new vertex position.

In order to discern the primary vertex from the hard-scattering event that we are interested in, vertices are listed in descending order of p_T . As we expect events from pile-up to have a lower sum of p_T to the hard process, we take the primary vertex with the highest p_T and assume the rest are from pile-up.

3.7.3 Calorimeter Clustering Algorithm

The granularity of the CMS HCAL is 25 times coarser than ECAL in order to separate spatially charged and neutral hadrons in jets with a transverse momentum above $100\text{ GeV}/c$. Therefore, we must employ a different algorithm for the identification and reconstruction of tracks within the different calorimeter systems. The calorimeter clustering algorithm is implemented to perform this task [28]. The calorimeter clustering algorithm works independently of the PF algorithm, however energy deposits from charged hadrons are matched using the PF algorithm in order to provide a more accurate measurement of the energy. The two algorithms working in tandem are able to resolve high- p_T and collinear tracks and thus reconstruct energy deposits from neutral hadrons and photons. Clusters of energy are formed by taking information from energy deposits within each calorimeter system, excluding the

forward HCAL where each cell is large enough that it is considered a cluster.

3.7.4 The Particle Flow Algorithm

3.7.5 Electron reconstruction

For the $t\bar{t} + \gamma$ analysis it is imperative that the identification and energy-momentum requirements are measured extremely accurately as final state electrons, which are required in two of the three decay modes, can imitate a photon by passing the same selection, and therefore contaminating signal events. In order for electrons to be reconstructed with these strict requirements they are processed in a specific way.

The CMS ECAL and tracker systems are extremely accurate detectors, however reconstruction of energy deposits within the ECAL is still a complex task due to the density of material. As a charged particle traverses the detector volume, the energy lost due to interaction with the material is not negligible. Most charged particles are heavy enough such that the energy lost manifests in the form of multiple Coulomb scattering as the particle passes between material. However, in the case of electrons, the dominant process by which the energy is lost is due to Bremsstrahlung, the process by which a photon is emitted upon interaction within a material.

Kalman fitters are a key tool used for the fitting of tracks in CMS due to their ability to incorporate noise amongst other inconsistencies, such as multiple scattering in track fitting, as Gaussian fluctuations. However, Bremsstrahlung radiation is non-Gaussian, and as a result electron tracks are poorly reconstructed with the standard Kalman filter fitting. To account for this, CMS provides a dedicated tool for the reconstruction of electrons using a Gaussian-sum filter [47]. This method is implemented by calculating the trajectory of electrons using a ‘relaxed’ Kalman filter, then re-fitted using a Gaussian-Sum Filter (GSF). The GSF method differs from the standard Kalman fitting method by computing uncertainties as the sum of multiple Gaussians rather than individual Gaussians. The downside to this method lies in the additional CPU time needed to process the events.

The PF algorithm uses two different techniques of electron identification that are used as seeds when reconstructing electron [48]. The first makes use of ECAL superclusters as seeds and projects back from the centre of the supercluster to the innermost layer of the pixel detector, and is thus known as ‘ECAL-seeding’. It makes use of ECAL track properties, such as a narrow width in η and a wide spread in the azimuthal angle, ϕ . The energy deposits of the object and its associated Bremsstrahlung form a single supercluster, where the performance

of this method lies the ability to correctly identify this. The method performs much more accurately in the high p_T region of the electrons, such that there are much fewer potential track seeds and energy clusters in the ECAL are less likely to overlap with deposits from other objects, such as jets. This is especially true if the electron is part of a jet structure, and thus not isolated. Moreover, a high track multiplicity can complicate the backwards propagation from a supercluster and mimic the signal of another object.

The second, tracker-driven, method is much more suited for the efficient reconstruction of non-isolated and low p_T electrons as they will most likely emit negligible amounts of Bremsstrahlung, and thus be fully reconstructed by extrapolating the tracks to superclusters. The cluster energy is then measured along with the track momentum, and if the ratio E/p is close enough to unity then the track is selected. However, Bremsstrahlung is not always negligible and other track properties are used in the calculation. For example, the number of hits recorded in the inner tracker and the χ^2_{KF} of the Kalman filter fit are taken into account before refitting using a GSF, before being characterised by a multivariate estimation tool [49].

After both procedures have been performed the two collections of electron candidates are merged into one where a GSF is run in order to determine the final properties of the object. It is important for the GSF to be run at this stage in the process such that more hits can be included in the reconstruction, thus providing a more accurate description of the electron's energy and momentum lost by interacting with the detector material. Upon completion of the process, the GSF electrons are passed to the PF algorithm.

3.7.6 Electron identification

Variable	Description
<i>Track quality variables</i>	
p_T	Transverse momentum of the GSF track
η	Pseudorapidity of the GSF track
$GSF \sigma_{p_T}/p_T$	Transverse momentum resolution of the GSF track
$\#hits_{KF}$	Number of reconstructed KF track hits
χ^2_{GSF} and χ^2_{KF}	GSF and KF goodness-of-fits
<i>ECAL shower variables</i>	

$\sigma_{i\eta i\eta}$	Cluster shape variable that gives a measure of the width of the cluster in η , using the distribution of energy in a 5 block of crystals around the seed crystal (the one with the highest energy) [50]: $\sigma_{i\eta i\eta}^2 = \sum_{5 \times 5 crystals} (\eta_i \eta_{seedcluster})^2 E_i / E_{seedcluster}$
$\sigma_{i\phi i\phi}^2$	Cluster shape in ϕ
$\eta_{SC}(\phi_{SC})$	Width of the super-cluster in $\eta(\phi)$
$1 - E_{1 \times 5} / E_{5 \times 5}$	$E_{1 \times 5}$ is the energy in the central 1×5 strip of the 5×5 electron cluster, and $E_{5 \times 5}$ is its total energy.
$E_{3 \times 3} / E_{SC,raw}$	Ratio of the energy in the preshower detector to the raw super-cluster energy (only in the endcap region).
<i>Longitudinal shower shape variables</i>	
H/E	Ratio of the hadronic energy associated with the electron candidate to the super-cluster energy. The hadronic energy is found by summing the HCAL towers in a cone of radius $\Delta R = 0.15$, centred at the super-cluster position.
$H/(H + E_e)$	Hadron fraction of the shower, where H is the energy of the hadron cluster linked to the GSF track.
<i>Track/super-cluster matching variables</i>	
$\Delta\eta_{in}\Delta\phi_{in}$	Distance in $\eta(\phi)$ between the super-cluster position and the extrapolated track position
$\Delta\eta_{vtx}\Delta\phi_{vtx}$	Distance in $\eta(\phi)$ between the super-cluster position and the position of the GSF track at vertex
$(E_e + E_\gamma) / p_{in}$	Ratio of the super-cluster energy to the inner track momentum
E_e/p_{out}	Ratio of the electron cluster energy to the outer track momentum
$1/E_{SC} - 1/p_T$	Difference between inverse super-cluster energy and inverse track momentum
<i>Bremsstrahlung variables</i>	
f_{brem}	Measured bremsstrahlung fraction, defined as: $f_{brem} = (p_{in} - p_{out})/p_{in}$, where p_{in} is the initial track momentum at the vertex and p_{out} is the track momentum at the last hit.
$E_\gamma/(p_{in-p_{out}})$	Ratio between the Bremsstrahlung photon energy as measured by ECAL and by the tracker

<i>EarlyBrem</i>	Flag of $E_e + \Sigma E_\gamma) > p_{in}$ inequality, corresponding to an electron emitting an “early” Bremsstrahlung photon, i.e. before it has crossed at least three tracker layers
<i>LateBrem</i>	Flag of $E_e > p_{out}$ inequality, corresponding to an electron emitting a “late” Bremsstrahlung electron, when the ECAL clustering is not able to disentangle the overlapping electron and photon showers

Table 3.2: Variables used in electron identification algorithms

[14].

3.7.7 Muon reconstruction

Reconstructing muons efficiently was one of the major focuses in the design of the CMS experiment, as can be seen from the name. A muon has a lifetime of around $\tau \sim 2.2 \mu\text{s}$, thus allowing them to travel a relatively large distance through the detector. Higher transverse momentum muons, $p_T > 20 \text{ GeV}$, are energetic enough to traverse the entire detector without feeling any effects from Coulomb scattering, while muons with a transverse momentum of $p_T < 5 \text{ GeV}$ are generally stopped within the detector volume and decay to an electron and two neutrinos via the electroweak force. Muons are also reconstructed using the particle flow algorithm (Section 3.7.4), however there are subtle differences between the standard usage of the PF algorithm and PF for muon identification, as will be described in this section. In relation to a top quark analysis, muons are produced from the W boson produced from the decay of the top. As a result, the muon will be produced at a close proximity to the primary vertex, and also well isolated.

In order to reconstruct muons, two types of tracks are used: tracks reconstructed from hits in the inner tracker (tracker hits), and in the muon system known as ‘standalone-muon’ tracks. Reconstruction in the muon system begins with hits in either the DC or CSC which have been merged to create short track segments. The Kalman Filter technique is then used to combine the segments into full tracks, as described in Section ???. We can then define muons in two ways [51]:

Tracker Muons are reconstructed such that every track found in the inner tracker with $p_T > 0.5 \text{ GeV}$ and $|p| > 2.5$ is regarded as a muon candidate. The standalone-muon

candidate is then extrapolated to the muon system, while taking into account factors such as the magnetic field, energy losses from traversing the detector volume, and uncertainties from multiple scattering. If the muon candidate satisfies both conditions of being in the tracker and extrapolated to the muon system, then we define the object as a muon. The method is also known as the ‘inside-out’ method.

Global Muons use the ‘outside-in’ method to reconstruct muons, conversely to tracker muons. The method works by taking a standalone-muon reconstructed in the muon system and matching it with a tracker muon by propagating onto a common surface. A global fit is then performed with the properties of both muon candidates, thus giving a *global muon*.

The two techniques for muon reconstruction are used in conjunction with one another due to the efficiency for tagging muons at different p_T thresholds. For the case of global muons, a muon with a higher p_T ($p_T > 20\text{ GeV}$) is more efficiently reconstructed due to the tracking depositing energy in each of the muon chambers. Conversely, muons with a lower p_T ($p_T < 5\text{ GeV}$) are more efficiently reconstructed in the tracking system, requiring a single muon segment. For the $t\bar{t} + \gamma$ analysis, the requirement of one or two reasonably high p_T muons means that it benefits from the use of both types of reconstruction method.

Once reconstructed, muons must undergo a series of quality checks in order to suppress punch through hadrons (hadrons that have a high enough energy to penetrate into the muon system from the HCAL). The following properties are used to select good muons for analysis.

- The number of hit in the muon chamber in the global muon fit.
- The number of muon stations with muon segments.
- Normalised chi-squared ($\chi^2/N_{d.o.f}$) of the global muon fit
- The number of pixel detector hits.
- The number of hits in the tracker layers.
- Transverse impact parameter d_{xy} (closest approach of the track to the primary vertex with respect to direction).
- longitudinal distance d_z of the tracker track with respect to the primary vertex.

Muons can also be produced within jets. The PF algorithm reconstructs such particles and identifies them in order to reduce the fake rate of charged hadrons being misidentified. This technique is essential for the measurement of missing transverse energy.

3.8 Higher-level Object Reconstruction

??

3.8.1 Jet reconstruction

We define a jet when a quark or gluon hadronises in an event, producing a narrow cone of collinear objects moving in the same direction. As described in Section ??, particles can not exist in an unbounded state as they carry a property known as colour and are confined within a hadron, and as a result they must fragment and form a hadron to be observed directly. Essentially, by reconstructing the sub-structures of jets, we are reverse-engineering the quantum-mechanical process of fragmentation and hadronisation. Therefore, if we wish to measure the initial energy and momentum, we must measure the combined properties of the reconstructed jet.

Jets play an integral part in the identification of top quark events where, at the very least, there are two jets in the final state due to the hadronisation of b quarks. In the di-lepton channels the case is as described previously, however in the semi-leptonic and fully hadronic top decay channels, at least four jets are present in the final state. The jets originate from the hadronisation of the b quark, and W boson decaying to quarks, which in turn hadronise to create jets.

In order to reconstruct jets, we first implement the PF algorithm in order to reconstruct single objects (muons, electrons, charged hadrons, neutral hadrons, and photons) before using clustering techniques to combine collinear particles to form jets. Various algorithms have been developed to compute the clustering of objects to form jets, such as the k_t [52], Cambridge/Aachen [53], and SISCone [54] jet clustering algorithms. However, the most widely used jet clustering algorithm is the anti- k_t [55] algorithm, which is used by the majority of the CMS collaboration. There are two main requirements that reconstructed jets must comply with:

- Collinear safe: should not be affected by collinear parton splitting.
- Infra-red safe: should not be affected by soft gluon emissions.

A low sensitivity to pileup and underlying events is also preferred in a jet clustering algorithm.

