

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*

January 28, 2016

### **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

<b>Introduction</b>	1
<b>1 Theoretical Motivation</b>	2
1.1 The Standard Model of Particle Physics . . . . .	2
1.1.1 Gauge theory . . . . .	2
1.1.2 The Electroweak Theory . . . . .	4
1.1.3 Quantum Chromodynamics . . . . .	4
1.1.4 Electroweak Symmetry Breaking . . . . .	4
1.2 The Top Quark . . . . .	6
<b>2 The LHC and the CMS Detector</b>	10
2.1 The Large Hadron Collider . . . . .	10
2.1.1 Pre-LHC accelerator complex . . . . .	10
2.1.2 Design of the LHC . . . . .	12
2.1.3 Physics goals . . . . .	12
2.1.4 Luminosity at the LHC . . . . .	13
2.1.5 Performance throughout run I . . . . .	14
2.2 The CMS Detector . . . . .	15
2.3 Inner Tracking System . . . . .	17
2.3.1 Tracker performance in Run I . . . . .	19
2.4 Electromagnetic Calorimeter . . . . .	21
2.4.1 Overview . . . . .	21
2.4.2 Composition of the ECAL . . . . .	22
2.4.3 Photodetectors . . . . .	24
2.4.4 Performance of the ECAL throughout Run I . . . . .	24
2.5 Hadron Calorimeter . . . . .	25

2.5.1	Overview	25
2.5.2	Performance of the HCAL in Run I	27
2.6	Superconducting Solenoid	27
2.7	Muon System	28
2.7.1	Performance of the muon system in Run I	31
2.8	Trigger and Data Acquisition	33
2.8.1	Performance of the Trigger in Run 1	34
<b>3</b>	<b>Event Reconstruction &amp; Simulation</b>	<b>37</b>
3.1	Event Reconstruction	39
3.2	Computing	39
3.2.1	Event Data Model	39
3.2.2	Analysis Software	40
3.3	Simulation of the CMS detector	40
3.4	Monte Carlo Simulation	41
3.4.1	Monte Carlo event generators	41
3.5	Simulated samples for the $t\bar{t} + \gamma$ analysis	43
3.6	Simulation of the $t\bar{t} + \gamma$ Signal Sample	44
3.6.1	Official $t\bar{t} + \gamma$ 2 → 7 process MADGRAPH Sample Production	47
3.7	Physics Object Reconstruction	50
3.7.1	Charged Particle Tracking	52
3.7.2	Primary Vertex Reconstruction	53
3.7.3	Calorimeter Clustering Algorithm	54
3.7.4	The Particle Flow Algorithm	55
3.7.5	Electron reconstruction	56
3.7.6	Electron identification	59
3.7.7	Muon reconstruction	62
3.8	Higher-level Object Reconstruction	64
3.8.1	Jet reconstruction	64
3.8.2	Jet Identification	66
3.8.3	Jet energy corrections	66
3.8.4	B-tagged Jets	67
3.8.5	Reconstruction of Missing Transverse Energy	69

<b>4 Event Selection &amp; Background Estimation</b>	<b>71</b>
4.1 Event Selection . . . . .	71
4.2 Pre-selection: Selection of $t\bar{t}$ Events . . . . .	72
4.3 Skim . . . . .	73
4.4 Trigger and Event Cleaning . . . . .	73
4.4.1 Trigger selection . . . . .	73
4.4.2 Filtering . . . . .	75
4.5 Di-lepton Selection and Vetoies . . . . .	76
4.5.1 Electrons . . . . .	76
4.5.2 Muons . . . . .	78
4.5.3 Corrections to simulated events . . . . .	79
4.6 Jet Selection and b-tag Requirements . . . . .	79
4.7 Missing Transverse Energy . . . . .	80
4.8 Comparison of Data to MC . . . . .	80
4.9 Selection of $t\bar{t} + \gamma$ events . . . . .	80
4.9.1 Cut based photon ID . . . . .	82
4.9.2 FSR suppression . . . . .	83
4.9.3 Super-cluster footprint-removal for photon isolation . . . . .	84
4.9.4 Signal template construction using the random cone method . . . . .	85
4.10 Background Estimation . . . . .	87
4.10.1 Background estimation using the $\sigma_{inj}$ side-band method . . . . .	87
<b>5 Measurement of the inclusive <math>t\bar{t} + \gamma</math> cross-section</b>	<b>95</b>
5.1 Signal Definition and Background Processes . . . . .	95
5.1.1 Signal definition . . . . .	95
5.1.2 Background processes . . . . .	95
5.1.3 Backgrounds of photon signature . . . . .	95
5.2 $t\bar{t} + \gamma$ Signal Simulation . . . . .	95
5.3 Phase Space Overlap Removal . . . . .	95
5.4 Measurement of the cross-section ratio $R = \sigma_{t\bar{t}+\gamma}/\sigma_{t\bar{t}}$ . . . . .	100
5.5 Number of top pair events . . . . .	102
5.6 Photon Purity Estimation . . . . .	102
5.7 Number of signal events in data . . . . .	104
5.8 Signal Acceptance Calculation . . . . .	104

<b>6 Systematic Uncertainties</b>	<b>107</b>
6.1 Flat Rate Uncertainties . . . . .	108
6.1.1 Luminosity . . . . .	108
6.1.2 Lepton Efficiencies . . . . .	109
6.2 Shape Uncertainties . . . . .	110
6.2.1 Parton Distribution Function . . . . .	110
6.2.2 Pile-up Re-weighting . . . . .	111
6.2.3 Jet Energy Corrections . . . . .	111
6.2.4 Missing Transverse Energy . . . . .	112
6.2.5 B-tagging Efficiency . . . . .	113
6.2.6 Data-driven Re-weighting . . . . .	113
6.3 Modelling Uncertainties . . . . .	113
6.3.1 QCD Renormalisation and factorisation scales . . . . .	113
6.3.2 Parton-level matching thresholds . . . . .	113
6.3.3 Analysis-dependent modelling uncertainties . . . . .	113
6.4 Impact of uncertainties . . . . .	113
<b>7 Results</b>	<b>116</b>
<b>8 Measurement of the anomalous couplings of the photon to the top quark</b>	<b>118</b>
<b>9 Electron Conversion Veto</b>	<b>119</b>
<b>Conclusions</b>	<b>120</b>
<b>Appendices</b>	<b>122</b>
A.1 Electron & Muon Efficiencies . . . . .	122
<b>Bibliography</b>	<b>131</b>

# List of Figures

1.1	The “Mexican Hat Potential” describing the vacuum expectation value of the Higgs in the real and imaginary planes, such that the minima lies below zero.	5
1.2	Top-photon vertex. Left: Leading Order (LO). Right: One-Loop correction (NLO).	6
1.3	The Standard Model of particle physics.	7
1.4	Lowest level diagrams for $t\bar{t}$ production at the LHC. Gluon scattering processes, (a), (b), and (c), are the dominant processes at LHC energies, while quark scattering, process (d), is the dominant one at TeVatron energies. [1]	7
1.5	Leading-order level diagrams for single top production at the LHC. (a) s-channel, (b) and (c) represent the t-channel, (d) and (e) both represent the two tW channels. [1]	8
1.6	Branching fractions of the W decays within top quark pairs. [2]	8
1.7	.....	9
2.1	A full schematic of the full CERN accelerator complex.	11
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 [3].	15
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.	16
2.4	A cross-sectional view of the CMS detector.	16

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. [4]. . . . .	18
2.6 Muon reconstruction efficiency in thee tracker as functions of pseudorapidity (left) and the number of proton-proton interaction vertices (right) [5]. . . . .	20
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 [5]. . . . .	20
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) [4]. . . . .	22
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) [4]. . . . .	23
2.10 Left: The mass resolution of the Z peak, reconstructed from its di-electron decay mode, as a function of time for the barrel. Right: Relative electron (ECAL) energy resolution unfolded in bins of pseudo-rapidity $\eta$ for the barrel. [6]. . . . .	25
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) [4]. . . . .	26
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) [4]. . . . .	29
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) [4]. . . . .	30
2.14 RPCMuon efficiencies on top of the semi-Tight muon selection in $p_T$ and $ \eta $ . [7]	32
2.15 Architecture of the Level 1 trigger [4]. . . . .	34
2.16 Top Triggers: Efficiency of the hadronic leg for the electron plus jets paths in 2012 versus the $p_T$ of the 4th jet (left) and the dependence with respect to the number of vertices (right). [8] . . . . .	36
3.1 Schematic overview of a hadron collision process [9]. . . . .	38

3.2 A radiography of a quarter of the simulated tracker geometry in the (a) fast and (b) full simulation [10]. . . . .	41
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 [11]. . . . .	46
3.4 Comparison between the full and the factorised $2 \rightarrow 7$ process using samples generated with WHIZARD. All distributions are normalised to unity. [12] . . . . .	49
3.5 Comparison between the MADGRAPH and WHIZARD generators for the factorized $2 \rightarrow 7$ process. All distributions are normalized to unity. [12] . . . . .	51
3.6 The $(x, y)$ view [13]. . . . .	57
3.7 . . . . .	58
3.8 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. . . . .	68
3.9 Logarithmic distribution of the partons as a function of the CSV b-tagging discriminant [14]. . . . .	69
4.1 . . . . .	82
4.2 Graphic representation of the PF candidate footprint (in red) from the primary interaction vertex, within the isolation cone (green). [15] . . . . .	86
4.3 Illustration of the random cone technique. Under the assumption that the footprint of the photon (red) is completely removed from the isolation sum, the energy deposited in the random-cone area (green in right image) predicts the energy deposited in the isolation cone (green area in left image). [15] . . . . .	87
4.4 . . . . .	89
4.5 . . . . .	90
5.1 Graphic representation of the signal and background definitions [?]. . . . .	96
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 [16]. . . . .	97
5.3 Flow chart showing each stage of the analysis. The box numbers represent the outlined analysis steps. . . . .	98
5.4 Background of photon identification: Initial state radiation (ISR). Left: Quark fusion. Right: Gluon fusion does not give rise to ISR (based on [17]). . . . .	99

5.5	Background of photon identification: Final state radiation (FSR). All charged particles in the top-pair decay tree contribute to FSR [17]. . . . .	99
5.6	. . . . .	100



# List of Tables

1.1	[18] . . . . .	6
2.1	LHC design parameters [19]. . . . .	13
2.2	Tower segmentation in azimuthal and polar angle for the hadronic barrel, end-cap and forward calorimeter [20]. . . . .	27
2.3	Parameters of the LHC superconducting solenoid [21]. . . . .	28
3.1	Dataset information for signal and background MC samples. . . . .	45
3.2	Variables used in electron identification algorithms [1]. . . . .	62
4.1	Triggers for each di-lepton channel. . . . .	74
4.2	The number of expected events in MC and events observed in data for the $\mu^+\mu^-$ channel, before the fitting process, including statistical uncertainties. . . . .	91
4.3	The number of expected events in MC and events observed in data for the $e^+e^-$ channel, before the fitting process, including statistical uncertainties. . . . .	92
4.4	The number of expected events in MC and events observed in data for the $e\mu$ channel, before the fitting process, including statistical uncertainties. . . . .	93
4.5	Dataset information for each run in each respective decay channel. . . . .	94
6.1	Systematic uncertainties and their contribution to the cross-section ratio. . . . .	114
6.2	. . . . .	115
7.1	Values used for calculating the cross section ratios in the $\mu^+\mu^-$ , $e^+e^-$ , and $e\mu$ channels. . . . .	116
7.2	. . . . .	117
A.1	Electron ID & isolation efficiencies [22]. . . . .	122
A.1	Muon ID & isolation efficiencies [23]. . . . .	122

# 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?

hard-scattering pp collisions

$$\sigma(pp \rightarrow X) = \int_0^1 dx_1 \int_0^1 dx_2 \sum_{a,b} f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) \hat{\sigma}_{ab \rightarrow X}(Q^2, \mu_F^2) \quad (1.1)$$

### 1.1 The Standard Model of Particle Physics

#### 1.1.1 Gauge theory

Almost all of the physics within Standard Model arises directly from imposed symmetries. Interactions are produced by requiring local gauge invariance under specific symmetry groups. We define the Standard Model in group theory as the unification of two gauge symmetry groups, describing the electroweak and strong interactions. The Glashow-Weinberg-Salam electroweak model sees the electromagnetic and weak interactions combined to create the electroweak gauge symmetry group  $SU(2)_L \otimes U(1)_Y$ , where the gauge symmetry group for strong interactions is defined as  $SU(3)_C$ . Thus, we can define the gauge symmetry group of the Standard Model to be

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \quad (1.2)$$

In order to explain the concept of gauge invariance we start by considering changes to the phase of the wavefunctions or fields. Let us take the Lagrangian density of a free Dirac

field,  $\psi$ , describing all free-moving spin- $\frac{1}{2}$  fermions with a mass  $m$ :

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.3)$$

where  $\gamma^\mu$  are the Dirac matrices. We can show that this is invariant under phase rotations, defined by

$$\psi \rightarrow \psi' = e^{i\theta}\psi, \quad \bar{\psi} \rightarrow \bar{\psi}' = e^{-i\theta}\bar{\psi}, \quad (1.4)$$

as the exponential factors will cancel each out, thus we have a  $U(1)$  symmetry and corresponding conserved current. This is what is known as a global phase transformation due to the time dependency of  $\theta$ . However, we require a local gauge invariance, and therefore a local phase transformation must be applied such that  $\theta$  is different at every space-time point, which we now define as  $\theta(x_\mu)$ . The local phase transformations are applied to the wave function,  $\psi$ , and are defined in the following way:

$$\psi \rightarrow \psi' = e^{i\theta(x_\mu)}\psi, \quad \bar{\psi} \rightarrow \bar{\psi}' = e^{-i\theta(x_\mu)}\bar{\psi}, \quad (1.5)$$