The anti- k_t algorithm is particularly insensitive to pileup and underlying events

We start by defining a distance parameter, $d_{i,j}$, between entities (particles, pseudojets), and distance parameter d_{iB} between entity i and the beam, B. The clustering method then begins by identifying the smallest of the distances, and if the smallest is $d_{i,j}$ then it recombines objects i and j, otherwise if the smallest distance is d_{iB} then i is defined as a jet and is removed from the list of objects. The procedure is then repeated in an iterative fashion until there are no more objects to run over. The distances $d_{i,j}$ and d_{iB} are defined as

$$d_{ij} = \min \left(\frac{1}{k_{t,i}^2}, \frac{1}{k_{t,j}^2} \right) \cdot \frac{\Delta_{i,j}^2}{R^2} \quad (3.1)$$

$$d_{iB} = \frac{1}{k_{t,i}^2} \quad (3.2)$$

where $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, $k_{t,i/j}$, $y_{i/j}$, and $\phi_{i/j}$ are the transverse momentum, rapidity, and azimuthal angle of each particle i and j, respectively. R is defined as the cone radius parameter. The anti- k_t algorithm outputs conical jets such that their boundaries are less susceptible to soft radiation.

3.8.2 Jet Identification

PF jet identification is used in order to reduce the amount of noise and the rate of electrons misreconstructed as jets. In order to correctly reconstruct jets, a number of observables are used as listed below:

- Number of particles in jet
- Neutral hadron energy fraction (NHF)
- Neutral electromagnetic energy fraction (NEF)
- Charged EM energy fraction (CEF)
- Charged hadron energy fraction (CHF)
- Charged hadron multiplicity (NCH)

3.8.3 Jet energy corrections

Upon reconstruction of jets, we find that the jet energy is not the same when comparing MC generator and detector-level jets, even when reconstructed using the same algorithm. We find this to be due to the non-linear and non-uniformity of the CMS calorimetry, and a mis-modelling between MC simulation and performance of the detector including areas where we see a large amount of electronic noise and pileup. In order to correct for such discrepancies, we must apply corrections to the reconstructed jet energy - Jet Energy Corrections (JEC). The overall goal is to achieve a detector response that is linear and uniform in η , where we define the detector response as the average amount of signal per unit energy deposited.

CMS has developed a factorised three tier system in order to apply jet energy corrections [56], such that each level corrects for a different effect. The corrections are then applied to the four-momentum of the jet. The three levels are described below.

- **L1 Pile-up** corrects for any energy that does not originate from the hard-scattering event, including detector noise from electronics, and pile-up events.
- **L2 Relative Jet Correction** derives an η -dependent scale factor in order to flatten the detector response in order to account for the non-uniformity of the detector. MC and data-driven methods are used to calculate the relative correction scale factors using the dijet balance method. The corrections are measured using only the barrel region ($|\eta| < 1.3$) and in bins of p_T , such that they are completely uncorrelated with the L3 absolute correction described below.
- **L3 Absolute Jet Correction** is a p_T -dependent correction used to correct back to particle level by applying absolute p_T correction. This is derived using MC truth information, or data-driven di-leptonic decays from $Z/\gamma^* + Jets$ samples, with the aim of flattening the jet response in p_T .

There is also an extra component in the correction factor calculation that arises from discrepancies between MC modelling and data. We define this step as **L2L3 Residual** and is only applied to data [57]. Higher order jet corrections, due to flavour-dependency, are also computed, however these are not included in this analysis. Uncertainties from each of the stages of correction are then convoluted into a single JEC systematic uncertainty included in the final result.

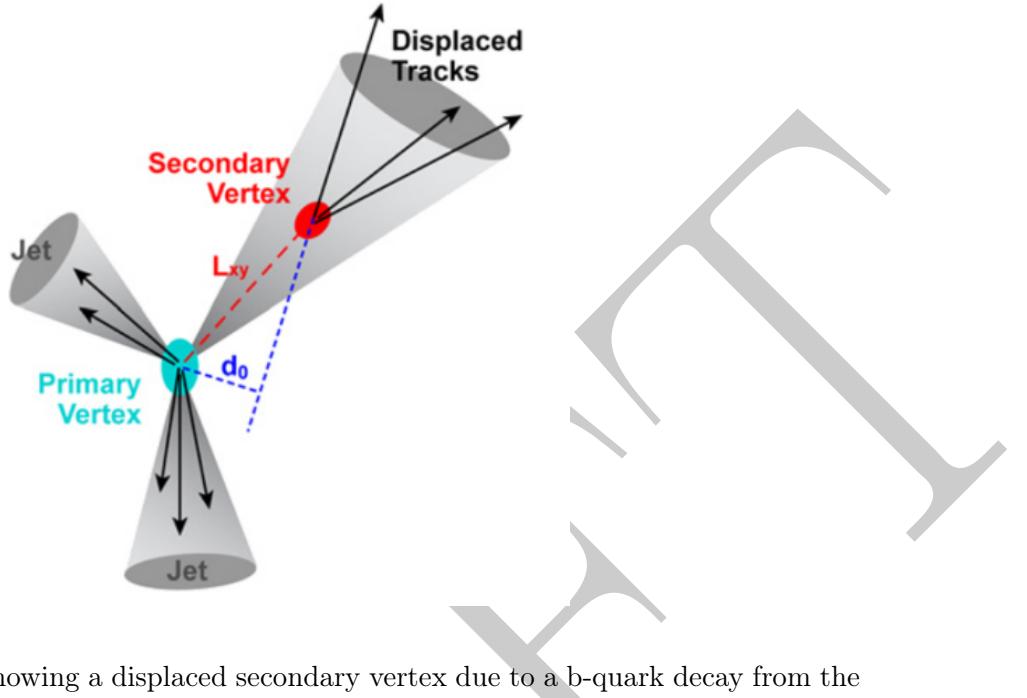


Figure 3.4: Schematic showing a displaced secondary vertex due to a b-quark decay from the primary vertex forming a b-jet. The impact parameter, d_0 , measures the displacement with respect to the primary vertex along the z-axis, and L_{xy} measures the displacement from the primary vertex in the transverse plane.

3.8.4 B-tagged Jets

We define a b-tagged jet to be a jet that manifests from the hadronisation of a b-quark. The process of identification of b-jets, b-tagging, is an extremely important and vital part of physics analysis, and top quark physics in particular. As the top quark decays to a W boson and a b quark almost 100% of the time, efficiently identifying b-jets significantly decreases the background contamination of our top quark signal process, and thus there is a strong need for b-tagging. Similar to non b-tagged jets, there are various algorithms that have developed by CMS that measure b-tagged jets [58], with the most successful being the Combined Secondary Vertex (CSV) algorithm. This method is derived from the prolonged life-time of the b-quark ($10^{-12} s$), and the fact that they decay to up and charm quarks, transitions which are Cabibbo suppressed by the standard model and can be seen in the CKM matrix. A highly-relativistic b-quark will travel through the detector around $450 \mu m$, as a product of its long life-time, which is an observable length that can be measured by the pixel detector. This allows for a displaced secondary vertex to be used as a tool for the identification of b-jets, as seen in Figure 3.8.4

B-tagged jets are particularly prominent due to their large mass, $4.18 \text{ GeV}/c^2$, and thus

a large multiplicity of charged particles are produced during hadronisation carrying the majority of the jet energy. The most commonly used b-tagging algorithm is the Combined Secondary Vertex [59] algorithm, which takes the following input parameters:

- Track multiplicity in jet.
- Track multiplicity at secondary vertex.
- Secondary vertex category.
- The invariant mass of charged particles from the secondary vertex must not exceed $6.6 \text{ GeV}/c^2$ which would surpass the mass of the bottom quark when convoluted with its uncertainties.
- The ratio of the sum of the track energy at the secondary vertex to the sum of track energy of all tracks in jet.
- η of the secondary vertex tracks with respect to the jet axis.
- The distance, L_{xy} , in the transverse plane between the primary and secondary vertex must be greater than $100 \mu\text{m}$ and be less than 2.5 cm .
- Impact parameter significance ($L_{xy}/\sigma_{L_{xy}}$) of the first track, increasing the invariant mass above the threshold for charm production of 1.5 GeV .

The algorithm begins by looking for a secondary vertex with the properties listed above. Using all particles in the jet, the algorithm uses a similar Kalman Filter [60] as for the reconstruction of primary vertices in order to reconstruct secondary vertices. If a secondary vertex is not found with this information, then a ‘pseudo-secondary vertex’ is computed using information from the tracks that were found to not be compatible with a primary vertex. The algorithm combines all of the information from the observables and calculates a discriminating value which is used to designate whether a jet is a ‘light jet’ (a jet that manifests from a u,d, or s quark), a charmed jet (from a c quark), or a b-tagged jet. The CSV discriminator output can be seen in Figure 3.8.4, showing the efficiency of correctly identifying a b-jet.

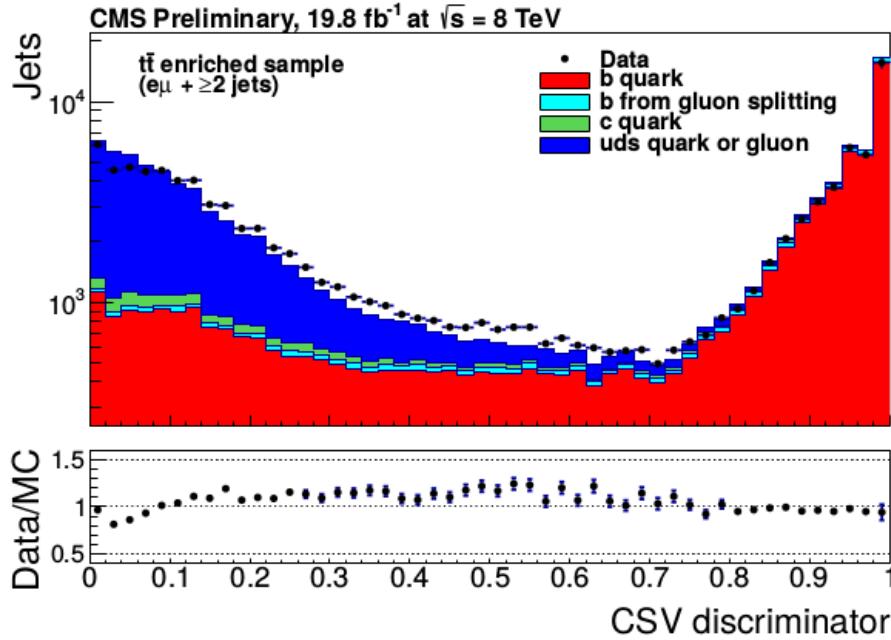


Figure 3.5: Logarithmic distribution of the partons as a function of the CSV b-tagging discriminant [8].

3.8.5 Particle flow jet identification

3.8.6 Reconstruction of Missing Transverse Energy

The conservation of momentum dictates that in an event, the sum of the p_T of all final state particles (partons) must be equal to 0, as would be the case in a perfect detector. However, this is not the case within the CMS detector as particles, such as underlying events and proton remnants, are able to traverse the detector without being reconstructed. This introduces an imbalance in the measured transverse energy deposited in the detector, which we call the Missing Transverse Energy (MET). We can define the MET in an event to be:

$$E_T^{miss} = - \sum_i (p_x^2 + p_y^2)^{1/2} \quad (3.3)$$

where p_x and p_y are the transverse momenta of reconstructed particles (partons) in the x and y axes.

The reconstruction of MET is of particular importance when dealing with a measurement where the presence of weakly-interacting standard model particles are measured in the final state, such as neutrinos. With respect to the $t\bar{t} + \gamma$ analysis, we expect a large amount of MET due to the presence of two neutrinos in the final state for each decay mode (excluding

the $e\mu$ channel) of the di-lepton channel, as the W from the top quark will for decay to a lepton-neutrino pair 1/9th of the time. These channels, and other similar analysis, rely on the efficient reconstruction of MET in order to accurately measure certain processes, and any misidentification or poorly reconstructed particles add to the MET sum further reducing the accuracy of a measurement.

Various methods on the reconstruction of MET in events have been conducted and are available [61]. The current most accurate method for the measurement of the MET is by using the PF algorithm, and is used throughout this analysis. The PF algorithm calculates the value of the MET from the list of PF particles, thus producing a ‘raw’ MET output which is systematically different from the final MET calculation. The ‘raw’ MET value is calculated without taking into account the non-linearity of the calorimeters, noise from electronics, pile-up, along with any other effects. We, therefore, implement a set of corrections as described below.

Type-0 corrects for the effects of pile-up on the MET.

Type-1 propagates the jet energy corrections into the MET (See Section 3.8.3).

xy-shift corrects for ϕ modulation in the MET.

The source of the ϕ modulation arises from misalignment of the detector relative to the beam, beam-spot displacement, an anisotropic detector response, and also broken or inactive cells within the calorimetry system. The correction for the shift in x-y is implemented in order to restore the spherical symmetry of events, and flatten the distribution in ϕ , thus correcting the measured distribution to the true distribution. This analysis makes use of all correction types using PF calculated MET.

Chapter 4

Event Selection & Background Estimation

When we finally arrive at a point when we have identified and reconstructed all the objects in the data, as described in Section 3, we impose a set of kinematic and topological cuts to each of the objects in order to provide a subset of the data containing mainly our signal events. The signal sample still contains background events and is corrected in several ways. Simulated MC samples are produced and used to optimise the number of signal events within this sample, and thus rejecting as much background as possible. The selection process, including kinematic, topological, and fiducial cuts on our final state objects is described in detail in the first part of this chapter.

Even though we model data using simulated MC samples, which are an essential tool for modelling distributions in particle physics, this is not always enough to provide a robust and accurate measurement of a process. In this case, additional methods for the calculation of background processes are used to purify our signal sample by removing events that are not in fact our final state signal event. These methods can be MC-driven and data-driven, and both are described in the second part of the chapter in greater detail pertaining to the $t\bar{t} + \gamma$ analysis.

4.1 Event Selection

Skim Preselection then photon selection

For the $t\bar{t} + \gamma$ analysis the selection of objects is computed in three stages; A **skim** is implemented when processing signal and background MC samples in order to reduce the

rate of events at analysis level and become much more manageable, a **pre-selection** of $t\bar{t}$ events is then computed for each final state respectively following the recommended top event selection group reference, and finally the full **selection** which includes an isolated photon radiated from top quark or its decay products.

It is important to reconstruct the number of $t\bar{t}$ events before we can include our radiated photon such that our sample is much cleaner. The pre-selection events have been constructed by following the CMS recommendation for cut-based selection of top-quark pair events with the requirement of at least two jets of which at least one is a b-tagged jet. Individual objects are reconstructed based on specific criteria, such as electrons, loose electrons, muons, loose muons, jets, and photons. Then an additional set of selection requirements is applied based on the relative positions of the objects (ΔR cuts). After that, the final decision is made if the event is to be considered in the further analysis. Pre-selection cuts are described in Section 4.2.

4.2 Pre-selection: Selection of $t\bar{t}$ Events

The pre-selection steps define our $t\bar{t}$ events before the addition of a radiated photon. The selection follows the recommended selection from the TOP Reference Selections and Recommendations (Run1) [62] designed to select di-lepton final states with two isolated oppositely charged leptons, at least 2 jets where at least one is a b-tagged jet, and two neutrinos from missing transverse energy. All objects in the selection are reconstructed using the PF algorithm as described in Section 3.7.4.

4.3 Trigger and Event Cleaning

4.3.1 Trigger selection

As the $t\bar{t} + \gamma$ analysis is studied as a di-lepton final state, the requirement of at least two oppositely charged leptons (electrons or muons only) is essential. These datasets are identified by the trigger system (as described in Section 2.8) to contain two leptons. Triggers are generally divided into two categories; single object triggers fire on one or more objects of the same flavour passing certain pre-selection requirements, such as p_T and η , and cross-triggers which select two objects of different flavours as predetermined by the user. For this analysis, both types of triggers are implemented in order to select the three final states in question.

The list of trigger paths can be seen in Table 4.3.1.

Final State	High Level Trigger Path
$\mu^+ \mu^-$	HLT_Mu17_Mu8_v*
$e^+ e^-$	HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL _Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*
$e\mu$	HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*, HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*s

Table 4.1: Triggers for each di-lepton channel.

Each trigger path name explains the selection requirements on the objects that it triggers on. The term Mu refers to a reconstructed muon and Ele refers to a reconstructed electron, where the succeeding number represents an associated energy threshold of the particle. For example, the the di-muon channel uses a single flavour object trigger to select two muons and is used with the requirements that one of the muons has a p_T greater than 8 GeV and the second greater than 17 GeV. The version of the trigger is denoted in the trigger path as v*, as the trigger path changes with the trigger table used. It should be noted that a different trigger version does not in fact require a change in trigger path. At this level the number of energy deposits within the calorimetry is still too large for the trigger rate to be usable, and thus extra selection requirements on the trigger system must be imposed.

A way to reduce the trigger rate is to impose cuts on the energy threshold of the particles in question greater than that required by the trigger, however this holds drawbacks for analyses who then wish to implement tighter cuts within offline analysis. Another method, and one that is used primarily in electron trigger studies, is to reduce trigger rates to a more feasible level is to introduce isolation and Id cuts, the ‘Iso’ and ‘Id’ terms that can be seen in the di-electron and $e\mu$ trigger path names. The objects must then also pass simple isolation and id criteria and thus reducing the trigger rate. The information for each is obtained from both the calorimeter (e.g CaloIso) and tracker (e.g TrkId) by placing requirements on such parameters as the shape of the energy cluster, the total number of energy depositions, and the angular separation between the ECAL and tracker energy deposits. Three categories of selection are implemented for each kinematic cut, and are listed as Tight (T), Loose (L), and Very Loose (VL) as can be seen in the trigger paths. These signify the harshness of the cuts when applied. This can be visualised, for example, in the the di-electron channel where the HLT calls for two electrons, where one must pass an energy threshold of 8 GeV with a tight requirement on calorimeter Id, and very loose requirements on calorimeter isolation and

tracker Id and isolation. The other electron must pass an energy threshold of 17 GeV with the same calorimeter and tracker isolation and Id cuts.

HLTs are used for the $t\bar{t} + \gamma$ analysis such that if the event does not pass the requirement of the trigger, then it is not included in the calculation. Single object triggers are used for both the di-muon and di-electron channels, and two cross-triggers were used for the $e\mu$ channel as the final state selection requires two oppositely charged leptons (electrons or muons) where one is an electron and the other a muon. The triggers were processed specifically for the $\sqrt{s} = 8$ TeV data-taking period with 19.6 fb^{-1} .

4.3.2 Filtering

Known anomalies derived from detector and accelerator effects are a prominent feature in the processing of data. To counter these effects we incorporate several ‘cleaning’ filters after trigger selection, but before any further selection cuts are applied. The first of which vetoes on **beam scraping** and includes a **tight CSC Beam Halo Filter**. We find that, even with the accuracy and precision that the LHC provides on accelerating bunches of protons, protons have a tendency to diverge radially from the bunch and form what is known as the beam halo that circulates the accelerator with the bunch. In early analyses it was found that the beam halo particles can be picked up in the detectors and be reconstructed as being part of an event in which it is not. Due to sensitivity to beam halo particles in the muon detectors, a filter is introduced based on muon tracking kinematics and thus veto on such events. Beam halo particles can also be removed from the beam within the LHC by introducing collimating blocks around the beam line at various points on the accelerator. This, however, presents another problem in the form of showering as the particles interact with the collimator blocks, beam scraping, and be detected by the experiments. These events are accounted for and removed from analyses by introducing the requirement that at least 25% of reconstructed tracks within the inner detector pass the high purity threshold (see Section 3.7.1).

Similarly, we employ a **HCAL noise filter** in order remove events with anomalous noise within the HCAL. CMS expects a certain degree of noise, stemming from the electronic of the detector, to be present when recording data, however the majority of anomalous noise is found to originate in the Hybrid Photo-Triodes (HPT) and their corresponding read-out boxes. At the current energy scale this is not a problem as the noise appears as large, isolated energy deposits. Anomalous events have easily-identifiable signatures such as the isolation and isolation of the HCAL readout, and the multiplicity in the read-out boxes. So, if a signal demonstrates very little change in the pulse shape over time, and the read-out boxes display

a high multiplicity, then an event is rejected. The next filter system comes in the form of the **HCAL laser filter**. The need for the laser filter first manifested during the 2011 data taking period when a much greater number of events were observed than was expected - ~ 5000 hits per event. The HCAL laser filter was then designed and introduced for the 2012 data taking period.

4.4 Di-lepton Selection and Veto

For the $t\bar{t} + \gamma$ analysis leptons mark our signature for our final state events. The leptons are taken from the list of PF-reconstructed objects and then required to pass additional selection cuts to refine the events further after trigger selection in Section 4.3. Along with our signal leptons, a set of ‘looser’ PF objects are selected as veto objects, such that our signal leptons are a subset of these objects. If an event has multiple loose selection leptons then the event is removed from the list of possible signal candidates.

The number of selected leptons differs for each decay mode, and thus three separate selections must be created for the di-muon, di-electron, and mixed channels, respectively. Cuts on leptons vary depending on the channel, but are taken from the recommended values produced by the central ‘Top Event Selection’ group [62].

4.4.1 Electrons

PF electron candidates are selected for di-electron and $e\mu$ final states if they have been identified by the GSF method, as described in Section 3.7.6 and pass the HLT for each channel respectively. The rates are then reduced by imposing a further set of cuts taken from the recommended top reference selection cuts [63]. The cuts are as follows:

- Passes a p_T threshold of > 20 GeV.
- Lies within the pseudorapidity region $|\eta| < 2.5$, excluding the EB-EE transition region $1.4442 < |\eta| < 1.5660$.
- The transverse IP of the electron (GSF) track with respect to the first offline primary vertex must be less than 0.04 cm.
- Combined relative Particle Flow (PF) ρ corrected isolation in cone 0.3 less than 0.15.
- Trigger version of electron multivariate discriminator Trigger MVAID greater than 0.5.

- Conversion rejection: there should be no extra tracks pointing in the same direction.
- The ratio of energy deposited in the HCAL over the energy deposited in the ECAL to be less than 0.05.

For electrons, an additional identification process is included, which uses a multivariate analysis to combine the information of several variables to produce a discrimination value between -1 and 1, such that the greater the number the more likely the event is to be an electron, (as described in Section 3.7.6). Depending on whether the HLT requires an electron or not, a different version of the discriminant is used.

One of the main criteria for lepton selection is the requirement of isolation. Generally, we define the isolation to be the sum of the p_T of the reconstructed objects within a cone by which we define the radius, and then divide by the p_T of the object. If we find that this produced a small number, then we say that the object is isolated. We must, however, include the effect of event PU into the calculation of isolation, and thus we introduce a correction factor. We are able to remove charged hadron tracks from the isolation sum if they do not originate from the event's primary vertex. For the case of neutral hadrons and photons which originate from PU, an effective area is defined for the electron and then an average energy is subtracted over this area. The effective area is extended into the ECAL due to the emittance of Bremsstrahlung radiation. For the case of electrons, we can define the isolation as so

$$I_\rho = \frac{I_{ChargedHadron} + \max(I_{NeutralHadron} + I_\gamma - \rho \cdot Eff.Area_{electron}, 0)}{p_T} \quad (4.1)$$

such that $I_{ChargedHadron}$, $I_{NeutralHadron}$, and I_γ are the are the isolation cones with a fixed radius of $\Delta R = 0.3$ containing the energy deposits for each category of particle; charged hadron, neutral hadron, and photon isolation. The ρ and $Eff.Area_{electron}$ parameters are the energy density of the event and the effective area for the electron which is calculated by taking the η_{SC} and electron p_T .

When a photon produced in collisions interacts with the detector material of the inner tracker it can pair-produce two electrons, thus mimicking the signature of an electron. The false signature has been calculated to represent a large source of fake electrons. The CMS EGamma working group have developed two methods in order to mitigate the fake electron signatures from the hard scattering process: measuring missing hits within the tracker system and measuring associated secondary tracks. The first technique measures the number of hits in the layers of the tracker and looks for any missing hits in the electrons associated track. If there are any missing hits, then the electron is considered as converted and discarded. The

second method requires a secondary electron/positron track such that it reconstructs the pair under a certain criteria. If a second track is not found to be within 0.02 cm in the $r - \phi$ plane and the $\cot \theta$ differs by less than 0.02 then the electron is considered a conversion.

We define a set of loose electron candidates by applying the recommended cuts as for our signal electrons but with less stringent requirements, such that our signal electrons are a subset of the loose electrons. The difference between loose and signal electrons is given by a p_T cut of > 10 GeV, non-triggering ID greater than zero, and several other cuts which are implemented for the identification of signal electrons are not included for loose. The **loose electrons** are defined to have the following cuts:

- Transverse momentum p_T greater than 10 GeV
- Absolute value of pseudorapidity less than 2.5
- Combined relative Particle Flow (PF) ρ corrected isolation in cone 0.3 less than 0.15
- Trigger version of electron multivariate discriminator Non-Trigger MVAID greater than 0

4.4.2 Muons

For our **signal muons**, once they have passed PF selection as described in Section 3.7.7, additional cuts taken from the recommended muon physics object group (POG) [64] are imposed as follows:

- Transverse momentum p_T greater than 20 GeV.
- Lies within the pseudorapidity region $|\eta| < 2.4$, excluding the EB-EE transition region $1.4442 < |\eta| < 1.5660$.
- Combined relative Particle Flow (PF) ρ corrected isolation in cone 0.4 less than 0.2
- Identified as a particle flow muon
- Identified as both a tracker and global muon

The analogous relative isolation for the signal muon candidates is pileup correction and much less complicated to compute. The technique is known as $\Delta\beta$ correction and removes

the neutral hadron and photon isolation from the isolation sum within a fixed cone of radius $\Delta R = 0.4$. The relative isolation can be computed as so

$$I_{\Delta\beta} = \frac{I_{ChargedHadron} + \max(I_{NeutralHadron} + I_\gamma - 0.5 \cdot I_{Pileup}, 0)}{p_T} \quad (4.2)$$

where the I_{Pileup} parameter is the neutral hadron energy within the cone, and the factor of 0.5 is a rough estimation of the ratio of neutral hadron to charged hadron in pileup events. We categorise a muon as being isolated if the isolation sum is less than $I_{\Delta\beta} < 0.2$ within an isolation cone of $\Delta R = 0.4$.