Dropping the space-time component  $\mu$  from  $x_\mu$  for simplicity, we can see that the Dirac equation is now not invariant under the local gauge transformation as

$$\partial_\mu(e^{i\theta(x)}\psi) = i(\partial_\mu\theta(x))e^{i\theta(x)}\psi + e^{i\theta(x)}\partial_\mu\psi \quad (1.6)$$

which we can then express in terms of the Lagrangian as

$$\mathcal{L} \rightarrow \mathcal{L} - [\partial_\mu\theta(x)]\bar{\psi}\gamma^\mu\psi. \quad (1.7)$$

Yang-Mills theory of  $SU(2)$  later extended to  $SU(3)$  [24]  
covariant derivative,  $D_\mu$

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu \quad (1.8)$$

$$A_\mu \rightarrow A_\mu - \partial_\mu\theta(x) \quad (1.9)$$

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \quad (1.10)$$

$$\mathcal{L}_{QED} = \bar{\psi}(iD_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.11)$$

where  $\not{D}_\mu$  is equal to  $\gamma^\mu(\partial_\mu + iqA_\mu)$

### 1.1.2 The Electroweak Theory

Sheldon Glashow 1961 [25] Steven Weinberg 1967 [26] Abdus Salam 1951 [27] Gell-Mann 1964 [?]

violates parity and thus distinguishes between right- and left-handed [28]

$$q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad l_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \quad (1.12)$$

$$\mathcal{L}_{EWK} = \bar{\psi}_L \gamma^\mu \left( i\beta \partial_\mu - g \mathbf{T} \cdot \mathbf{W}_\mu - \frac{g'}{2} Y B_\mu \right) \psi_L + \bar{\psi}_R \gamma^\mu \left( i\beta \partial_\mu - \frac{g'}{2} Y B_\mu \right) \psi_R - \frac{1}{4} \mathbf{W}_{\mu\nu} \mathbf{W}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \quad (1.13)$$

### 1.1.3 Quantum Chromodynamics

asymptotic freedom David J. Gross and Frank Wilczek [29]

H. David Politzer [30]

$$r = \begin{pmatrix} r \\ b \\ g \end{pmatrix} \quad (1.14)$$

$$\mathcal{L}_{QCD} = \bar{q} (\gamma^\mu \partial_\mu - m) q + g_s (\bar{q} \gamma^\mu T_a q) G_\mu^a - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.15)$$

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.16)$$

$$V_{CKM} = \begin{pmatrix} 0.97425 \pm 0.00022 & 0.2253 \pm 0.0008 & 0.00413 \pm 0.00049 \\ 0.225 \pm 0.008 & 0.986 \pm 0.016 & 0.0411 \pm 0.0013 \\ 0.0084 \pm 0.0006 & 0.040 \pm 0.0027 & 1.021 \pm 0.032 \end{pmatrix} \quad (1.17)$$

### 1.1.4 Electroweak Symmetry Breaking

Peter Higgs [31]

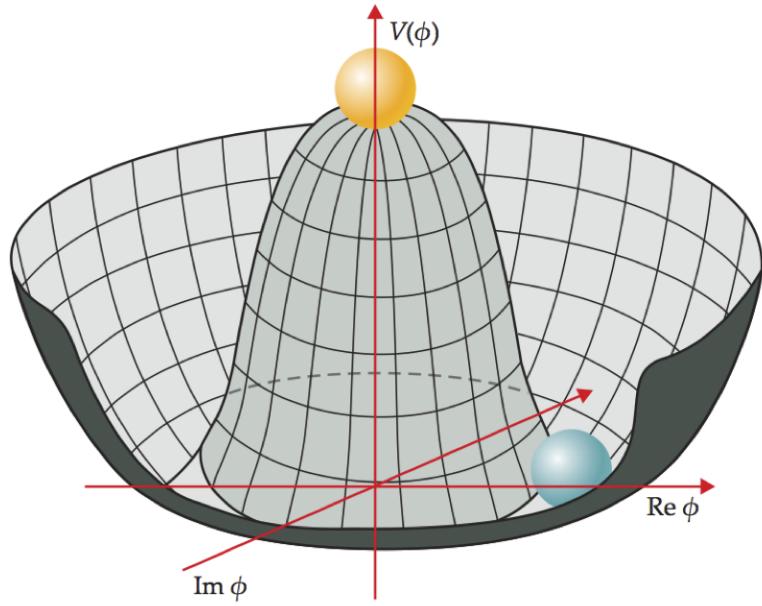


Figure 1.1: The “Mexican Hat Potential” describing the vacuum expectation value of the Higgs in the real and imaginary planes, such that the minima lies below zero.

Francois Englert and Robert Brout [32]

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.18)$$

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad (1.19)$$

$$V(\phi) = -\mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2 \quad (1.20)$$

$$\langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.21)$$

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \quad (1.22)$$

$$\begin{aligned} \mathcal{L}_{Higgs} = & \frac{1}{2} (\partial_\mu H) (\partial^\mu H) + \frac{1}{4} g^2 (H^2 + 2vH + v^2) W_\mu^+ W^{-\mu} \\ & + \frac{1}{8} (g^2 + g'^2) (H^2 + 2vH + v^2) Z_\mu Z^\mu - \mu^2 H^2 - \frac{\lambda}{4} (H^2 - v^2)^2 \end{aligned}$$

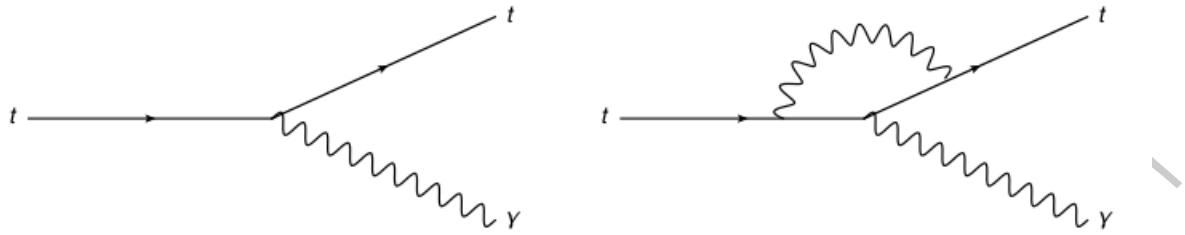


Figure 1.2: Top-photon vertex. Left: Leading Order (LO). Right: One-Loop correction (NLO).

$\sqrt{s}$	$\sigma_{tot.}$	scales [pb]	pdf [pb]
Tevatron 1.9 TeV	7.164		
LHC TeV	172.0	4.4.5.8	
LHC 8 TeV	245.8		
LHC 14 TeV	953.6		

Table 1.1: [18]

$$\mathcal{L}_{\gamma t \bar{t}} = -e Q_t \bar{t} \gamma^\mu t A_\mu - e \bar{t} \frac{i \sigma^{\mu\nu} q_\nu}{m_t} (d_V^\gamma + i d_A^\gamma \gamma^5) t A_\mu \quad (1.23)$$

$$\mathcal{L}^{eff.} = \frac{C_x}{\Lambda_x} O_x + \dots, \quad (1.24)$$

$$\delta d_V^\gamma = \frac{\sqrt{2}}{e} \text{Re}[c_W C_{uB\phi}^{33} + s_W C_{uW}^{33}] \frac{v m_t}{2} \quad (1.25)$$

$$\delta d_A^\gamma = \frac{\sqrt{2}}{e} \text{Im}[c_W C_{uB\phi}^{33} + s_W C_{uW}^{33}] \frac{v m_t}{2} \quad (1.26)$$

## 1.2 The Top Quark

mass → ≈2.3 MeV/c <sup>2</sup>	charge → 2/3	spin → 1/2	mass → 0	charge → 0
up	c	t	g	H
down	s	b	γ	
electron	μ	τ	Z	
electron neutrino	ν <sub>e</sub>	ν <sub>μ</sub>	W	
muon neutrino				
tau neutrino				

Figure 1.3: The Standard Model of particle physics.

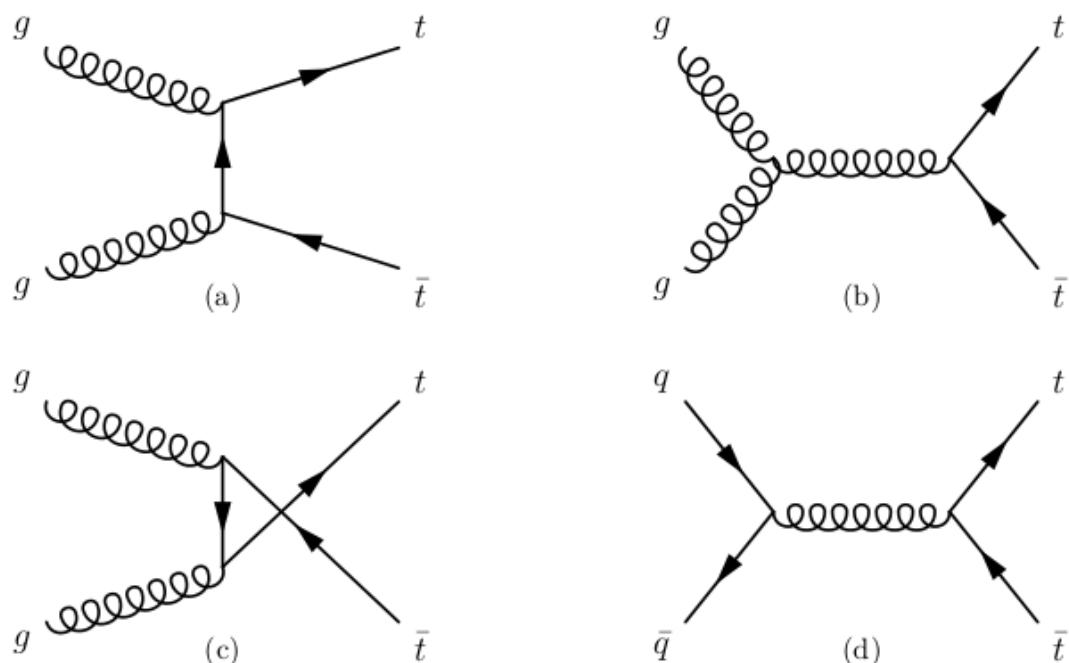


Figure 1.4: Lowest level diagrams for  $t\bar{t}$  production at the LHC. Gluon scattering processes, (a), (b), and (c), are the dominant processes at LHC energies, while quark scattering, process (d), is the dominant one at TeVatron energies. [1]

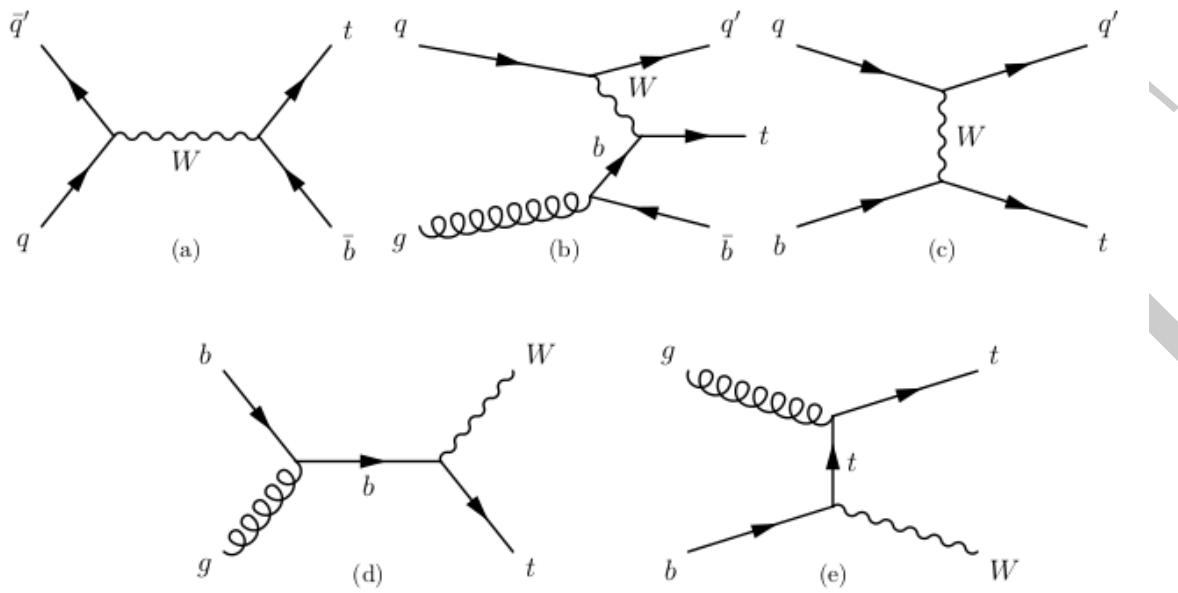


Figure 1.5: Leading-order level diagrams for single top production at the LHC. (a) s-channel, (b) and (c) represent the t-channel, (d) and (e) both represent the two tW channels. [1]

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

Figure 1.6: Branching fractions of the W decays within top quark pairs. [2]