The production of muons from inn-flight decays is found to be much more prominent in data than is modelled in simulation. This results in a number of fake muons being recorded. In order to account for the number of fakes, the implementation of further cuts is required such that at least one hit is required in each of the pixel detector and muon detectors, and at least 6 hits recorded in the inner tracking system with two corresponding hits in the outer muon system.

Loose Muons are selected from PF muons failing the muon selection that have, and are selected to have less severe requirements as listed below:

- Transverse momentum p_T greater than 10 GeV
- Absolute value of pseudorapidity less than 2.5
- Combined relative Particle Flow (PF) ρ corrected isolation in cone 0.4 less than 0.2
- Identified as a particle flow muon
- Identified as both a tracker and global muon

4.4.3 Corrections to simulated events

4.5 Jet Selection and b-tag Requirements

Jets are reconstructed with the PF AK5 algorithm, anti- k_T with jet size parameter (R) of 5 in the jet reconstruction model, and the following selections are made (PF loose Jet ID). Before applying any selection the following corrections are made to account for imperfect jet energy measurement: Jet Energy Scale correction, and Jet Energy Resolution smearing, as described in Section 3.8.1. After jets are identified and reconstructed, the following set of cuts are implemented with the requirements:

- Transverse momentum greater than 30 GeV
- Absolute value of pseudorapidity less than 2.4
- Number of constituents greater than 1
- Charge multiplicity greater than 0
- Neutral hadron fraction of energy less than 0.99
- Neutral electromagnetic energy fraction less than 0.99
- Charged EM energy fraction less than 0.99
- Charged hadron energy fraction greater than 0

These cuts help to avoid picking detector noise and ECAL spikes as jets. We also impose a selection criterion that if a lepton lies within a cone of $\Delta R = 0.3$, then the lepton is included within the jet calculation. CMS improves the quality of reconstructed jets by the requirement that the energy deposits from a jet are recorded in both the ECAL and HCAL, where jets that manifest from anomalous deposits of energy in just one of the sub-detector are able to be removed from the sample [65].

B-tagged jets are identified with the Combined Secondary Vertex b-tagging algorithm using the loose working point (CSVL). Event re-weighting is applied to correct for the difference in b-tagging efficiency in data and simulation as explained in Section 4.4.3. The loose working point refers to a b-tagging efficiency of and a misidentification probability of 24.4%.

In each channel the requirement of at least 2 good jets, where a good jet passes all the aforementioned selection requirements, and at least one b-jet. By applying these requirements the contribution from the most prominent backgrounds, $t\bar{t}$ and events with additional loose jets, are removed.

4.6 Selection of $t\bar{t} + \gamma$ events

From the set of $t\bar{t}$ preselected events in Section ??, we select only events with a photon candidate present. Main fiducial requirements are implemented as cuts in $|\eta|$ and E_T . CMS recommended cuts are added for fiducialisation and photon selection (tight cut-based photon ID 2012 with particle flow based isolation, [66]). In order to suppress FSR from final state particles a ΔR cut of photons between jets and leptons is applied. All simulated data is scaled to the luminosity of the recorded data used.

Transverse energy A transverse energy cut of $E_T > 25$ is implemented in order to suppress the numerous low energy fake photons and photons from other vertices other than the primary vertex. The transverse energy distribution for each channel can be seen on the left in Figs.?? — ??.

Pseudorapidity An acceptance cut on the fiducial region of just the CMS ECAL barrel (EB), $|\eta| < 1.4442$, is applied to verify that the electromagnetic shower of the photon will be fully reconstructed. We will not be including the ECAL Endcap (EE) in this analysis due to the difficulty in identifying a photon thus giving very low statistics.

4.6.1 Cut based photon ID

Electron Conversion Veto A boolean to help distinguish an electron from a photon. We should not see a track seed in the pixel detector when identifying a photon.

Tower Based H/E The ratio of energy deposited in the Hadronic Calorimeter (HCal) over the fraction of energy deposited in the Electromagnetic Calorimeter (ECal). This cut is set so that the energy ratio should be less than 5% (see Fig.?? right).

Shower Width $\sigma_{i\eta i\eta}$ The shower shape weighted by energy (see Fig.?? right), defined as:

$$\sigma_{i\eta i\eta} = \left(\frac{\sum (\eta_i - \bar{\eta}) \omega_i}{\sum \omega_i} \right)^{1/2}; \bar{\eta} = \frac{\sum \eta_i \omega_i}{\sum \omega_i}; \omega_i = \max \left(0, 4.7 + \log \frac{E_i}{E_{5x5}} \right). \quad (4.3)$$

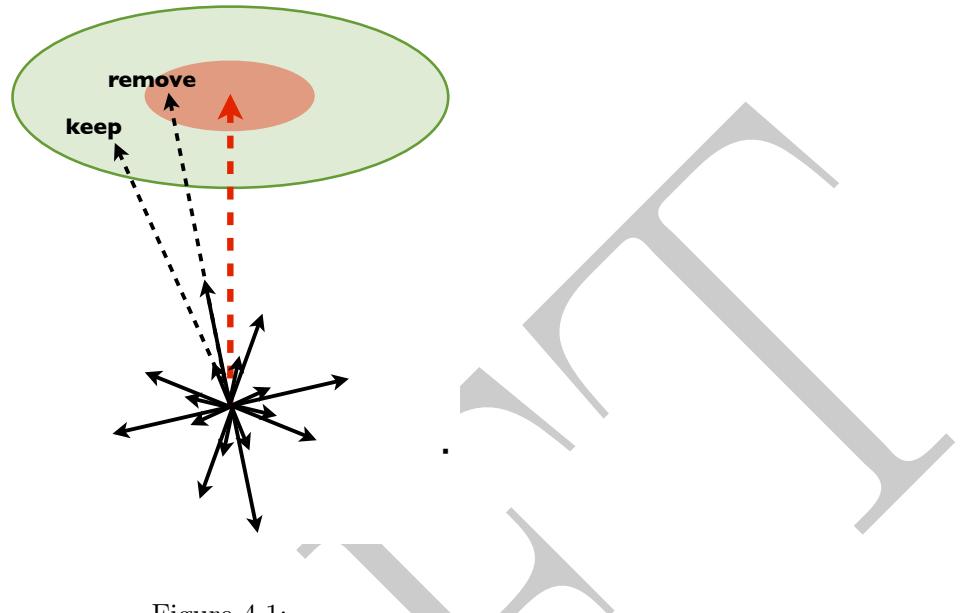
Charged Hadron Isolation The isolation of charged hadrons with energy density correction, ρ , applied. Cut given as $I_{char.had} < 1.5 GeV + 0.04 * E_T(\gamma)$. See Fig.??.

Neutral Hadron Isolation The isolation of neutral hadrons with energy density correction, ρ , applied. Cut given as $I_{neut.had} < 1.0 GeV + 0.005 * E_T(\gamma)$. See Fig.??.

Photon Isolation The isolation of the photon I_γ , with energy density correction applied. Cut given as $I_\gamma < 1.0 GeV + 0.005 * E_T(\gamma)$. See Fig.??.

super-cluster Footprint-removed Charged Hadron Isolation The super-cluster footprint-removed isolation of charged hadrons with energy density correction, ρ , applied. Cut given as $I_{char.had}(SCFR) < 5$ GeV. See Fig.??.

super-cluster Footprint-removed Neutral Hadron Isolation The super-cluster footprint-removed isolation of neutral hadrons with energy density correction, ρ , applied. Cut given as $I_{neut.had} < 1.0 GeV + 0.005 * E_T(\gamma)$. See Fig.??.



super-cluster Footprint-removed Photon Isolation The super-cluster footprint-removed isolation of the photon I_γ , with energy density correction applied. Cut given as $I_\gamma < 1.0\text{GeV} + 0.005 * E_T(\gamma)$. See Fig.??.

FSR suppression

$\Delta R(\gamma, \text{leptons})$ In order to diminish FSR in final state leptons, e.g photons radiated off high p_T muons, a minimum distance criterion in the $\eta - \phi$ plane is implemented. The cut is given as $\Delta R(\gamma, \text{leptons}) > 0.3$.

$\Delta R(\gamma, \text{jets})$ In order to reduce FSR in final state partons a minimum distance criterion in the $\eta - \phi$ plane is implemented. The cut is given as $\Delta R(\gamma, \text{jets}) > 0.3$.

$\Delta R(\text{leptons}, \text{jets})$ In order to reduce FSR in final state partons a minimum distance criterion in the $\eta - \phi$ plane is implemented. The cut is given as $\Delta R(\text{leptons}, \text{jets}) > 0.3$.

A significance test was performed in an attempt to optimise the ΔR cut between final state leptons and jets, however this proved inconclusive due to an initial cut of $\Delta R > 0.1$ at generator level.

4.6.2 Super-cluster footprint-removal for photon isolation

4.7 Background Estimation

Dataset	Run Range	Integrated Luminosity (pb^{-1})
/DoubleMuParked/Run2012A-22Jan2013-v1/AOD		876
/DoubleMuParked/Run2012B-22Jan2013-v1/AOD		4412
/DoubleMuParked/Run2012C-22Jan2013-v1/AOD		7017
/DoubleMuParked/Run2012D-22Jan2013-v1/AOD		7369
Total		19.7
/DoubleElectron/Run2012A-22Jan2013-v1/AOD		875
/DoubleElectron/Run2012B-22Jan2013-v1/AOD		4412
/DoubleElectron/Run2012C-22Jan2013-v1/AOD		7055
/DoubleElectron/Run2012D-22Jan2013-v1/AOD		7369
Total		19.7
/MuEG/Run2012A-22Jan2013-v1/AOD		876
/MuEG/Run2012B-22Jan2013-v1/AOD		4411
/MuEG/Run2012C-22Jan2013-v1/AOD		7055
/MuEG/Run2012D-22Jan2013-v1/AOD		7360
Total		19.7

Table 4.2: Dataset information for each run in each respective decay channel.

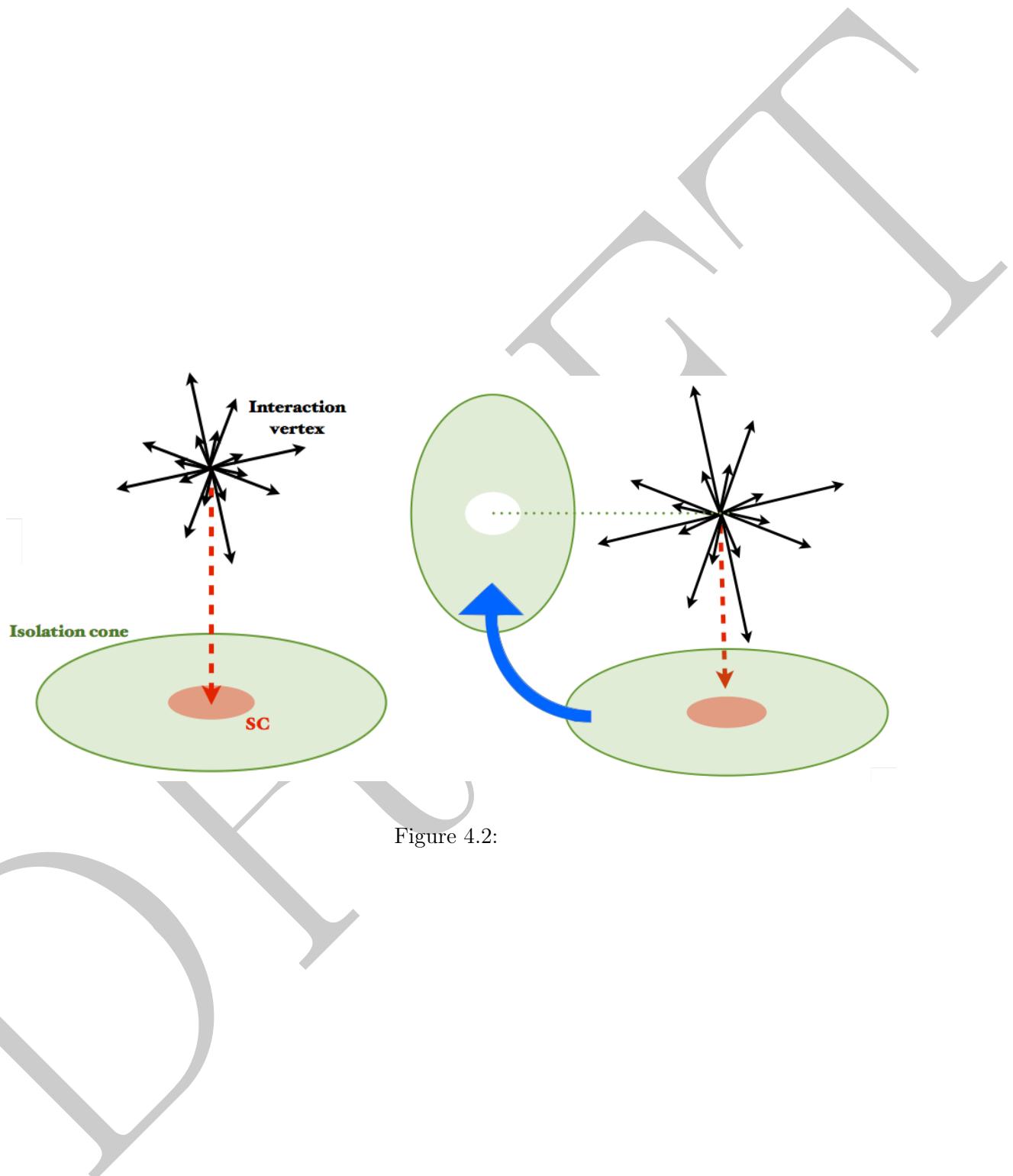


Figure 4.2:

Chapter 5

Measurement of the inclusive $t\bar{t} + \gamma$ cross-section

5.1 Signal Definition and Background Processes

5.1.1 Signal definition

5.1.2 Background processes

5.1.3 Backgrounds of photon signature

5.2 $t\bar{t} + \gamma$ Signal Simulation

WHIZARD

then

MADGRAPH

Factorised matrix element

5.3 Phase Space Overlap Removal

Events of the $t\bar{t} + \gamma$ process lie within a small region of $t\bar{t}$ phase space, and thus our signal sample events are expected to overlap with TTJets events in the case where a hard photon is radiated by initial state quarks, top quarks, b quarks, W and its decay products: electrons, muons, and their corresponding neutrino. In order to prevent the double counting of events we apply an overlap removal procedure to remove such events from our TTJets samples. In order for an event to be considered as overlapping with TTGamma, an event has to have at

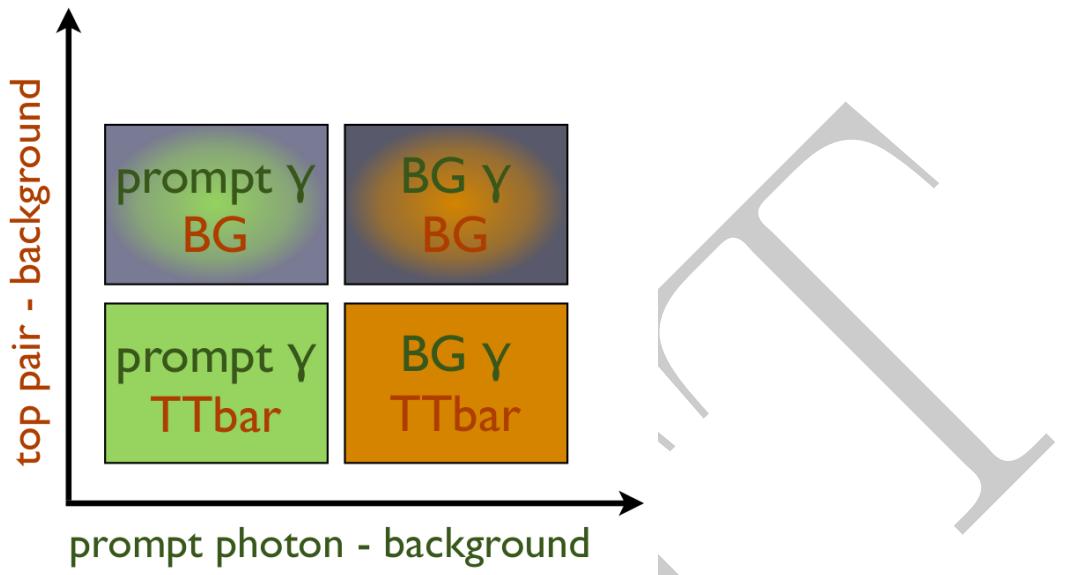


Figure 5.1: Graphic representation of the signal and background definitions [?].

least one generator-level photon with the following properties:

- $p_T(\gamma) > 13$ GeV
- $|\eta| < 3.0$
- Only gluons, bosons, or leptons are in the parents list. This ensures that photons from π^0 decays are not considered as signal
- $\Delta R(\gamma, \text{other}) > 0.2$ where other particles include leptons, b quarks and final state particles (hadrons, charged leptons, photons) with transverse momenta above 5 GeV.

The last cut is implemented in order to suppress photons from showers. In such cases the information from the parent particle will show that a photon is radiated by an electron, however the photon may be collinear with it. In particular in TTJets di-lepton events, such as described in this analysis, where a considerable fraction of the reconstructed photons comes from electrons radiating photons.

Similarly, we also observe an overlap between Z+Jets and ZGamma processes, and between W+Jets and WGamma samples, for the same reasons as described above. The phase space overlap removal procedure is applied on Z+Jets and W+Jets samples to remove events containing generator-level photons. Events containing generated photons are removed in the case in which they are from initial state radiation (emitted from the colliding partons) or

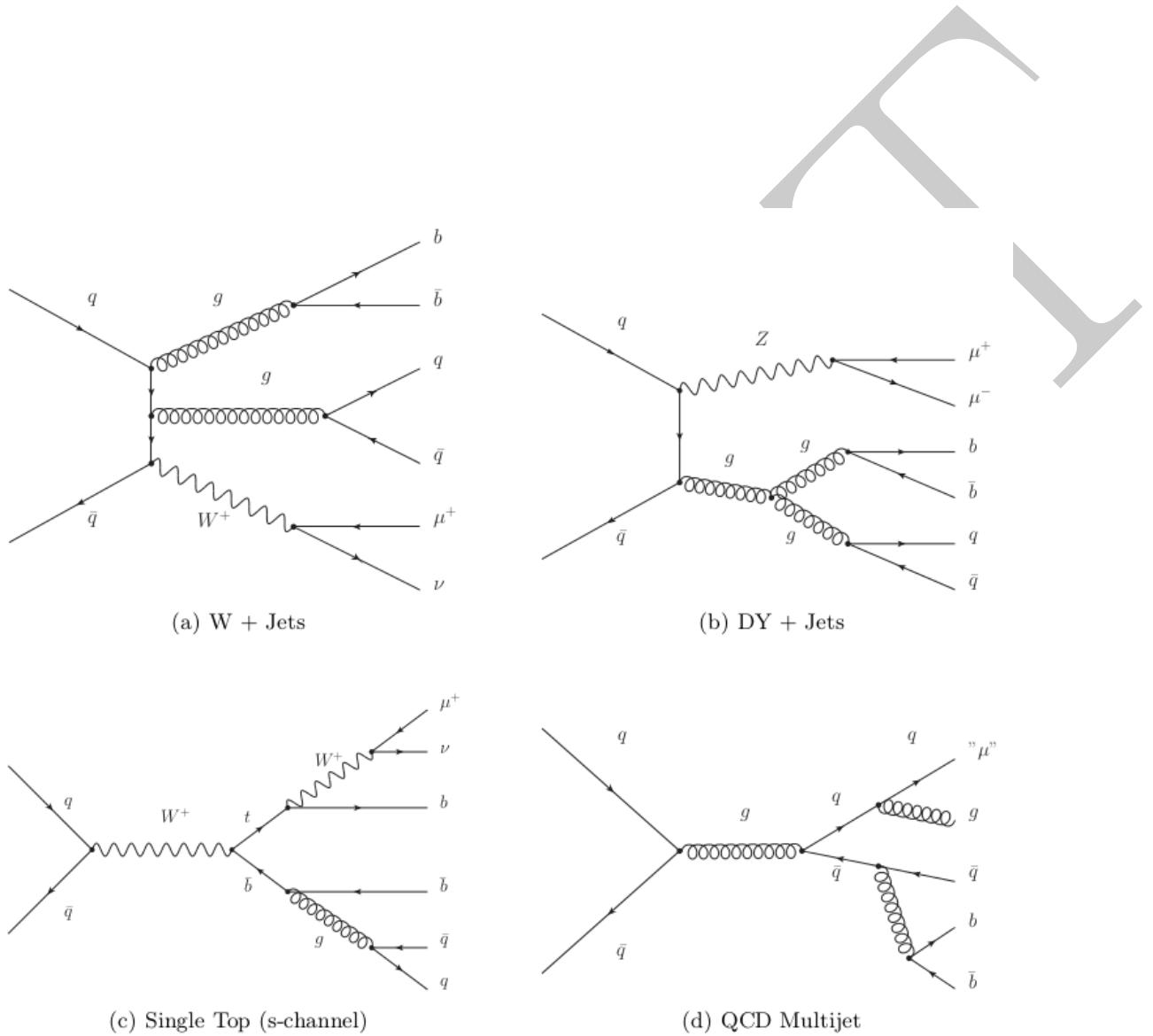


Figure 5.2: Examples of Feynman graphs of the considered $t\bar{t}$ backgrounds. QCD showering is necessary in all cases to obtain the required jet multiplicity [9].

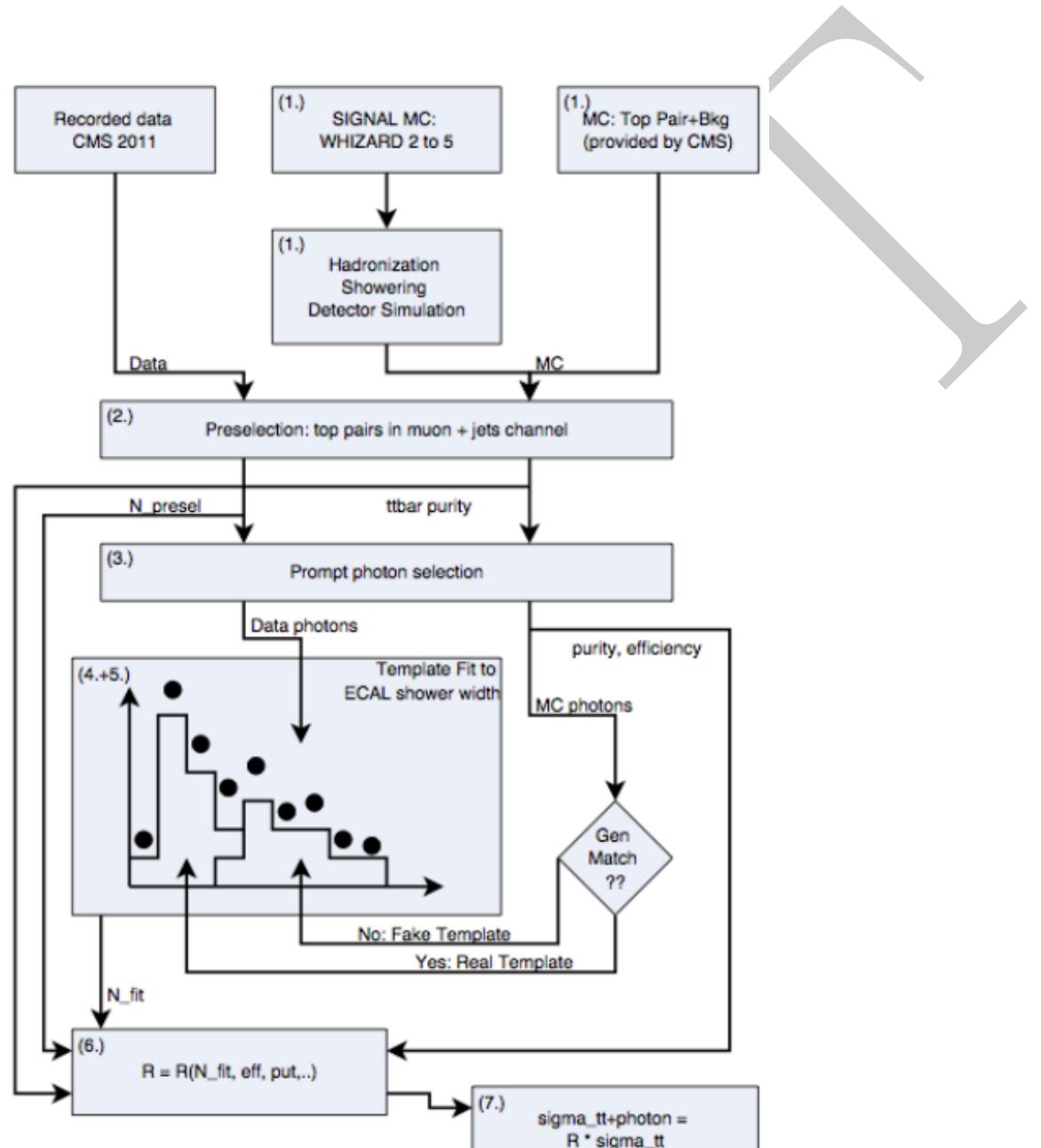


Figure 5.3: Flow chart showing each stage of the analysis. The box numbers represent the outlined analysis steps.