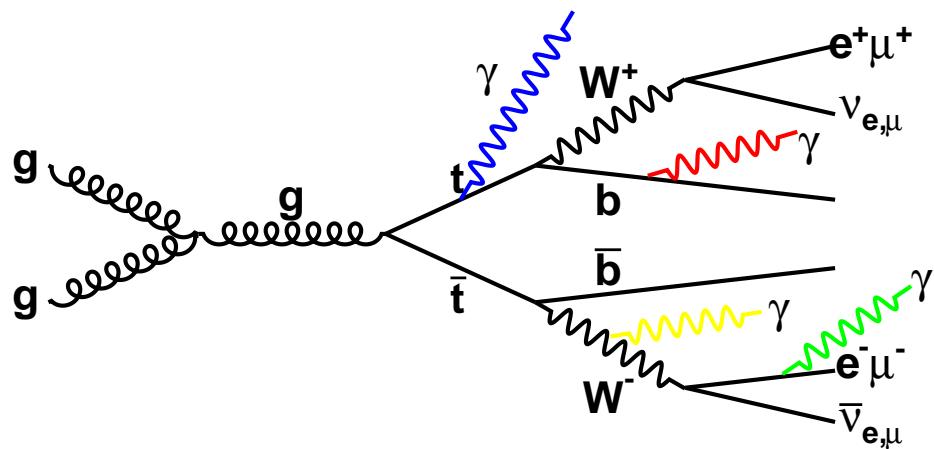


Figure 1.7

## 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 ??, 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

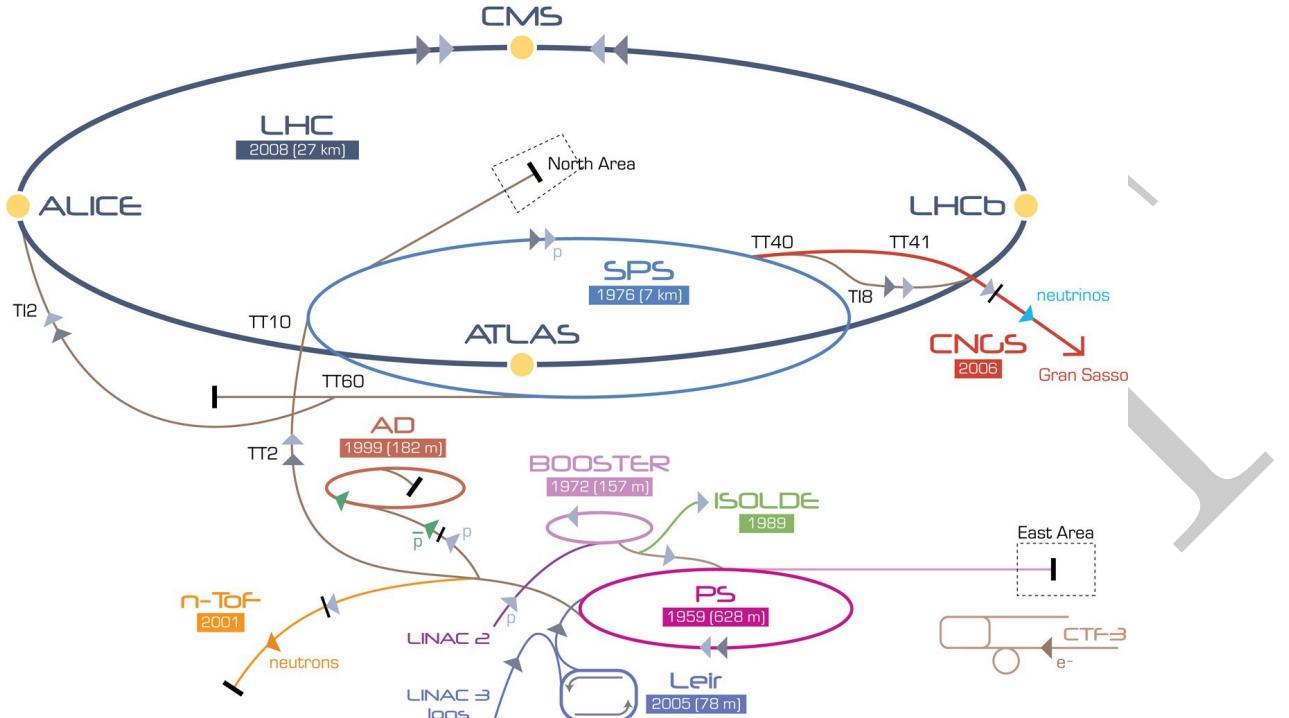


Figure 2.1: A full schematic of the full CERN accelerator complex.

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  $\sim 27$  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 [33] 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 ??.

### 2.1.3 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	$1.7 \times 10^{11}$	$1.15 \times 10^{11}$
Bunch spacing (ns)	50	25
Revolution frequency (kHz)	11.245	11.245
Transverse beam size ( $\mu\text{m}$ )	18.8	16.6
Peak luminosity ( $\text{cm}^{-2}\text{s}^{-1}$ )	$7.7 \times 10^{33}$	$10^{34}$
Stored beam energy (MJ)	140	362
Normalised emittance at start of fill (mm mrad)	2.5	3.75
$\beta^*$ in IP 1 and 5 (m)	0.6	0.55

Table 2.1: LHC design parameters [19].

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 ???. 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 [34] and CMS [35] 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  $\beta^*$  - the measure of how precisely the beam is focused at the interaction point. For ATLAS and CMS  $\beta^*$  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  $1.7 \times 10^{11}$  protons per bunch, which was then translated into a bunch intensity of  $1.6 \times 10^{11}$  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  $23.3 \text{ fb}^{-1}$  to ATLAS and CMS, and over  $2.1 \text{ fb}^{-1}$  to LHCb. The integrated and peak luminosity can be seen in Figure ??, 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 ??.

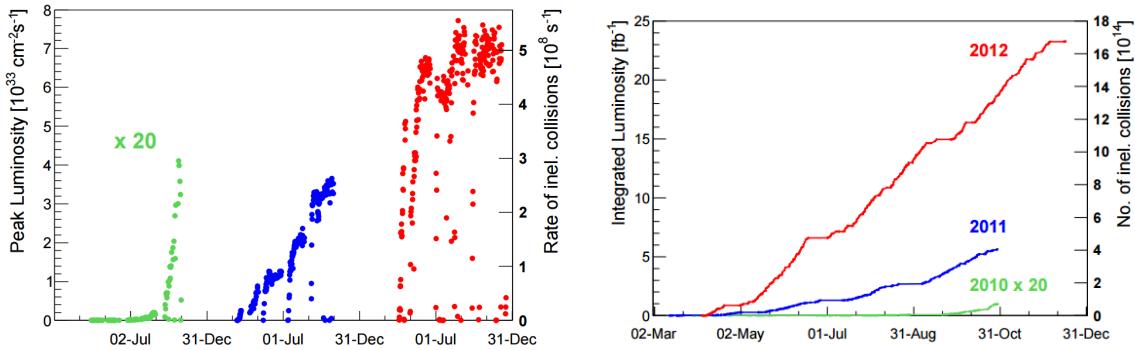


Figure 2.2: (Left) Peak and (right) integrated luminosity recorded by the LHC between 2010 and 2012 for proton operation. The 2010 luminosity values have been multiplied by a factor 20 [3].

## 2.2 The CMS Detector

The Compact Muon Solenoid (CMS) [4] 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 ???. 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,  $\phi$ , is measured from the x-axis in the xy plane where the radial component in this plane is

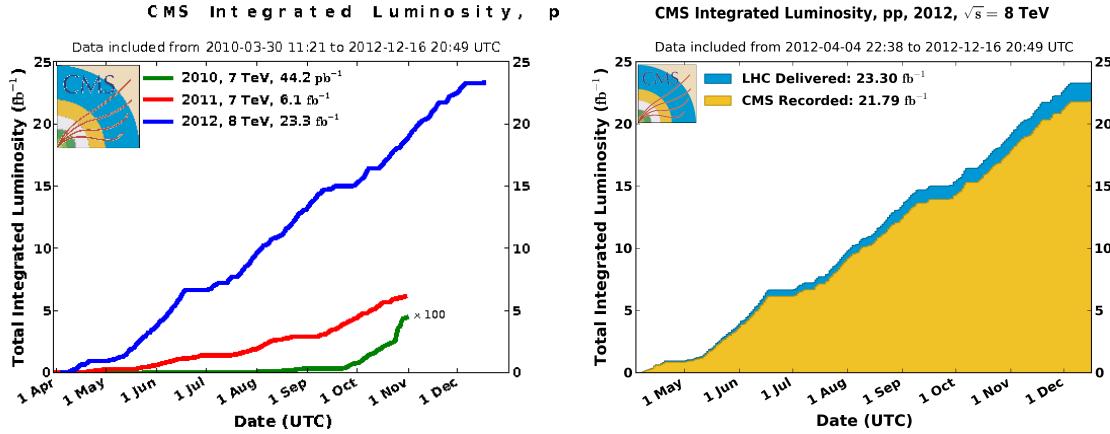


Figure 2.3: Left: The accumulation of the integrated luminosity produced at the LHC vs time for runs in 2010, 2011, and 2012. The 2010 integrated luminosity is multiplied by 100 in order for it to be visible on the plot. Right: Total integrated luminosity vs time for the 2012 run in CMS and the LHC.

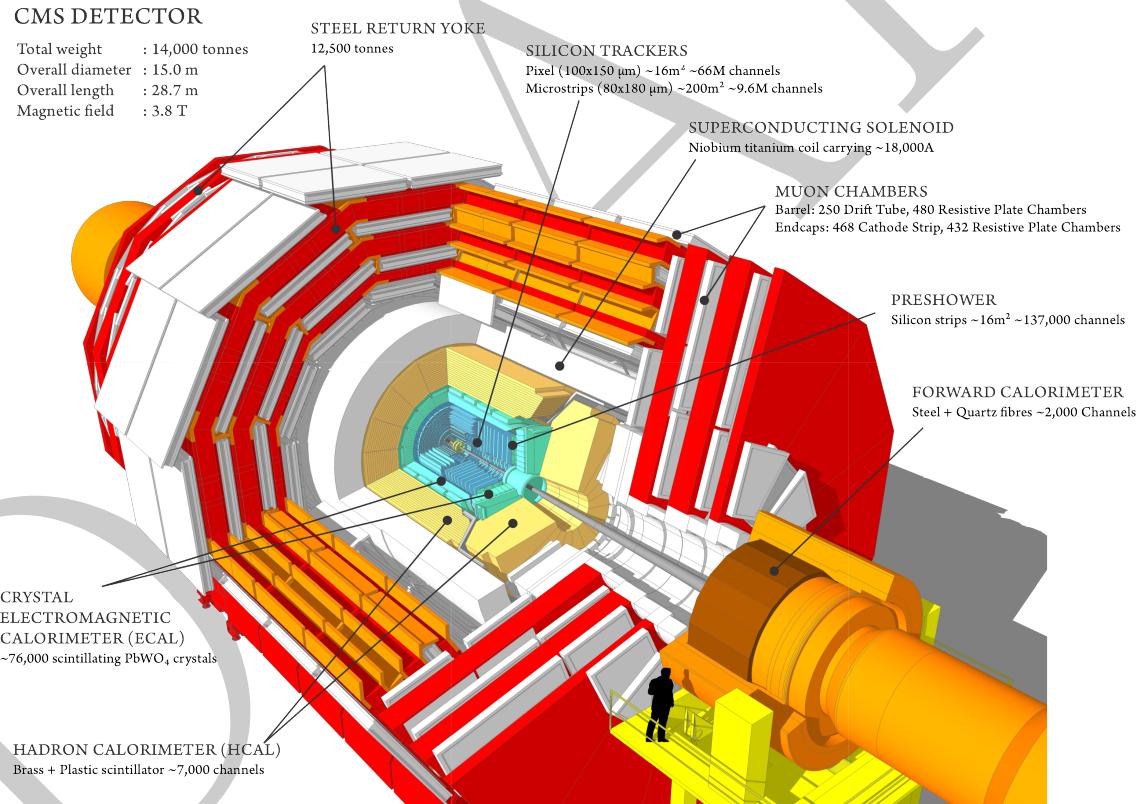


Figure 2.4: A cross-sectional view of the CMS detector.

define by  $r$ , and the polar angle  $\theta$  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 ??, 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 in 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  $1 \text{ m}^2$ .

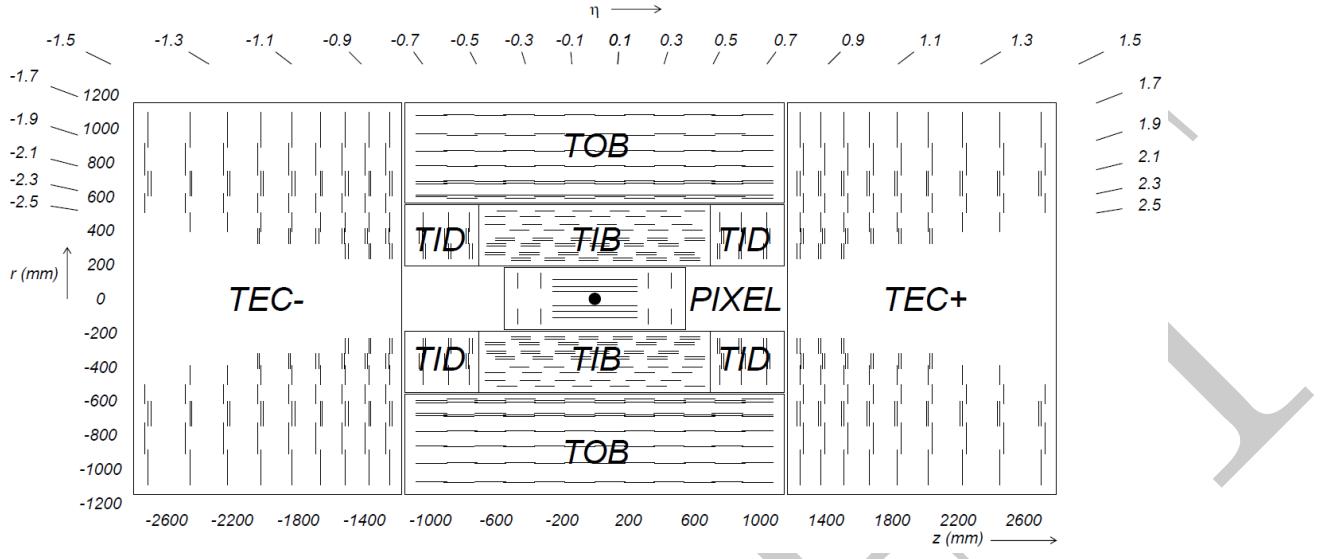


Figure 2.5: The sub-detectors of the CMS silicon tracker system: TOB=outer barrel, TIB=inner barrel, TID=inner disc, TEC=endcaps, PIXEL=pixeldetector. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits. [4].

The sensor elements in the silicon strip tracker system are single sided p-on-n type silicon micro-strip sensors [36, 37]. 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  $\phi$  measurements per trajectory.