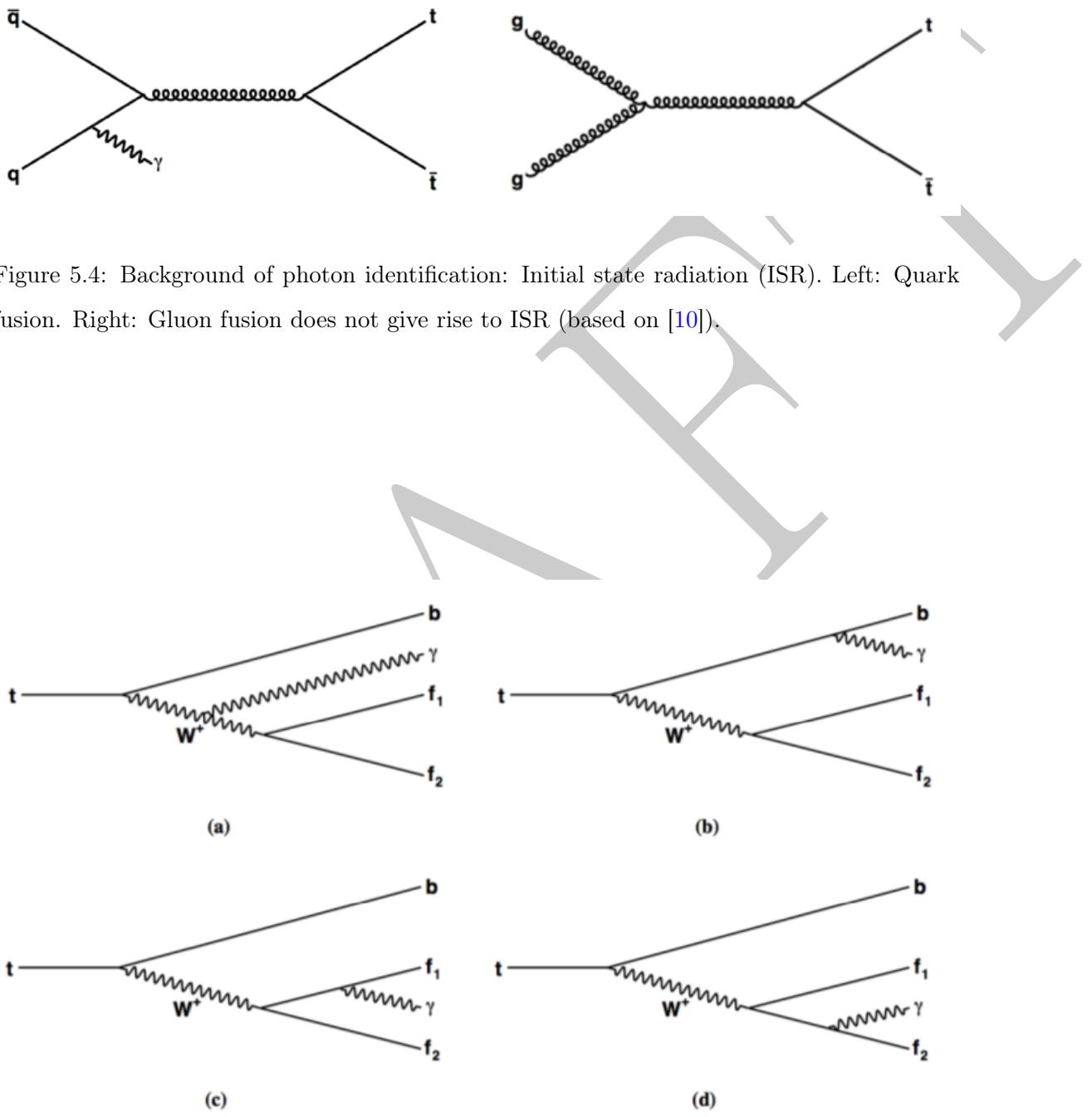


Figure 5.5: Background of photon identification: Final state radiation (FSR). All charged particles in the top-pair decay tree contribute to FSR [10].

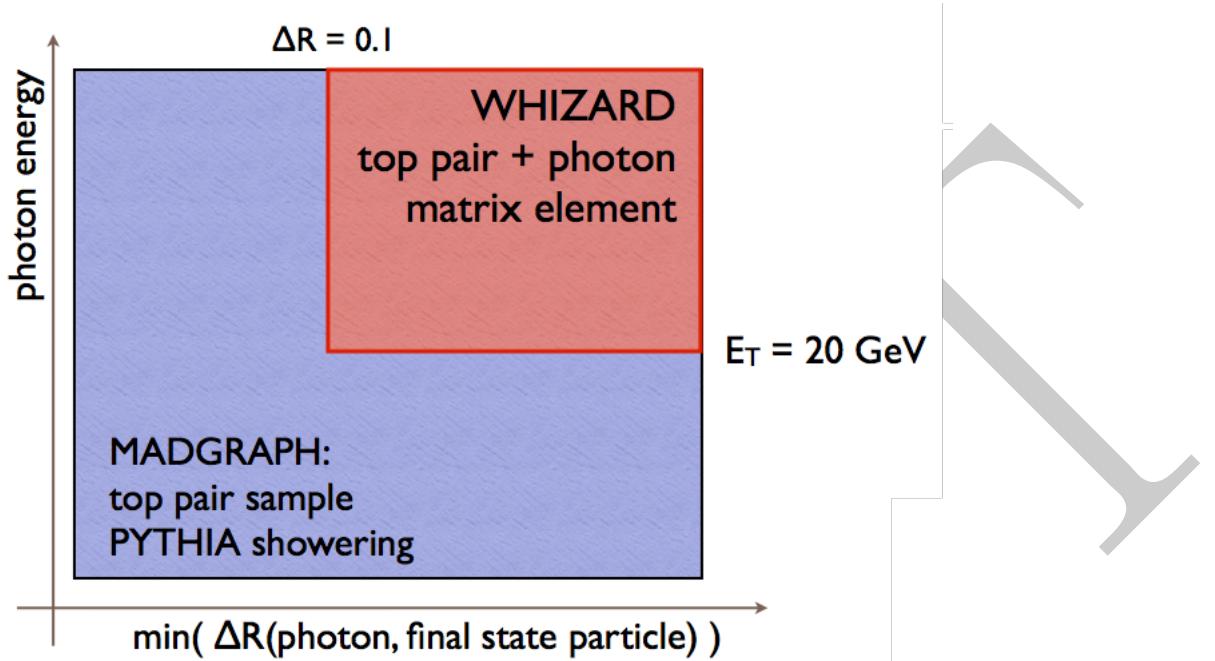


Figure 5.6:

final state radiation (emitted from W or Z bosons or their decay products), since these are already included in the WGamma and ZGamma simulations. The overlap removal procedure removes approximately one percent of the events in the W+Jets sample, and approximately three to four percent of the events from the TTJets and Z+Jets samples.

5.4 Event Selection

5.5 Photon Purity Estimation

5.6 Signal Acceptance Calculation

Acceptance calculation for this analysis differs from usual inclusive cross section measurements because we measure ratio of cross sections. The event selection is chosen to make use of this fact: two steps (top selection and photon selection) are done sequentially. For the inclusive $t\bar{t}$ process, we start with number of generated events (within some fiducial phase space) and count how many events are left after top event selection. The acceptance times efficiency is defined for the $t\bar{t}$ top selection as:

$$\epsilon_{top}^{t\bar{t}} \cdot A_{top}^{t\bar{t}} = \frac{N_{t\bar{t}.preselection}}{N_{t\bar{t}.generated}} \quad (5.1)$$

This gives the acceptance times efficiency of the inclusive $t\bar{t}$ process to be $\epsilon_{top}^{t\bar{t}} \cdot A_{top}^{t\bar{t}} = \pm(stat.)$ in the di-muon channel, $\epsilon_{top}^{t\bar{t}} \cdot A_{top}^{t\bar{t}} = \pm(stat.)$ in the di-electron channel, and $\epsilon_{top}^{t\bar{t}} \cdot A_{top}^{t\bar{t}} = \pm(stat.)$ in the mixed final state.

The same can be done for the $t\bar{t} + \gamma$ (signal sample). To get the acceptance times efficiency we have to divide the number of events passing top selection by the total number of events considered. However, in this case we have to choose what we take as the denominator. The signal $t\bar{t} + \gamma$ sample is inclusive, but theoretical calculations for the cross section are done for final states with 1 and 2 leptons [67]. To make comparison with theory easier we consider the fiducial space for signal when 1 or 2 leptons are present. The $t\bar{t} + \gamma$ acceptance times efficiency of the top selection is defined for the signal samples with 1 or 2 leptons as:

$$\epsilon_{top}^{t\bar{t}+\gamma} \cdot A_{top}^{t\bar{t}+\gamma} = \frac{N_{t\bar{t}+\gamma.\text{preselection}(1\text{or}2l)}}{N_{t\bar{t}+\gamma.\text{generated}(1\text{or}2l)}} \quad (5.2)$$

Giving the acceptance times efficiency of the inclusive $t\bar{t} + \gamma$ process to be $\epsilon_{top}^{t\bar{t}+\gamma} \cdot A_{top}^{t\bar{t}+\gamma} = \pm(stat.)$ in the di-muon channel, $\epsilon_{top}^{t\bar{t}+\gamma} \cdot A_{top}^{t\bar{t}+\gamma} = \pm(stat.)$ in the di-electron channel, and $\epsilon_{top}^{t\bar{t}+\gamma} \cdot A_{top}^{t\bar{t}+\gamma} = \pm(stat.)$ in the mixed final state.

The acceptance and efficiency for the $t\bar{t} + \gamma$ sample includes a term for photon selection. This is found based on the ratio of the number of events in $t\bar{t} + \gamma$ passing photon selection and the reconstructed photon matched to a generated photon over the number of events passing the top selection. The $t\bar{t} + \gamma$ photon selection acceptance times efficiency is defined as

$$\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \frac{N_{t\bar{t}+\gamma.\text{photonselection}(1\text{or}2l)}}{N_{t\bar{t}+\gamma.\text{preselection}(1\text{or}2l)}} \quad (5.3)$$

Thus, yields of the acceptance times efficiency of the inclusive $t\bar{t} + \gamma$ process to be $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \pm(stat.)$ in the di-muon channel, $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \pm(stat.)$ in the di-electron channel, and $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \pm(stat.)$ in the mixed final state.

The calculation of signal acceptance explained above is done for the sake of comparison with theoretical prediction. The biggest difference in the generated phase space and analysis selection is the transverse energy cut on the photon (13 GeV in generated sample and 25 GeV in analysis). We are measuring the value for the higher E_T cut on the photon and using the efficiencies to extrapolate back to the full generated phase space. In order to avoid this propagation of the result into the larger phase space we also quote the *visible cross section ratio*, where the cross section is measured in the fiducial region with generated photons having transverse energy of at least 25 GeV and $|\eta| < 1.4442$. For the visible cross section ratio, the photon selection term includes only the photon reconstruction efficiency because, by

definition, we are considering events where generator photon passes analysis level p_T and η cuts.

The visible top selection efficiency in $t\bar{t} + \gamma$ is taken to be the ratio of the number of $t\bar{t} + \gamma$ events passing the preselection (with a generated photon having $p_T > 25$ GeV and $|\eta| < 1.4442$)

$$\epsilon_{top}^{t\bar{t}+\gamma Vis} = \frac{N_{t\bar{t}+\gamma.\text{preselection}(1lor2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma^{gen})| < 1.4442)}{N_{t\bar{t}+\gamma.\text{generated}(1lor2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma^{gen})| < 1.4442)} \quad (5.4)$$

The $t\bar{t} + \gamma$ visible photon selection efficiency is calculated as the ratio of $t\bar{t} + \gamma$ events passing to photon selection with a reconstructed photon matched to a generated photon over the number of events passing top selection and with an isolated generator level photon passing the $p_T > 25$ GeV and $|\eta| < 1.4442$ cuts used in the photon selection.

$$\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = \frac{N_{t\bar{t}+\gamma.\text{photonselection}(1lor2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma^{gen})| < 1.4442)}{N_{t\bar{t}+\gamma.\text{preselection}(1lor2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma^{gen})| < 1.4442)} \quad (5.5)$$

The visible top selection efficiency is found to be $\epsilon_{top}^{t\bar{t}+\gamma Vis} = 0.0712 \pm 0.0005$, $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = 0.0928 \pm 0.0006$, and $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis}$ in the di-muon, di-electron and mixed channels, respectively. The value of the visible photon selection efficiency is found to be $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = 0.287 \pm 0.004(\text{stat.})$ for the di-muon final state, $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = 0.286 \pm 0.004(\text{stat.})$ for the di-electron channel, and $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = 0.287 \pm 0.004(\text{stat.})$ for the mixed final state.

Chapter 6

Systematic Uncertainties

Upon studying such a decay, large statistical uncertainties, comparable to the systematic uncertainties on the measurement, arise due to the small cross-section of the $t\bar{t} + \gamma$ process and small branching fraction of the decay channel. In order to perform a scientifically solid measurement we must take into account and fully understand all systematic uncertainties associated with the analysis. To begin with, we can categorise the errors into two broad categories:

Flat Rate Uncertainties - These uncertainties manifest in the form of detector performance factors, event reconstruction algorithms, and other such aspects as theoretical cross-sections which affect the overall rate of a particular process. Each uncertainty is almost universal in that it affects nearly all analyses within the collaboration, and are thus studied within their own dedicated performance group. A more detailed description can be found in [6.1](#)

Scale-factor Uncertainties - In analyses there are often scale factors applied to scale Monte Carlo to data in order to correct for inconsistencies between the two. These can arise due to such aspects as the theoretical input parameters of the Monte Carlo generators, which are used to model signal and background processes, not taking the true shape of the data. These types of scale factors affect all distribution shapes in an analysis and therefore must be accounted, and thus an uncertainty on the scale factor is applied by varying the value up and down by one standard deviation, $\pm\sigma$, and measuring the impact that this variation has on the final result. An in-depth description of each of these types of systematic uncertainties is given in [6.2](#).