### 2.3.1 Tracker performance in Run I

Over the Run I period, from 2010 to 2013, the LHC delivered around  $6 \text{ fb}^{-1}$  at 7 TeV and  $23.3 \text{ fb}^{-1}$  at 8 TeV (Figure ??), 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 shut-down 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 [38].

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 ??.

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 [5]. 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 ?? . LHC proton-proton events are reconstructed by firstly identifying the tracks, then grouping in accordance with

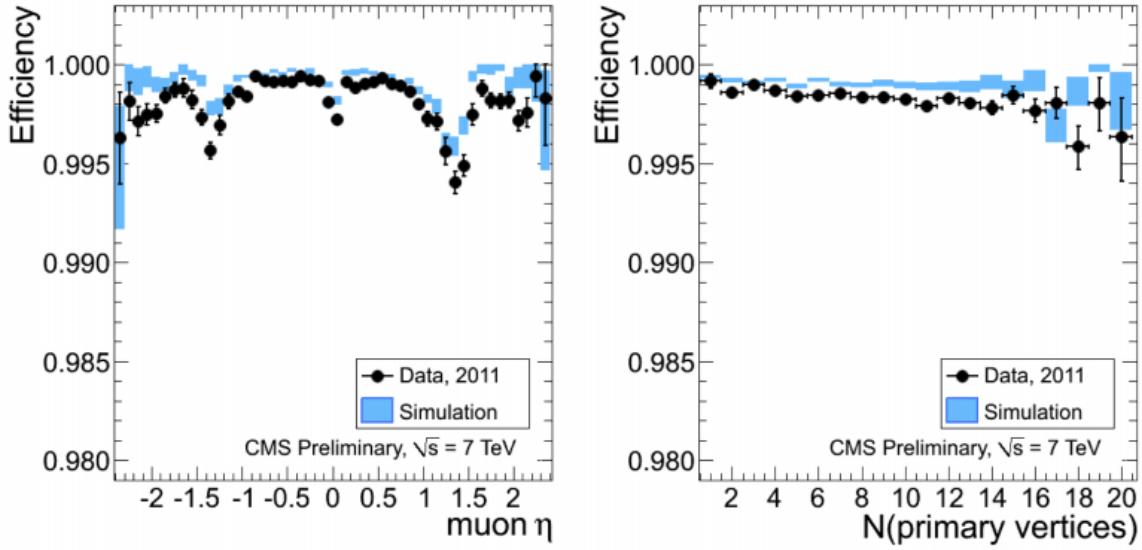


Figure 2.6: Muon reconstruction efficiency in the tracker as functions of pseudorapidity (left) and the number of proton-proton interaction vertices (right) [5].

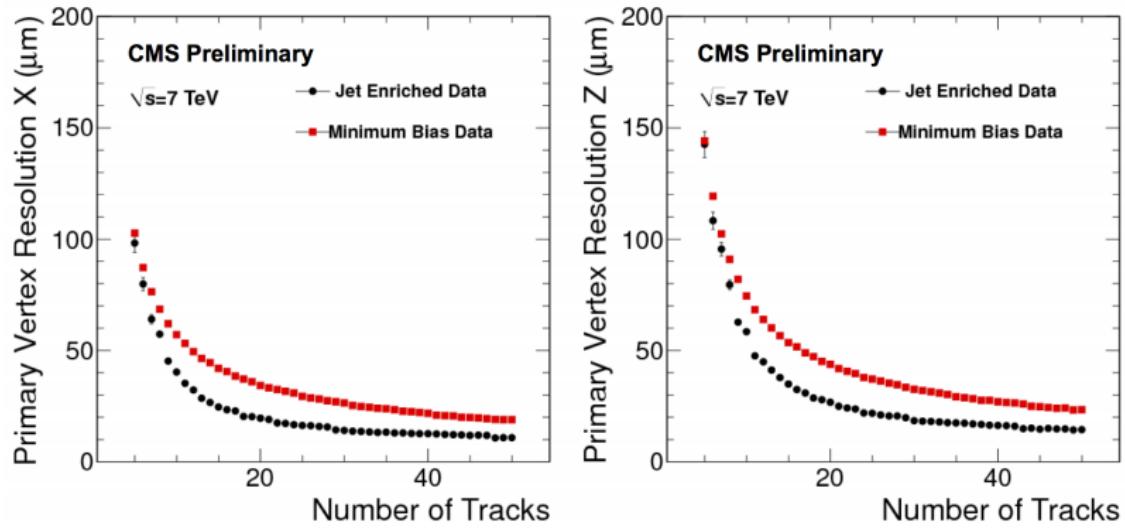


Figure 2.7: Primary vertex resolution in the transverse plane (left) and along the beam-line (right) as functions of the number of tracks attached to the vertex [5].

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 [39]. 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

[40]. 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 ( $\text{PbWO}_4$ ) crystal calorimeter [41], shown in Figure ???. The  $\text{PbWO}_4$  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 [41].

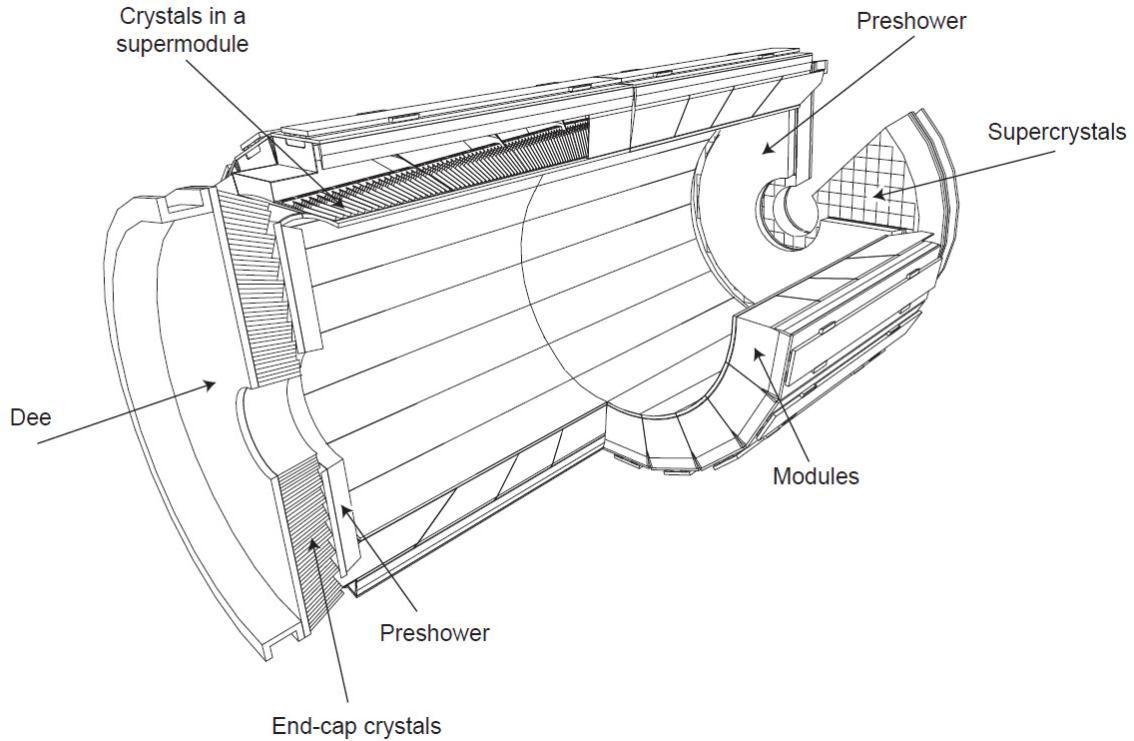


Figure 2.8: Geometric view of one quarter of the ECAL (top). Layout of the CMS electromagnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (bottom) [4].

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

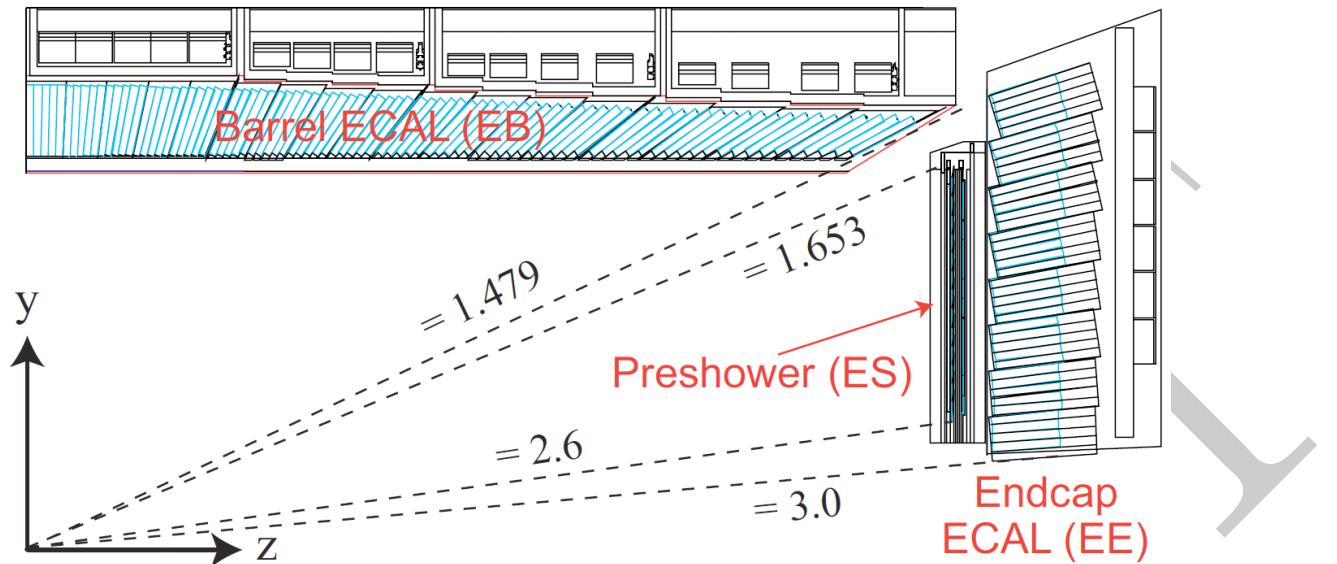


Figure 2.9: Geometric view of one quarter of the ECAL (top). Layout of the CMS electromagnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front (bottom) [4].

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  $\sim 26$  (EB) and  $\sim 25$  (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  $11 \text{ m}^3$  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 ???. 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  $\sim 3$  radiation lengths. The purpose of the ES is provide improved separation of photons to  $\pi^0$ s.

### 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  $\sim 50$  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  $280 \text{ mm}^2$  with a 20% QE and gain of  $\sim 10$ . 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 PbWO<sub>4</sub> crystals. We find that the performance of the ECAL can be attributed to inconsistencies in longitudinal light collection, noise due to the read-out electronics, and dead material within the detector causing energy leakages. The resolution was measured from test-beam data, using electrons within the energy range of between 20 and 250 GeV. The values of A, B, and C have been measured to be 2.8%, 12%, and 0.3%, respectively [4]. The results were obtained in the absence of magnetic field, very little inert material in front of the calorimeter, and where the beam was aligned to the centres of the ECAL crystals.

For unconverted photons within the energy range of interest for physics analysis ( $\sim 100$  GeV) we see that the energy resolution of the ECAL will be dominated by the constant term<sup>1</sup> in Equation 2.4. The result signifies that the performance of the ECAL is largely dependent on the quality of calibration and monitoring during data-taking periods.

Figure ?? shows the invariant mass resolution as a function of time for prompt and re-reconstructed electrons (left), and the relative electron resolution as a function of pseudo-

<sup>1</sup>For photons and electrons converting at small radii within the silicon tracker, the dominant contribution to the energy resolution arises due to radiation in the silicon tracker material combined with the effect from the magnetic field

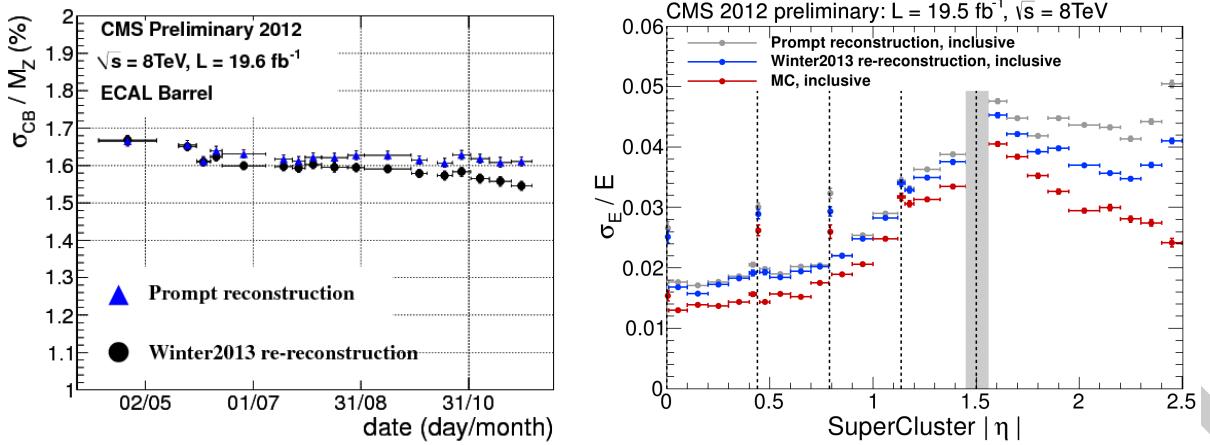


Figure 2.10: Left: The mass resolution of the Z peak, reconstructed from its di-electron decay mode, as a function of time for the barrel. Right: Relative electron (ECAL) energy resolution unfolded in bins of pseudo-rapidity  $\eta$  for the barrel. [6].

rapidity,  $\eta$ , (right) both used a  $Z \rightarrow e^+e^-$  sample.