Once computed, the systematic uncertainties are introduced as nuisance parameters within the fitting process. The final uncertainty to be considered in the fit is the statistical uncertainty that dominates this particular decay mode. This is discussed in greater detail in section 7.

Blah blah [68] [69]

6.1 Flat Rate Uncertainties

6.1.1 Luminosity

The CMS collaboration measures instantaneous and integrated luminosity in two ways; one way is by means of a coincidence trigger in the forward hadron calorimeter sub-detector, and also by counting the number of clusters measured by the pixel detectors. The former method was used at the beginning of runs in the LHC, but then ran into difficulties when the number of PU increased and shifts in calibration. This lead to the development and implementation of the pixel-based calculation in 2011 - the *Pixel Cluster Counting* (PCC) method [70].

The PCC method evaluates the number of pixel clusters that occur on average for a zero-bias event (an event triggered by the requirement of only two bunches crossing at the CMS IP). It assumes that there is a small probability that each pixel within the silicon inner detector is part of more than one track per bunch crossing, and thus it is assumed that the number of pixel clusters scales linearly with the number of interactions in any given bunch crossing. This gives an excellent measure of the luminosity within the detector. Measured rates are calibrated by the method of a Van de Meer scan [71]. The total calculated integrated luminosity for the entire 2012 dataset was measured to be 23.27 fb^{-1} .

Although the total integrated luminosity is measured to be the value described above, the true value that we measure is less due to a number of technical reasons. Quite often a sub-detector may encounter problems at the start of the run and may require rebooting or re-calibration, thus a period of "dead time" is induced such that data is unable to be used for physics analysis, and therefore given the title of 'bad' data. The remaining measured luminosity entitled 'good' is provided to analysts by the Run Coordination team, and is measured to be 19.7 fb^{-1} for the CMS experiment with the full 2012 dataset at, where a flat rate associated uncertainty of 2.6% is assigned [72]. Each simulated sample used in analysis is scaled to the luminosity of the dataset used, and thus the associated uncertainty affects the normalisation of every physics process.

6.1.2 Lepton Efficiencies

Lepton efficiencies are measured and implemented in order to correct Calculated by TP

6.2 Shape Uncertainties

6.2.1 Parton Density Function

Parton Density Functions (PDFs), denoted as $f_i(x, Q^2)$, give the probability of finding a parton of flavour i (quark or gluon) carrying momentum fraction x of the proton momentum where Q is the resolution scale of the hard interaction. Cross-sections are calculated by convoluting parton level cross-sections with PDFs. Due to the non-perturbative nature of partons not being observed as free particles, we cannot fully obtain PDFs by perturbative QCD alone. The shapes of PDFs are determined from global fits to data from experimental observables in various processes, such as deep inelastic scattering (DIS), Drell-Yan, and jet data using the DGLAP evolution equation [73]. PDFs are updated by the collaborations who perform the fits, such as CTEQ [74], each time new data or theoretical predictions become available.

The set of PDFs used in this analysis are taken from the CT10 [75] set. CT10 provides the nominal PDF weight along with 25 eigenvalues, which provide 50 alternative weights for each event

6.2.2 Pile-up Re-weighting

Another example of a process that is not described well in simulation compared to data is PU. Additional pileup interactions are included within the simulated samples, however the true number of primary vertices in simulation does not match the number observed in data correctly. This discrepancy between simulation and data gives rise to an incorrect estimation of signal and/or background events in an analysis. In order to correct for this effect, additional corrections must be applied to all simulated samples. The PU re-weighting comes into fruition when dealing with the ever changing instantaneous luminosity of the beams, and thus the change in number of primary vertices in a single data-taking period. In order to implement the PU re-weighting, the number of primary vertices is re-weighted to match the current running conditions in the LHC, for example the number of primary vertices changes with the energy and luminosity of the beams. The obtained uncertainty is then included in the systematic uncertainty on the final results of the analysis.

6.2.3 Jet Energy Corrections

As described in Section ??, it is necessary to apply corrections to reconstructed jet energies in order to counteract the discrepancies between generator level and detector level jets. These jet energy corrections are a set of tools included to account for non-linearities in the calorimeter, and to give a flat jet response in η and E_T as it is not trivial to translate the measured jet energy to the true particle or parton energy. The resulting jet energy corrections and associated uncertainties are measured by the JEC group who then provide the results to the collaboration to be used in analysis [57, 76].

When we change the jet energy scale (JES) in analysis, the kinematics of each jet within an event are also modified. As a result, the number of jets that pass, or fail, the event selection requirements is likely to change, whereby altering the final topology of an event. This will have a significant impact on the final result. In order to measure the JES significance, the correction factors are varied up and down by one standard deviation, σ , and propagating the effects through to the MET.

The jet energy resolution (JER) is measured as the standard deviation of a Gaussian that is fitted to the jet response of the detector. The JER in data has been found to be worse than the JER in simulation, $\sim 10\%$ broader, and has an associated uncertainty of a similar size [77]. We correct for this effect by smearing the 4-momentum of jets in MC as a function of the true and reconstructed p_T and η . To obtain our up and down systematic samples for the jet energy resolution, which are then included within the analysis as nuisance parameters, the smearing is applied twice for up and not at all for down.

6.2.4 Missing Transverse Energy

Events that contain neutrinos in the final state are affected mostly by uncertainties from modelling of MET from simulation. The way that the MET is calculated is by taking the sum of the p_T of all PF-reconstructed objects, including ‘unclustered’ energy deposits, and thus uncertainties from these propagate into the calculation of the MET. Unclustered energy is defined as recorded energy deposits that have a low p_T and/or not included in a calorimeter energy deposit cluster due to isolation requirements. PF-reconstructed objects are already corrected for during the reconstruction process (ρ -correction, etc), however this is not the case for unclustered energy deposits. Because the unclustered energy is not corrected for during reconstruction, this is where the largest, most prominent source of uncertainty arises. In order to measure the uncertainty on the MET, we remove the p_T of all PF-reconstructed

objects from the MET calculation, the residual energy is scaled up and down by 10%. Other uncertainties that affect the MET, such as JES and JER, are propagated on calculation and thus included in their respective uncertainties. Effects ee and mumu channels due to MET cut and neutrino final states

6.2.5 B-tagging Efficiency

Studies of b-tagging efficiencies and misidentification rates are conducted by the b-tag and vertexing group (BTV) and scale factors are produced to correct for discrepancies between data and MC simulation. For the Run I data-taking period at 8 TeV, the BTV group performed studies using $t\bar{t}$ and multijet samples [78]. The given samples were chosen such that studies could be performed using events with at least two jets, and a choice of the number of leptons.

6.2.6 Data-driven Reweighting

6.3 Modelling Uncertainties

6.3.1 QCD Renormalisation and factorisation scales

6.3.2 Parton-level matching thresholds

6.3.3 Analysis-dependent modelling uncertainties

6.4 Impact of uncertainties

Source	Uncertainty (%)	
	Statistical	Systematic
Pileup (PU)		
Out-of-time Pileup (OOT)		
Top P_T		
b-tag		
Photon E_T		
JEC		
JER		
Electron Efficiency		
Electron P_T		
PDF		
Total		

Table 6.1: Systematic uncertainties and their contribution to the cross-section ratio.

$t\bar{t}$ scale up	/tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TTJets_scaleup_TuneZ2star_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TTJets_scaledown_TuneZ2star_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	5009488 5387181	
$t\bar{t}$ scale down	Drell-Yann, $10 < m_{ll} < 50$ Drell-Yann, $m_{ll} > 50$ matching up Drell-Yann, $m_{ll} > 50$ matching down Drell-Yann, $m_{ll} > 50$ scale up Drell-Yann, $m_{ll} > 50$ scale down	/DYJetsToLL_M-50_matchingup_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /DYJetsToLL_M-50_matchingdown_8TeV- madgraph/Summer12_DR53X-PU_S10_START53_V7A- v1/AODSIM /DYJetsToLL_M-50_scaleup_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /DYJetsToLL_M-50_scaledown_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TTToDilepton_tW-channel-DR_scaleup_8TeV-powheg- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TTToDilepton_tW-channel-DR_scaledown_8TeV-powheg- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TBarToDilepton_tW-channel-DR_scaleup_8TeV-powheg- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TBarToDilepton_tW-channel-DR_scaledown_8TeV-powheg- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /WJetsToLNu_matchingup_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /WJetsToLNu_matchingdown_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /WJetsToLNu_scaleup_8TeV-madgraph- tauola/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM	1985529 2112387 2170270 1934901 1492816 497658 1492534 1493101 21364637 21364637 20784770

Chapter 7

Results

The values used for measuring the cross section ratio can be found in Table 7.

Value	$\mu^+ \mu^-$	$e^+ e^-$	$e\mu$
Number of signal events, N_{signal}			
$t\bar{t} + \gamma$ Top Selection $\epsilon_{top}^{t\bar{t}+\gamma} \cdot A_{top}^{t\bar{t}+\gamma}$			
$t\bar{t} + \gamma$ Photon Selection $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma}$			
$t\bar{t}$ Number of $t\bar{t}$ events $N_{t\bar{t}}$			
$t\bar{t}$ Top Selection $\epsilon_{top}^{t\bar{t}} \cdot A_{top}^{t\bar{t}}$			
$t\bar{t} + \gamma$ Visible Photon Selection $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} \cdot A_{\gamma}^{t\bar{t}+\gamma Vis}$			
Total			

Table 7.1: Values used for calculating the cross section ratios in the $\mu^+ \mu^-$, $e^+ e^-$, and $e\mu$ channels.

We calculate the cross section ratios for each of the di-lepton channels to be:

$$R_{\mu^+ \mu^-} = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma} A_{top}^{t\bar{t}+\gamma} \epsilon_{\gamma}^{t\bar{t}+\gamma} A_{\gamma}^{t\bar{t}+\gamma}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} = \quad (7.1)$$

$$R_{e^+ e^-} = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma} A_{top}^{t\bar{t}+\gamma} \epsilon_{\gamma}^{t\bar{t}+\gamma} A_{\gamma}^{t\bar{t}+\gamma}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} = \quad (7.2)$$

$$R_{e\mu} = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma} A_{top}^{t\bar{t}+\gamma} \epsilon_{\gamma}^{t\bar{t}+\gamma} A_{\gamma}^{t\bar{t}+\gamma}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} = \quad (7.3)$$

Another result is visible cross section ratio where we do not extrapolate the measured result to the phase space used for signal simulation. In this way we do not rely on kinematic properties of simulated signal dataset. We will need photon reconstruction and identification

Parameter	Value
$t\bar{t} + \gamma events$	
$t\bar{t} + \gamma eff.$	
$t\bar{t} events$	
$t\bar{t} eff$	

Table 7.2:

efficiency, but not signal acceptance for photon selection. Efficiency is calculated as ratio of number of generated signal events that passed photon selection to number of generated signal events with a generator level signal photon in the region of $p_T - \eta$ space used for photon selection (the same p_T and η cuts are applied to generated photons as used for the reconstructed photon). The visible cross section ratio is measured in each respective di-lepton channel to be:

$$R_{\mu^+\mu^-}^{Vis.} = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma Vis.} A_{top}^{t\bar{t}+\gamma Vis.} \epsilon_\gamma^{t\bar{t}+\gamma Vis.} A_\gamma^{t\bar{t}+\gamma Vis.}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} = \quad (7.4)$$

$$R_{e^+e^-}^{Vis.} = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma Vis.} A_{top}^{t\bar{t}+\gamma Vis.} \epsilon_\gamma^{t\bar{t}+\gamma Vis.} A_\gamma^{t\bar{t}+\gamma Vis.}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} = \quad (7.5)$$

$$R_{e\mu}^{Vis.} = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma Vis.} A_{top}^{t\bar{t}+\gamma Vis.} \epsilon_\gamma^{t\bar{t}+\gamma Vis.} A_\gamma^{t\bar{t}+\gamma Vis.}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} = \quad (7.6)$$