The electron energy resolution,  $\sigma_e/E$ , is derived from the peak width of  $Z \rightarrow e^+e^-$  decays and fitting in bins of  $\eta$  using an unbinned maximum likelihood fit to the invariant mass distribution of  $e^+e^-$  pairs. We find that the energy resolution is greater than 2% in the eta range  $|\eta| < 0.8$ , and ranges from 2 - 5% in other regions. We can also see the effect created by the excess of material upstream of the ECAL in the regions where  $|\eta| > 1$ , also near the detector cracks (shown as a grey line) between ECAL modules. The difference between data (blue) and MC (red) is thought to arise due to flaws in simulation, such as a mismodelling of the upstream material.

## 2.5 Hadron Calorimeter

### 2.5.1 Overview

The Hadron Calorimeter (HCAL), shown in Figure ??, 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  $\phi$ .

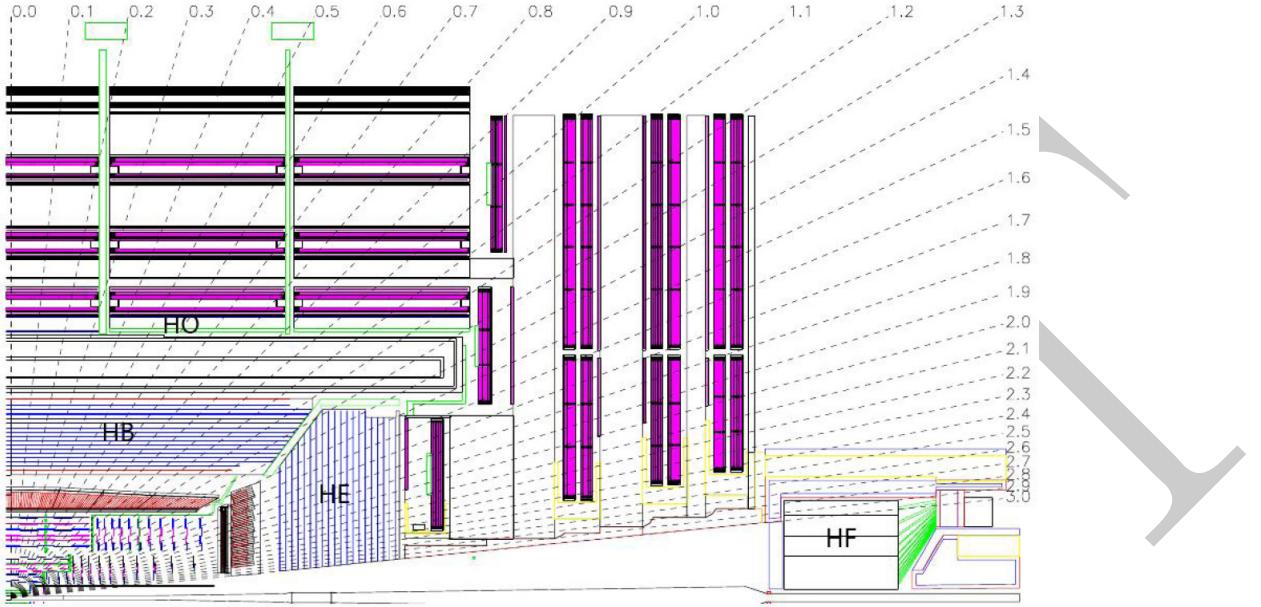


Figure 2.11: Longitudinal view of one quarter of the detector in the  $r - \eta$  - plane, showing the positions of the HCAL parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF) [4].

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 ( $\lambda_l$ ). 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  $\Delta\eta \times \Delta\phi =$

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

Table 2.2: Tower segmentation in azimuthal and polar angle for the hadronic barrel, endcap and forward calorimeter [20].

$0.17 \times 0.17$  for  $|\eta| \geq 1.6$ , whereas the forward segment changes with respect to pseudorapidity 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  $|\eta| = 5.0$  [20], as seen in Table ???. 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 [42]. 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 ??, 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, 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  $-268.5^{\circ}\text{C}$  – 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

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 [21].

requires a return yoke, which can be viewed in the CMS schematic in Figure ??, 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 ?? [43].

## 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, []) highlights the discovery potential for muon final state decay modes and the

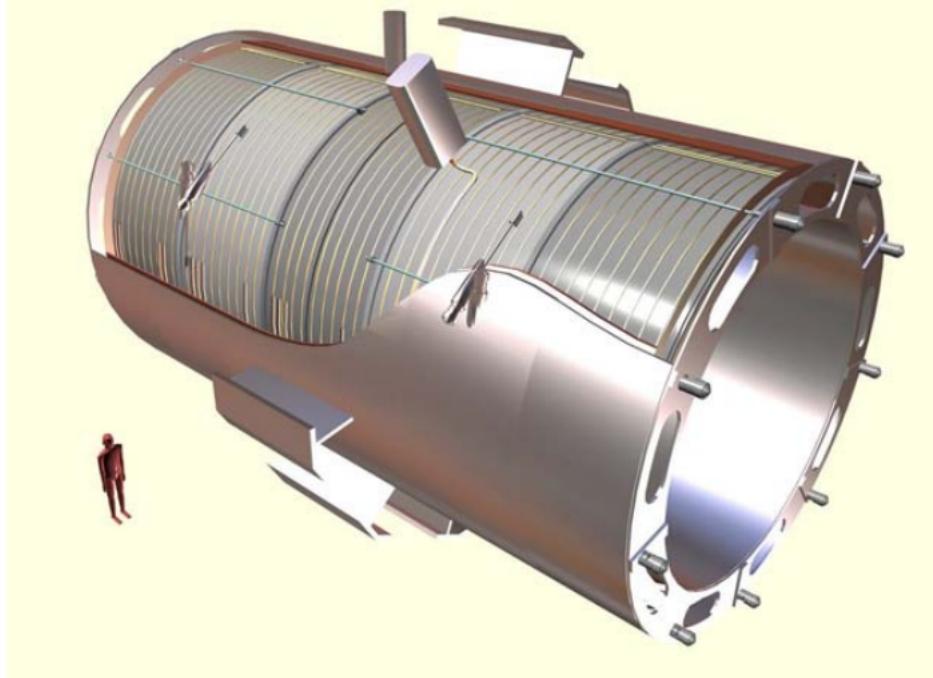


Figure 2.12: General artistic view of the 5 modules composing the cold mass inside the cryostat, with details of the supporting system (vertical, radial and longitudinal tie rods) [4].

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 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  $25,000 \text{ m}^2$  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

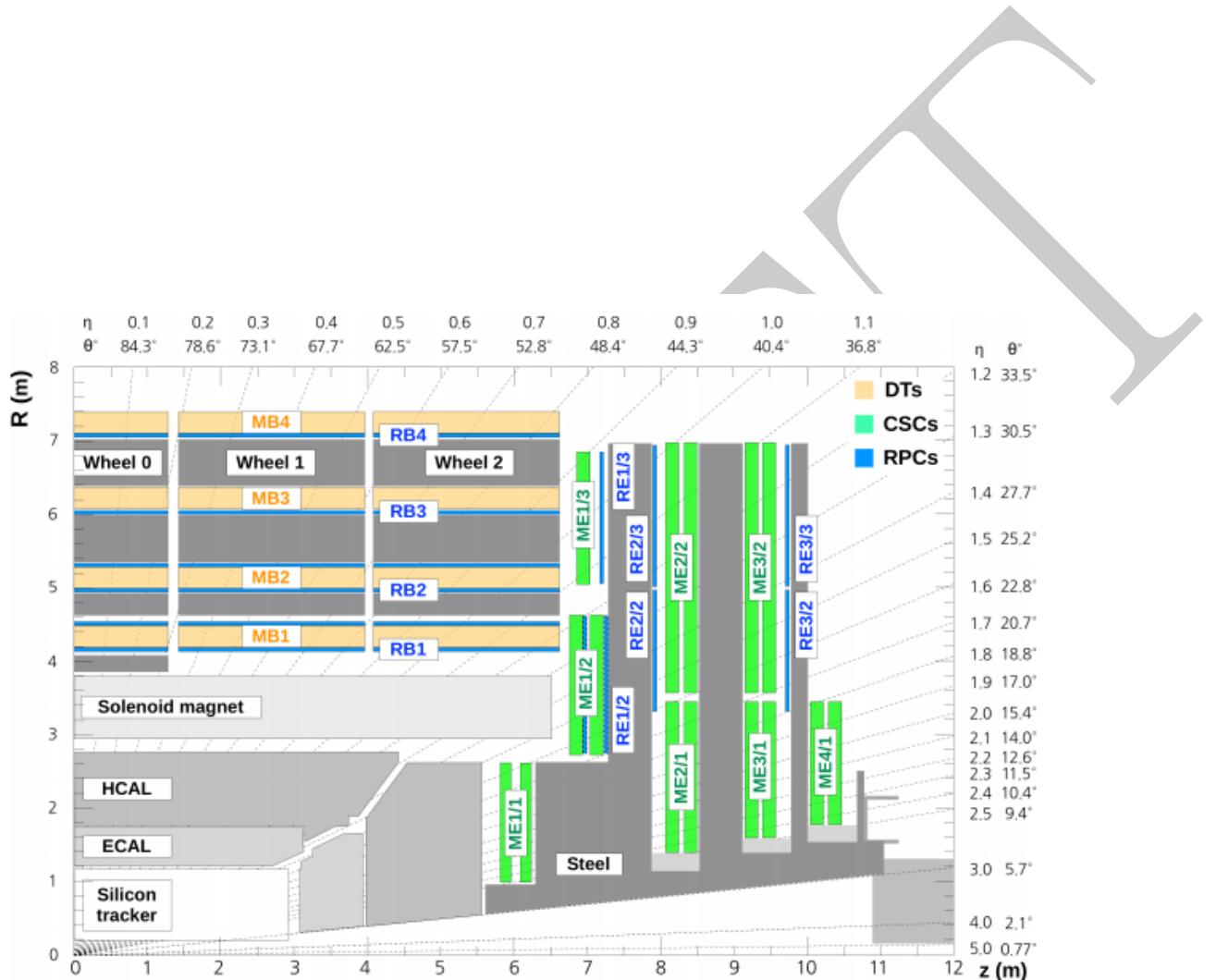


Figure 2.13: Layout of one quadrant of CMS. The figure shows the four DT stations in the barrel (MB1-MB4, yellow), the four CSC stations in the endcap (ME1-ME4, green), and the RPC stations (RB1-RB4 and RE1-RE3) [4].

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

During the Run I data taking period, CMS muon reconstruction was performed using the silicon inner tracker in coincidence with up to four of the available gas-ionisation muon chambers: DTs, CSCs, and RPCs. Standard CMS muon tracks can be reconstructed independently both in the inner tracker and in the muon system, where tracks are labelled tracker and standalone-muon tracks, respectively. When reconstructing a global muon, we rely on local reconstruction of objects within each individual muon chamber. Each muon chamber uses different techniques in reconstructing charged particles that cross the chambers, however every case relies on the ionisation of gas within the volume of the detector.

Basic objects that traverse the ionising gas within the chambers are then read out and reconstructed as hits, which are defined as spatial points with assigned uncertainties, and segments, which are obtained by fitting straight lines to the reconstructed hits [44]. For

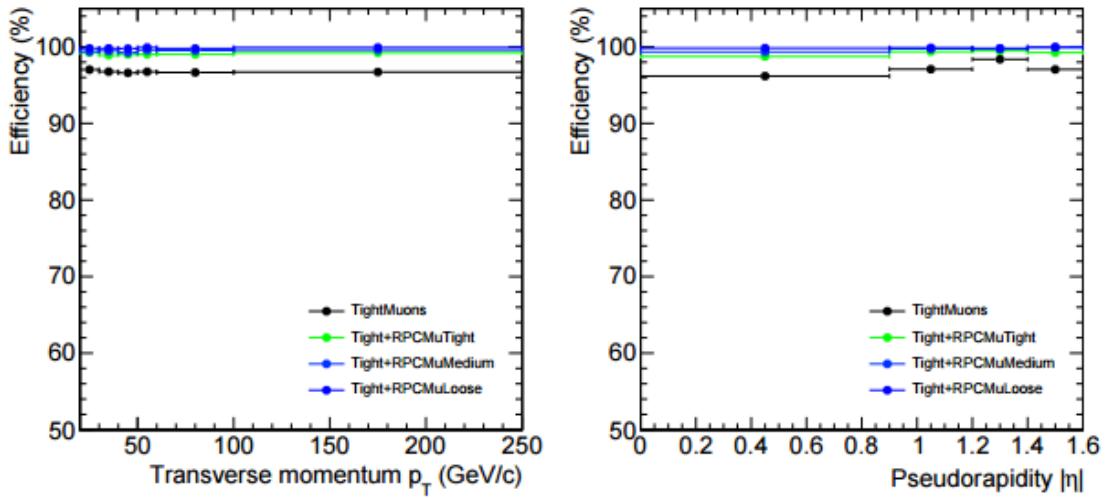


Figure 2.14: RPCMuon efficiencies on top of the semi-Tight muon selection in  $p_T$  and  $|\eta|$ . [7]

the reconstruction of standalone-muon tracks, a matching tracker track is pinpointed by comparing the parameters of each and propagating onto a common surface — which is known as the outside-in method. Global muons are fitted by combining hits from each of the tracker and standalone-muon tracks and using the Kalman-filter technique.

RPC muons are reconstructed in a similar fashion to tracker muons by requiring only two reconstructed hits in different layers of the RPC chambers. The matching criteria for each of these types of reconstructed hits is the distance and pull between their positions in a local x and y frame, where we define the pull as the residual between the extrapolated positions over the combined uncertainty [7]. The pulls and residuals for track segment matching has been studied previously [44]. By reconstructed tracker muons in this manner we drastically reduce the probability of reconstructing punch-through hadrons. By combining different algorithms we are able to increase the efficiency of correctly reconstructing a muon.

The efficiency of correctly reconstructing a muon track is studied by way of the tag-and-probe method (see Section 6.1.2) using a sample of Z bosons decaying to muons. The analysis is conducted in a semi-Tight muon selection regime, whereby the candidate is required to be identified as a RPC muon with hits matched in at least three RPC layers, and two matched hits in two muon stations. The semi-Tight muon selection does not include particle flow identification and isolation cuts. The results yield and increase of around 3% in the efficiency compared to previous measurements [7], and can be seen in Figure ???. The efficiency lies above 99% in for most  $p_T$  and  $\eta$  regions.

## 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  $\sim 1$  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 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 ???. 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  $\sim 30$  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.

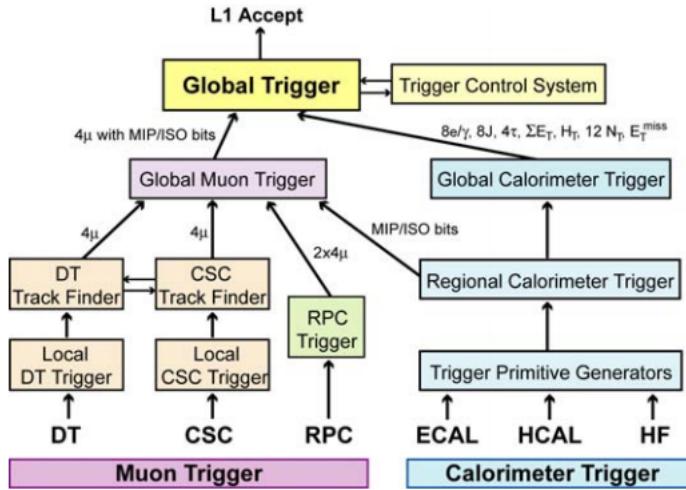


Figure 2.15: Architecture of the Level 1 trigger [4].

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 information 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  $\sim 300$  Hz and a final data rate of  $300 \text{ MBs}^{-1}$  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

The majority of physics analyses in CMS that published findings using Run I data at a centre-of-mass of  $\sqrt{s} = 8$  TeV used complicated triggers, where multiple categories of objects may be incorporated, such as jets and leptons. This category of trigger looks at the whole event topology and calculates such quantities as the total scalar energy,  $H_T$ , or missing transverse energy.

At the new unprecedented energies recorded by the LHC, gluon fusion is the most dominant process in proton-proton collisions, and as a result the rate of production for top-quark pairs increases and the LHC essentially becomes a top factory. As a result, the production cross-section for top pairs is one of the most fundamental measurements to compute. The measurement is computed using the semi-leptonic channel, such that one of the top quarks

will decay to a b-jet and a lepton (electron or muon) and a neutrino, and the other will decay to a b-quark and two jets [45, 46].

In order to accommodate the high instantaneous luminosity and pile-up throughout Run I, various triggers were used: Lepton triggers with very tight identification and calorimeter isolation requirements, jets and PF jets were limited to the central region. Similar techniques were implemented at the L1 stage of reconstruction. Muons were defined within the central region ( $|\eta| < 2.1$ ), charged hadron subtraction was introduced to deal with pile-up mitigation, and during the second half of the 2012 running period online jet energy calibrations were implemented which resulted in higher  $E_T$  thresholds in three-jet trigger paths.

The lepton plus jets efficiency is measured as the product of two independent efficiency measurements,  $\epsilon_{Lep.} \times \epsilon_{Had.}$  using the tag-and-probe method with simulated Drell-Yan and  $t\bar{t}$  samples, as described in Section 6.1.2. To estimate the top acceptance simulated events were used and are corrected for the trigger efficiency measured in data.

Figure ?? shows the turn-on curve for the efficiency for selecting the hadronic leg ( $p_T$  of the 4th jet) of the electron plus jets channel, and the dependency of the efficiency on the number of reconstructed primary vertices. The offline selection of the  $p_T$  for the jets was designed to assure a plateau behaviour of the scale factors, such that there is no variation with respect to MC sample or jet energy calibrations. From the variation of the scale factors, a systematic uncertainty of 2% (1.5%) for the electron (muon) scale factors cover the variation around their value of 0.995 (0.987) [8].

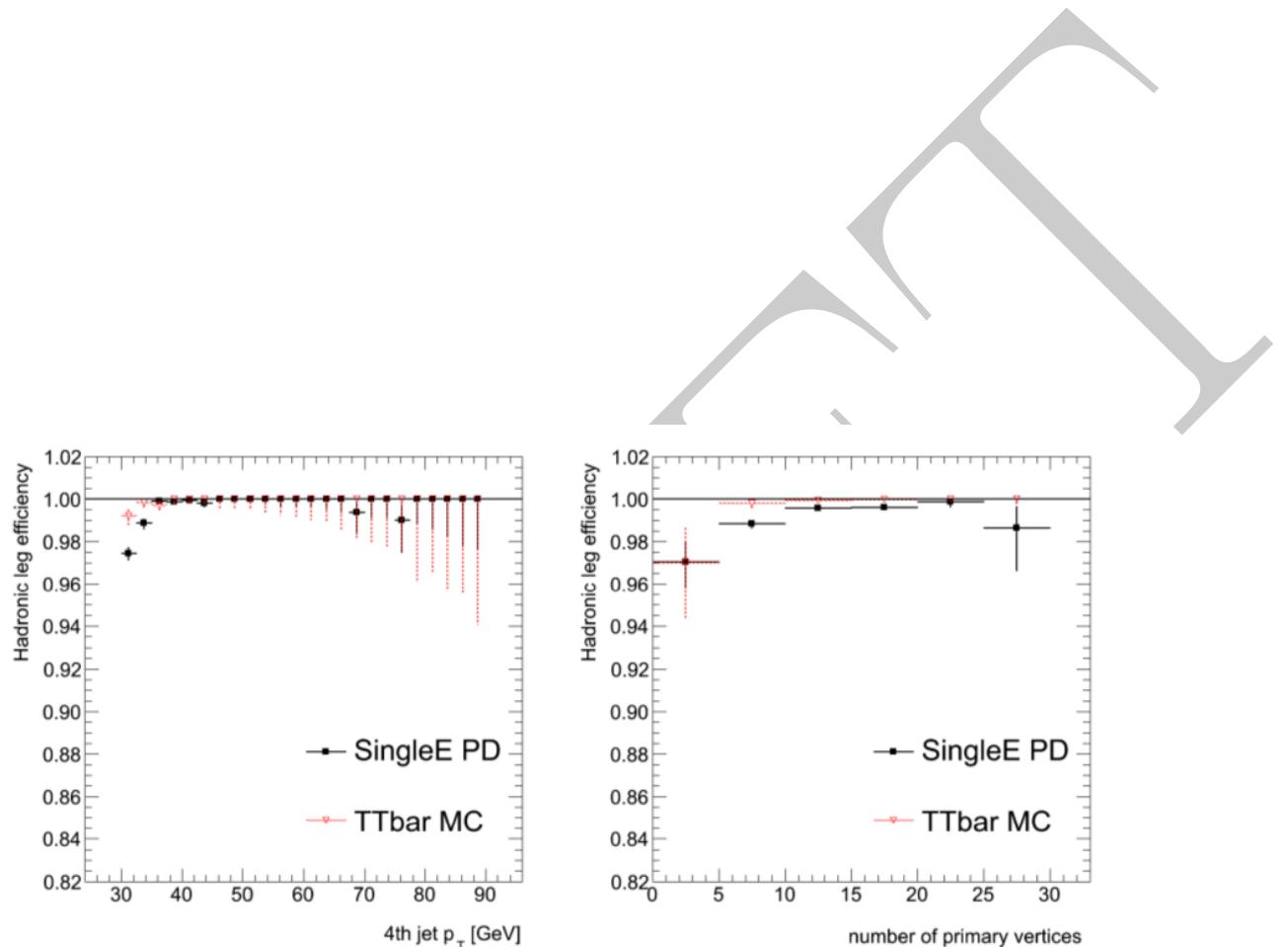


Figure 2.16: Top Triggers: Efficiency of the hadronic leg for the electron plus jets paths in 2012 versus the  $p_T$  of the 4th jet (left) and the dependence with respect to the number of vertices (right). [8]

## 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 [13, 47] as mentioned above.

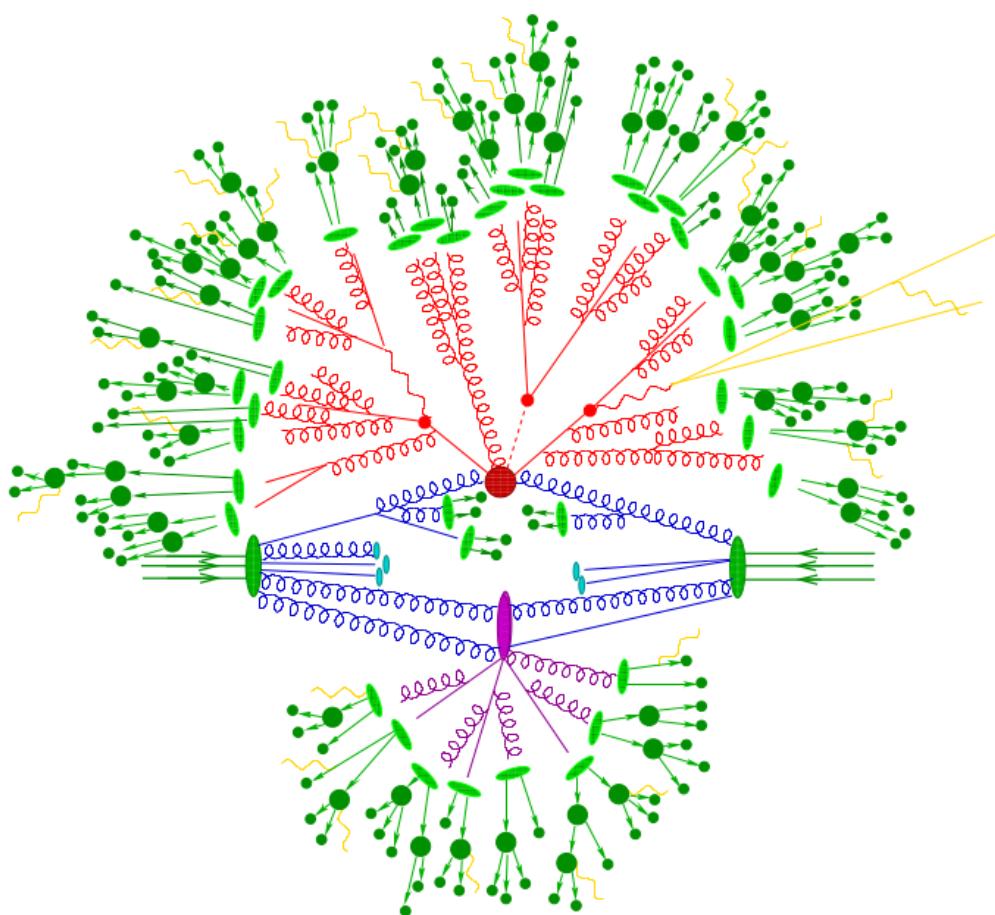


Figure 3.1: Schematic overview of a hadron collision process [9].

## 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 [48]. 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  $\sim 50$  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 ???. 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  $\sim 3$  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 [49], 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 [50] is used to simulate

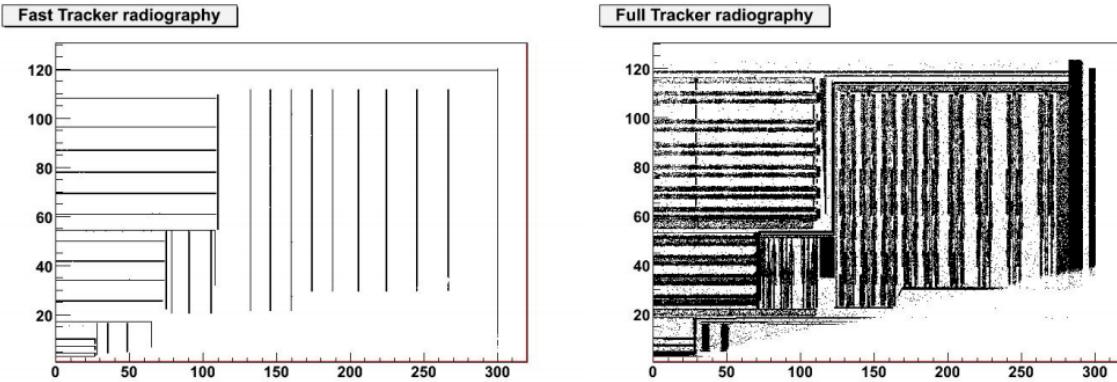


Figure 3.2: A radiography of a quarter of the simulated tracker geometry in the (a) fast and (b) full simulation [10].

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) [51] and Fast Simulation (FastSim, previously named FAMOS) [10]. 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 ??.