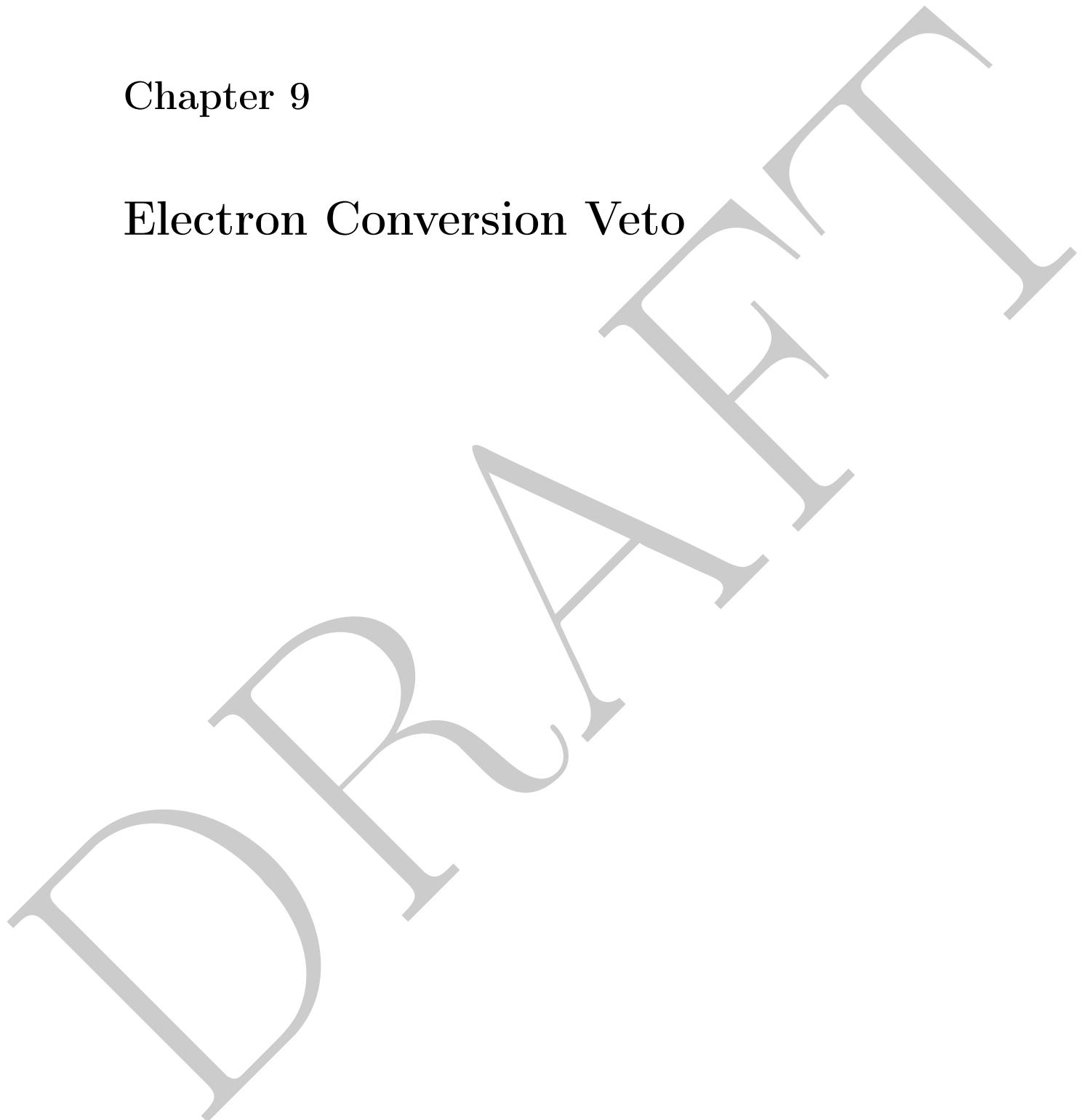
Chapter 8

Measurement of the anomalous couplings of the photon to the top quark

[67] [79]

Chapter 9

Electron Conversion Veto



Conclusions

DRAFT

Appendices



All ID and Isolation (MVA > 0.5)	$20 < p_T < 30$	$30 < p_T < 40$	$40 < p_T < 50$	$50 < p_T$
$ \eta < 0.8$	0.969 ± 0.007	0.926 ± 0.003	0.969 ± 0.002	0.975 ± 0.000
$0.8 < \eta < 1.4442$	0.935 ± 0.017	0.945 ± 0.004	0.964 ± 0.002	0.974 ± 0.002
$1.4442 < \eta < 1.5660$	1.032 ± 0.039	0.907 ± 0.015	0.957 ± 0.022	0.877 ± 0.020
$1.5660 < \eta < 2.5$	0.919 ± 0.014	0.926 ± 0.005	0.952 ± 0.003	0.950 ± 0.005

Table A.1: Electron ID & isolation efficiencies [15].

All ID and Isolation	$p_T > 20 \text{ GeV}/c$
$0 < \eta < 0.9$	0.9984 ± 0.0002
$0.9 < \eta < 1.2$	0.9990 ± 0.0002
$1.2 < \eta < 2.1$	0.9986 ± 0.0001
$2.1 < \eta < 2.5$	1.0000 ± 0.0003

Table A.1: Muon ID & isolation efficiencies [16].

A.1 Electron & Muon Efficiencies

Bibliography

- [1] Alemany-Fernandez et al. Operation and configuration of the lhc in run 1. *CERN Internal Note*, 25 Nov 2013.
- [2] CMS Collaboration. The cms experiment at the cern lhc. *JINST*, 3:S08004, 2008.
- [3] CMS Collaboration. Tracking results. (2012).
- [4] Cristina Biino. The cms electromagnetic calorimeter: overview, lessons learned during run 1 and future projections. *Journal of Physics: Conference Series*, 587(1):012001, 2015.
- [5] Stefan HÄuche. Introduction to monte carlo event generators. *Slides from 2012 CTEQ-Fermilab School on QCD and Electroweak Phenomenology*, 2012.
- [6] Andrea Giammanco. The fast simulation of the cms experiment. *Journal of Physics: Conference Series*, 513(2):022012, 2014.
- [7] H. J. A. Tholen and A. Stah. Study of the inclusive $t\bar{t} + \gamma$ cross-section with the cms experiment. <https://cds.cern.ch/record/1505276>, pages no. CERN–THESIS–2012–222, 2012.
- [8] Philip Symonds. Top pair production cross sections and differential cross sections in the semi-leptonic channel using the cms detector at $\sqrt{s} = 7$ and 8 tev. *PhD Thesis*, March 2015.
- [9] Thomas Steeger. Background estimation in a search for $t\bar{t} + \gamma$ events. *Bachelor Thesis*, August, 2012.
- [10] Thomas Hermanns. Studies at the cms experiment on silicon microstrip module defects and on photon identification in semileptonic $t\bar{T}$ events. *PhD Thesis*, CERN-THESIS-2008-111, 12/008.

Bibliography

- [11] Oliver Sim et al. BrAijning. Lhc design report. *Geneva : CERN, v.1 : the LHC Main Ring*:548, 2004.
- [12] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.
- [13] CMS Collaboration. Precise mapping of the magnetic field in the cms barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021, 2010.
- [14] Sergey Senkin. Top quark physics at the lhc with the cms detector. *PhD Thesis*, page 204, 04/2014.
- [15] CMS Collaboration. Electron efficiencies for top quark analysis. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/KoPFAElectronTagAndProbe>.
- [16] CMS Collaboration. Reference muon id and isolation efficiencies. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs>.
- [17] John P. Blewett. 200-gev intersecting storage accelerators. *Proceedings of The 8th International Conference on High-Energy Accelerators. CERN. Geneva, Switzerland*, 1971.
- [18] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Phys. Lett. B*, 716, 2012.
- [19] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716, 2012.
- [20] L. Borrello et al. Sensor design for the cms silicon strip tracker. *CMS-NOTE-2003-020*.
- [21] J.L. Agram et al. The silicon sensors for the compact muon solenoid tracker: design and qualification procedure. *Nucl. Instrum. Meth.*, A 517:77, (2004).
- [22] V. Veszpremi. Operation and performance of the cms tracker. *Journal of Instrumentation*, 9(03):C03005, 2014.
- [23] L. Rossi et al. Pixel detectors. *Springer-Verlag, Heidelberg Germany*, (2006).
- [24] CMS Collaboration. Letter of intent. *CERN-LHCC-92-003*, 1992.
- [25] CMS Collaboration. The e.m. calorimeter project: Technical design report. *CERN-LHCC-97-33*, 1997.

- [26] J et al. Mans. CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter. Technical Report CERN-LHCC-2012-015. CMS-TDR-10, CERN, Geneva, Sep 2012.
- [27] CMS Collaboration. CMS technical design report, volume II: Physics performance. *J. Phys. G*, 34:995–1579, 2007.
- [28] 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.
- [29] Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, 2010.
- [30] Rene Brun and Fons Rademakers. {ROOT} – an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1–2):81 – 86, 1997. New Computing Techniques in Physics Research V.
- [31] CMS Collaboration. Cmssw software framework.
- [32] GEANT4 Collaboration. Geant4: A simulation toolkit. *Nucl.Instrum.Meth.*, A506:250, 2003.
- [33] et al. D. Acosta, CMS Collaboration. Cms physics: Technical design report, volume 1: Detector performance and software. *CERN/LHCC, Geneva, Switzerland*, 2, 2006.
- [34] T. Ohl W. Kilian and J. Reuter. Whizard: Simulating multi-particle processes at lhc and ilc. *Eur.Phys.J.*, C71:1742, 2011.
- [35] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), 2011.
- [36] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, Andreas Schalicke, and Steffen Schumann. Matching parton showers and matrix elements. In *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*, 2006.
- [37] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

- [38] Stefano Frixione and Bryan R. Webber. Matching nlo qcd computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029, 2002.
- [39] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, 2007(11):070, 2007.
- [40] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.
- [41] M. R. Whalley, D. Bourilkov, and R. C. Group. The Les Houches accord PDFs (LHAPDF) and LHAGLUE. In *HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B*, 2005.
- [42] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [43] Z. Was. Precision simulations with TAUOLA and PHOTOS. *Nucl. Phys. Proc. Suppl.*, 169:16–21, 2007. [,16(2006)].
- [44] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009, 2014.
- [45] Pierre Billoir. Progressive track recognition with a kalman-like fitting procedure. *Computer Physics Communications*, 57(1–3):390 – 394, 1989.
- [46] W Erdmann. Vertex reconstruction at the cms experiment. *Journal of Physics: Conference Series*, 110(9):092009, 2008.
- [47] W Adam, R FrÄijhwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005.
- [48] Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV. Technical Report CMS-PAS-PFT-10-003, CERN, 2010. Geneva, 2010.
- [49] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 543(2–3):577 – 584, 2005.

- [50] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron reconstruction in CMS. *Eur. Phys. J.*, C49:1099–1116, 2007.
- [51] The CMS collaboration. Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$ tev. *Journal of Instrumentation*, 7(10):P10002, 2012.
- [52] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.
- [53] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.
- [54] Gerald C. Blazey et al. Run II jet physics. In *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pages 47–77, 2000.
- [55] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [56] Ricardo Eusebi and on behalf of the CMS collaboration). Jet energy corrections and uncertainties in cms: reducing their impact on physics measurements. *Journal of Physics: Conference Series*, 404(1):012014, 2012.
- [57] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011.
- [58] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
- [59] Christian Weiser. A combined secondary vertex based b-tagging algorithm in cms. *Technical Report*, CMS-NOTE-2006-014, CERN, Geneva, Jan 2006.
- [60] R Frhwirth Wolfgang Waltenberger Thomas Speer, Kirill Prokofiev and Pascal Vanlaer. Vertex fitting in the cms tracker. *Technical Report*, CMS-NOTE-2006-032, CERN, Geneva, Feb 2006.
- [61] The CMS collaboration. Missing transverse energy performance of the cms detector. *Journal of Instrumentation*, 6(09):P09001, 2011.
- [62] CMS Collaboration. Top reference selections and recommendations (run1) twiki. *CMS TWiki*.

- [63] CMS Collaboration. Cms top egamma coordination (run1). *CMS TWiki*.
- [64] CMS Collaboration. Top muon information for analysis (run1). *CMS TWiki*.
- [65] CMS Collaboration. Jet performance in pp collisions at $\sqrt{s}=7$ TeV. CMS Physics Analysis Summary CMS-PAS-JME-10-003, 2010.
- [66] CMS Collaboration. Simple cut based id 2012. *CMS Twiki*.
- [67] Kirill Melnikov, Markus Schulze, and Andreas Scharf. QCD corrections to top quark pair production in association with a photon at hadron colliders. *Phys. Rev.*, D83:074013, 2011.
- [68] Simon White. Luminosity scans at the lhc. luminosity scans at lhc. 2011.
- [69] Cms luminosity based on pixel cluster counting - summer 2013 update. Technical Report CMS-PAS-LUM-13-00, CERN, Geneva, 2013.
- [70] CMS Collaboration. Absolute calibration of the luminosity measurement at cms: Winter 2012 update. 2012.
- [71] Simon White. Luminosity Scans at the LHC. Luminosity Scans at LHC. 2011.
- [72] CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.
- [73] A. Vogt, S. Moch, and J. A. M. Vermaseren. The Three-loop splitting functions in QCD: The Singlet case. *Nucl. Phys.*, B691:129–181, 2004.
- [74] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C.-P. Yuan. Implications of cteq global analysis for collider observables. *Phys. Rev. D*, 78:013004, Jul 2008.
- [75] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C.-P. Yuan. Ct10 next-to-next-to-leading order global analysis of qcd. *Phys. Rev. D*, 89:033009, Feb 2014.
- [76] 8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS. Oct 2013.
- [77] Jet Energy Resolution in CMS at $\text{sqrt}(s)=7$ TeV. 2011.

- [78] Performance of b tagging at $\sqrt{s} = 8$ tev in multijet, $t\bar{t}$ and boosted topology events. Oct 2013.
- [79] Valentin Hirschi, Rikkert Frederix, Stefano Frixione, Maria Vittoria Garzelli, Fabio Maltoni, and Roberto Pittau. Automation of one-loop QCD corrections. *JHEP*, 05:044, 2011.

DRAFT