## 3.4 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 ?? 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** [52] 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** [53] 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 [54] only if a parton-jet pair satisfies a predefined  $\Delta R$  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** [55] 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** [56] 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** [57] (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 ?? 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 ?? and shows the final state of the process using each technique [11]. The parton distribution function CTEQ6L1 [58] is interfaced to WHIZARD via LHAPDF [59]. 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 [60], TAUOLA and PHOTOS [61] 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 levels 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-PU_S10_START53_V7A_ext-v1/AODSIM	245.8	31223821
$t\bar{t}(Semileptonic)$	/TTJets_SemiLeptMGDecays_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A_ext-v1/AODSIM	245.8	25424818
$t\bar{t}(Inclusive)$	/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM	245.8	6923652
Drell-Yann, $10 < m_{ll} < 50$	/DYJetsToLL_M-10To50_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	11050.0	37835275
Drell-Yann, $m_{ll} > 50$	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3350.0	30459503
$Z + \gamma$	/ZGToLLG_8TeV-madgraph/Summer12_DR53X-PU_RD1_START53_V7N-v1/AODSIM	159.12	6588161
Single Top tW	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	11.1	497658
Single TopBar tW $\bar{t}$	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	11.1	493460
Single Top t	/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v3/AODSIM	56.4	99876
Single TopBar t	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30.7	1935072
Single Top s	/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3.79	259961
Single TopBar s	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1.76	139974
W+Jets	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM	36257.2	57709905
$W + \gamma$	/WGToLNuG_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-	553.92	4802358

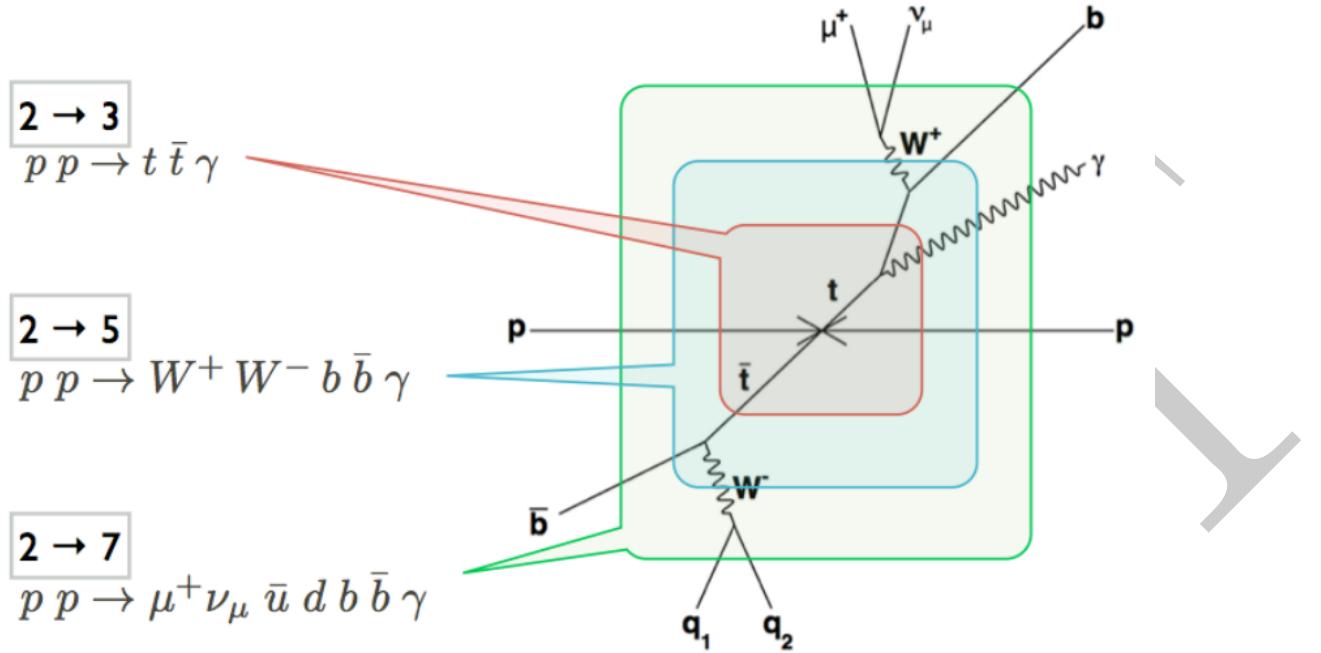


Figure 3.3: Process generation. The red, blue, and green boxes depict the matrix element calculation. Background processes with the same final state are included as well [11].

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 [52], 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  $1.8 \pm 0.5$  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

The final version of the analysis uses the factorised 2 → 7 process for the simulation of the  $t\bar{t} + \gamma$  signal sample. Instead of computing one single matrix element for the complete  $t\bar{t} + \gamma$  process, we factorise the process into individual processes and calculate the matrix element for each sub-process. The major restriction on the calculation of the matrix element lies in the computational time, such that using a factorised matrix element method vastly reduces the computation time to calculate the matrix element. The matrix element is divided in the following manner:

$$pp \rightarrow t\bar{t} + \gamma, \quad t \rightarrow bxx, \quad \bar{t} \rightarrow \bar{b}xx \quad (3.1)$$

$$pp \rightarrow t\bar{t}, \quad t \rightarrow bxx\gamma, \quad \bar{t} \rightarrow \bar{b}xx \quad (3.2)$$

$$pp \rightarrow t\bar{t}, \quad t \rightarrow bxx, \quad \bar{t} \rightarrow \bar{b}xx\gamma \quad (3.3)$$

where  $x = u, d, s, c, \bar{u}, \bar{d}, \bar{s}, \bar{c}, e^\pm, \mu^\pm, \tau^\pm, \nu_{e,\mu,\tau}, \bar{\nu}_{e,\mu,\tau}$ .

In order to verify the factorisation method, we generated a sample using WHIZARD with the full, un-factorised, 2 → 7 matrix element generation process, which we then compared to a sample simulated in WHIZARD using the factorised matrix element method. The results of this comparison can be seen in Figure ??, where we see a good agreement in the shapes of observables. We observe small fluctuations between bins within the full 2 → 7 sample, which we attribute to the number of events. We can therefore verify that the method can be used as a good approximation of the full 2 → 7 process.

Upon validation of the factorisation method for the full process, we observe the method to be a good compromise between accuracy and computational time. A comparison is then conducted between the MADGRAPH and WHIZARD event generators using the full 2 → 7 process. Within the WHIZARD framework, we are unable to compute the three factorised processes at the same time, and must therefore generate the individual processes separately and later combine them, whereby scaling is taken into account.

For both generators, the original  $pp \rightarrow t\bar{t} + \gamma$  process and the decay of the top quarks are distinct processes. The MADGRAPH framework handles both processes internally, however we must simulate the two must be handled individually in the WHIZARD framework and an extra set of kinematic cuts must be implemented, as shown below.

- $p_T(\gamma) > 1 \text{ GeV}$
- $\Delta R(\gamma, X) > 0.1$

These cuts are required to be softer than the final generator cuts, due to the dilution of kinematics by resolution effects at the value of the cuts. This is such that if the value of cuts used for decays is too close to the final generator cuts, resolution effects may be modelled incorrectly. When calculating the cross-section, WHIZARD does not take the final generator cuts into consideration. They are simply used to select events which are then written to the output file. The cross-section must be re-weighted when taking individual efficiencies into account, leading to the simulation taking more time than it would with the final generator cuts included, where a vast amount of generated events are rejected by the tighter cuts.

It must be noted that for simplicity, in the comparison we only select semi-leptonic events in the muon channel. However, for the full sample we incorporate all decay channels for the  $t\bar{t} + \gamma \rightarrow 7$  process in order to be inclusive as possible since the sample will be an official CMS sample.

We calculate the cross-sections to be

$$\text{WHIZARD : } \sigma_{t\bar{t}+\gamma} = 1408, \quad \text{MADGRAPH : } \sigma_{t\bar{t}+\gamma} = 1227.1 \pm 0.4 \text{ fb} \quad (3.4)$$

for each event generator, respectively. We note that there is no associated error with the WHIZARD cross-section, this is due to the cross-section calculated with this event generator not taking the final cuts on the decay products into account. The final cross-section is calculated from the remaining number of events. Additionally, the branching ratios for the top decays are not included within the cross-section calculation, as additional calculations are required by the user. This is more of a feature in channels where a photon is required.

The calculation steps, described previously, are estimated individually and the uncertainty on each usually contributes to the overall uncertainty by propagating them onto the final value. However, this is not calculated in this analysis, because the cross-sections are used to compare the event generators. Thus, an exact calculation of the errors is not necessary, and a NLO cross-section is used in the analysis which renders the differences inconsequential. The difference of 12.9% observed between the two generators is thus within the expected range of deviation.

A comparison between the two event generators for the factorised  $2 \rightarrow 7$  process, for three key variables, can be seen in Figure ???. We observe a good agreement between the two event generators except for small discrepancies in the angle between the photon and b quark at low values of  $\Delta R(\gamma, b)$ . We regard this discrepancy as irrelevant to our analysis due to the manner in which jets and isolation is handled in a proton-proton collider experiment such

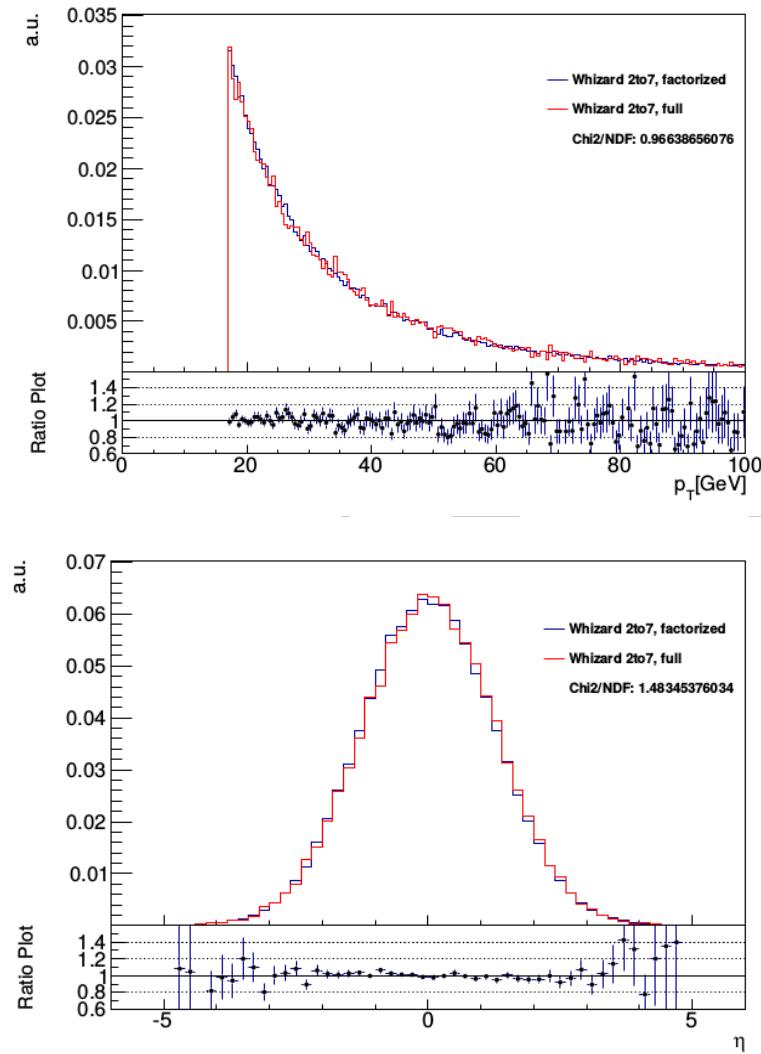


Figure 3.4: Comparison between the full and the factorised  $2 \rightarrow 7$  process using samples generated with WHIZARD. All distributions are normalised to unity. [12]

as CMS. Another explanation could lie within the normalisation process of the WHIZARD sample, and could affect the shapes of the distributions.

- $p_T(\gamma) > 13 \text{ GeV}$
- $|\eta(\gamma)| < 3.0$
- $\Delta R(\gamma, \text{all}) > 0.3$ , 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) [13], 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.

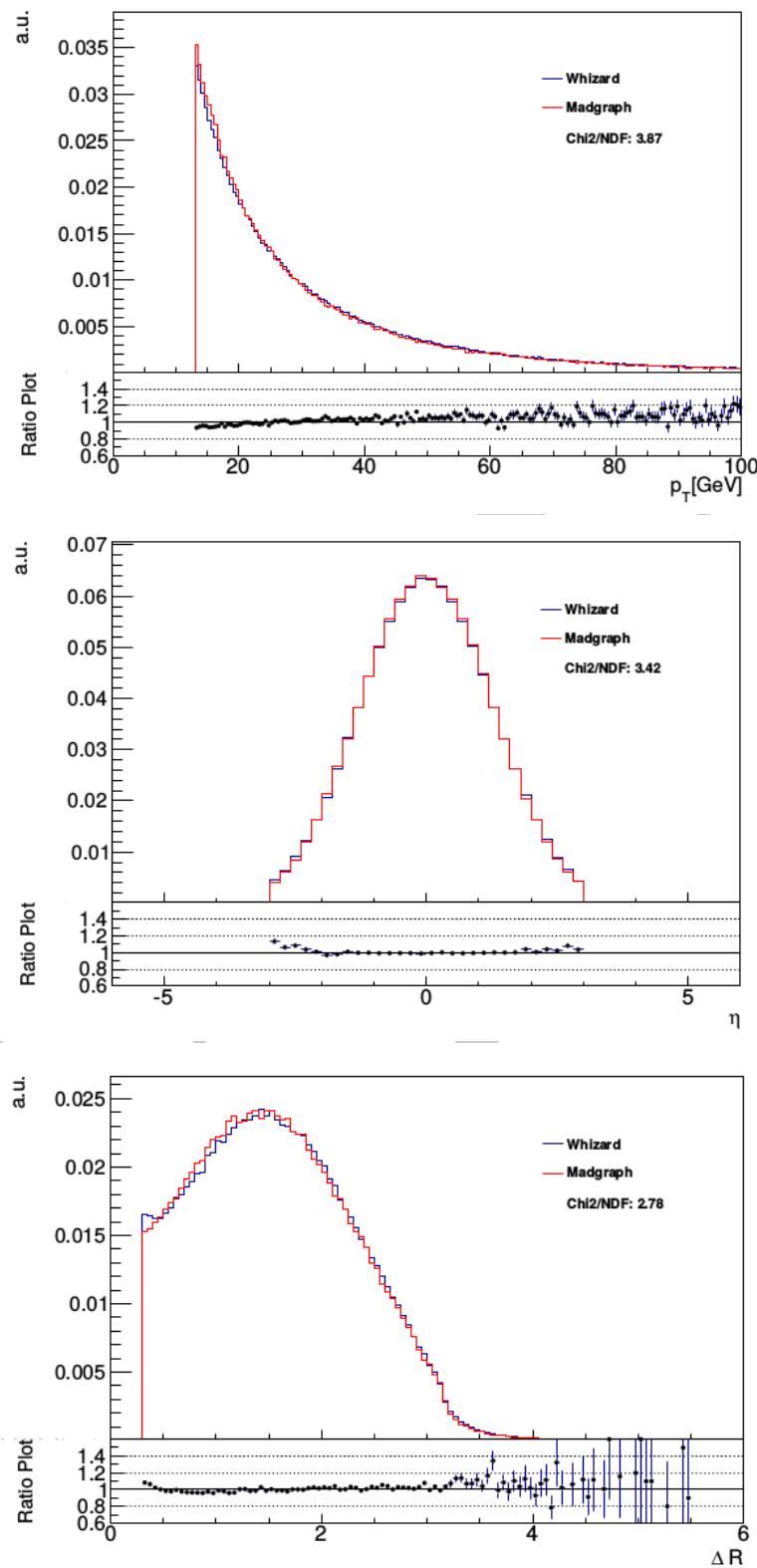


Figure 3.5: Comparison between the MADGRAPH and WHIZARD generators for the factorized  $2 \rightarrow 7$  process. All distributions are normalized to unity. [12]

### 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 [62]. 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 3.8) 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 [62]. 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 [63], 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  $\chi^2$ . 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 ( $\epsilon$ ) 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  $\tau$  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% [62].

### 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  $\tau$  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  $\chi^2$  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) [64] 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  $\chi^2$  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 [13]. 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

The particle flow (PF) algorithm [13] uses information from each sub-detector to reconstruct and identify all stable particles produced in proton-proton collisions at the LHC, such as electrons, muons, photons, charged hadrons, and neutral hadrons, and determine their direction, energy, and type. The information is then used in a similar manner to events from MC generators in order to reconstruct higher-order objects, such as jets (from which we infer the directions and energies of the quarks and gluons), missing transverse energy (giving a rough measurement of the direction and energy of invisible particles, such as neutrinos), and tau leptons, and to tag b-jets.

The design of the CMS detector is ideal for this type of reconstruction, due to it being completely hermetic around the interaction point with a 3.8 T magnetic field, thus the reconstruction of charged particle tracks (which make up around 2/3 of the total objects) is extremely efficient with a small fake-rate down to a low  $p_T$  of around 150 MeV/c and pseudorapidities as large as  $|\eta| < 2.6$ . Energy deposits are recorded as ‘blocks’ in the detector, which are then interpreted as particles within a particular sub-detector. The PF algorithm feeds the information obtained by each sub-detector into the ‘link’ step, where it then groups together combinations of blocks that are likely to originate from particles. The combinations are then categorised as individual objects and stores a list of particles of each type.

Particles are composed of several particle-flow elements from a number of CMS sub-detectors: one charged particle track, and/or various clusters in the calorimeters, and/or one muon track in the muon system. The link algorithm is then used to connect the tracks and energy clusters to reconstruct the particle, whilst removing possible double counting from any of the sub-detectors. Pairs of elements are found and ‘linked’, such that the link is extrapolated to the end of a shower for a typical electron in the ECAL or the typical interaction length of a hadron shower in the HCAL, where a distance parameter defines the quality of a ‘link’. Blocks of elements in each sub-detector are produced by the algorithm which are classes as directly or indirectly linked. The PF reconstruction process is made easier by its high granularity, such that each block contains a maximum of three elements, and thus the performance of the algorithm is independent of the event complexity. However, the size of a block may be increased by up to a cell to account for the non-uniformity of the calorimetry system.

Once blocks are processed by the link algorithm, they are then passed to the PF algorithm,

where the objects are reconstructed and identified on an event-by-event basis. When an object is fully reconstructed, the tracks and clusters from the block are then removed from the collection and the next object is reconstructed. The first object type to be reconstructed is the muon, due to the efficiency of identifying muons. We measure muons from their associated Global and Tracker tracks, such that if they are within 3 standard deviations from each other we call the object a PF muon. The tracks and clusters associated with the blocks are then removed from the collection.

Secondly, electrons are identified and reconstructed from the, now much smaller, collection of tracks. The GSF method of associating clusters is used to differentiate between charged hadrons and electrons, and the electromagnetic shower must be thin in  $\eta$ . Charged pions give rise to a source of fake electrons and are vetoed by other such parameters as the number hits and  $\chi^2$ . Many other variables are then input into a multivariate analysis tool which then provides a discriminant on whether or not the particle is an electron. If the particle passes the MVA (Table 3.2) it is then labelled as a particle flow electron.

By process of elimination, all that remains are charged hadron, neutral hadron, and photon candidates. By matching any charged tracks associated to a HCAL cluster, we can define these particles as PF charged hadrons. By implementing a distance parameter we make sure not to double count. We are then able to remove tracks associated with charged hadrons. Any remaining energy deposits in the HCAL is then known to originate from photons or neutral hadrons, differentiating between the two by the shower shape of the candidate. We can thus define PF neutral hadrons and PF photons.

### 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

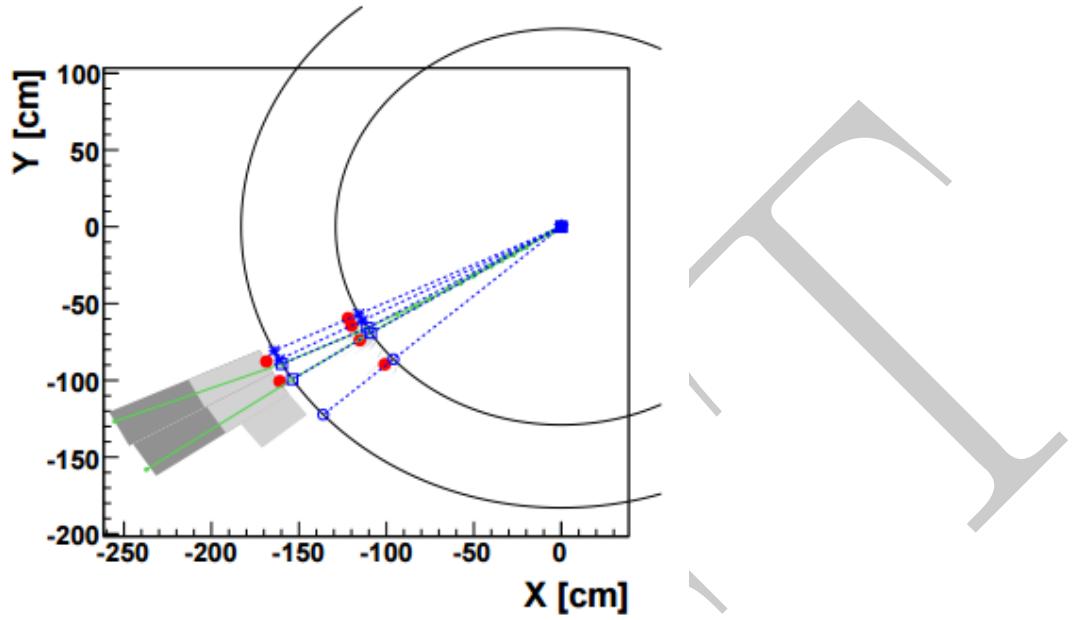
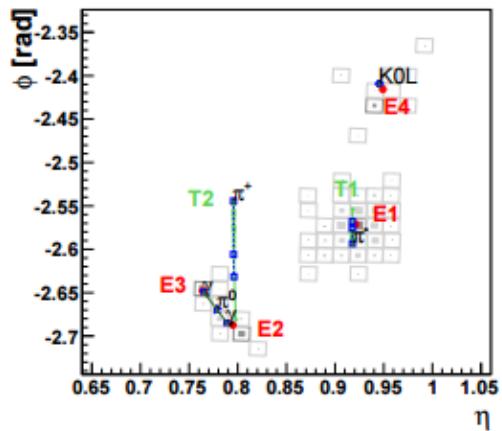


Figure 3.6: The (x, y) view [13].

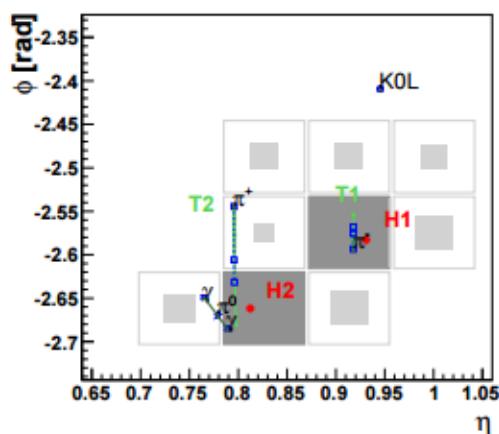
the energy is lost is due to Bremsstrahlung, the process by which a photon is emitted upon interaction within a material.

Kalman filters 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 [65]. 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 [66]. 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  $\eta$  and a wide spread in the azimuthal angle,  $\phi$ . The energy deposits of the object and its associated Bremsstrahlung form a single supercluster, where the performance of this method lies in the ability to correctly identify this. The method performs much more



(a) The  $(\eta, \phi)$  view on ECAL [13].



(b) The  $(\eta, \phi)$  view on HCAL [13].

Figure 3.7

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  $\chi^2_{KF}$  of the Kalman filter fit are taken into account before refitting using a GSF, before being characterised by a multivariate estimation tool [67].

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

Electrons are identified in the ECAL and can be recognised by their distinct shower shape deposited within the calorimeter crystals. However, other objects such as charged hadrons, jets, and converted photons can have a very similar signature to an electron deposit and may be wrongly reconstructed as such. Therefore, in order to correctly reconstruct an electron candidate, we must implement further selection criteria.

Individual physics analyses in CMS require different identification requirements tailored to personal needs, thus identification working points are initially required to be much looser than final cuts while still retaining a high efficiency of detecting electrons. When a further tight working point is implemented the efficiency for correctly selecting an electron increases. CMS uses various identification algorithms in order to correctly identify electrons: Simple cut-based ID, cuts in categories ID, particle flow ID, and MVA ID. Each of which are used in the analysis described in this thesis, and are described below.

**Simple cut-based ID** Cut-based identification techniques, while not providing the greatest performance compared with other techniques, can be a useful tool in the understanding of the data and for making comparisons with MC. Fewer variables from Table 3.7.6 are used with this method, such as  $H/E$ ,  $\Delta\eta_{in}$  and  $\Delta\phi_{in}$  between supercluster position and track direction at the vertex extrapolated to ECAL assuming no radiation,  $\sigma_{in\eta\eta}$  cluster and shape variables. [68]

**Cuts in categories ID** is a more complex version of the simple cut-based identification techniques. It makes use of electrons originating from different sources, such as  $Z$  and  $W$  decays, while reducing the likelihood of wrongly selecting “fake” electrons originating from photon conversions or jet mis-identifications, however the method has also been used to select electrons from alternative sources, such as the  $J/\psi$ . The problem of identifying electrons in CMS differs to that of other experiments due to the detector topology — the large amount of material from the tracker directly in front of the ECAL and the high magnetic field make it difficult to efficiently reconstruct electrons. It has been found that many of the problems faced in the identification of electrons can be overcome by dividing the problem into categories. The most basic category sees the ECAL split into barrel and endcap regions due to the differences in topology in the two regions for both the inner tracker and ECAL. The second comes from the large amount of bremsstrahlung radiation in the tracker material, then using the measurement for  $E/p$  to account for the poor reconstruction of the track. The final category divides electrons into high- and low- $E_T$  groups to correct for effects from the high magnetic field at high  $E_T$ .

The cuts in categories method uses the same variables as the simple cut-based method with the addition of track conversion rejection (number of missing hits near beginning of track), and track isolations with tracker isolation ( $\Delta R = 0.3$ ), ECAL Isolation (jurassic  $\Delta R = 0.4$ ), HCAL Isolation ( $\Delta R = 0.4$ ). The cuts are applied such that the signal to background ratio is optimised, where there are several levels of cut severity: *Very-Loose*, *Loose*, *Medium*, *Tight*, *SuperTight*, *HyperTight* (1-4). Each step of cut severity decreases the electron fake rate by roughly a factor of two for  $E_T > 20$  GeV. [69]

**Particle Flow ID** is the loosest of all the electron identification techniques as it is a universal particle collection for all physics analyses. The method takes information from all subdetectors and combines it in order to create new observables to aid in the identification of electrons. These variables are shown in Table 3.7.6 and are input into a

multivariate analysis tool which then results in a discriminating value. The MVA is trained with respect to signal and background MC samples. [70]

**MVA ID** is a multivariate analysis identification technique, and the most robust and common technique in physics analysis. It is used to identify electrons originating from W and Z bosons. The variables, show in Table 3.7.6, are used to produce a single discriminating value. The value is optimised by so called “training” of the MVA in different selection categories.

Variable	Description
<i>Track quality variables</i>	
$p_T$	Transverse momentum of the GSF track
$\eta$	Pseudorapidity of the GSF track
$\text{GSF } \sigma_{p_T}/p_T$	Transverse momentum resolution of the GSF track
$\#\text{hits}_{KF}$	Number of reconstructed KF track hits
$\chi^2_{GSF}$ and $\chi^2_{KF}$	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 $\eta$ , using the distribution of energy in a 5 block of crystals around the seed crystal (the one with the highest energy) [71]: $\sigma_{i\eta i\eta}^2 = \sum_{5 \times 5 \text{ crystals}} (\eta_i - \eta_{\text{seedcluster}})^2 E_i / E_{\text{seedcluster}}$
$\sigma_{i\phi i\phi}^2$	Cluster shape in $\phi$
$\eta_{SC}(\phi_{SC})$	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 $1 \times 5$ strip of the $5 \times 5$ electron cluster, and $E_{5 \times 5}$ is its total energy.
$E_{3 \times 3}/E_{SC,\text{raw}}$	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 [1].

### 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 [72]:

**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

In this section we describe different techniques to reconstruct higher-level objects, such as jets and missing transverse energy, a process which happens post identification and reconstruction of objects. Higher-level objects must also be corrected for detector defects and other inconsistencies, which is also described.

### 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  $k_t$  [73], Cambridge/Aachen [74], and SISCone [75] jet clustering algorithms. However, the most widely used jet clustering algorithm is the anti- $k_t$  [76] 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

$$d_{ij} = \min \left( \frac{1}{k_{t,i}^2}, \frac{1}{k_{t,j}^2} \right) \cdot \frac{\Delta_{i,j}^2}{R^2} \quad (3.5)$$

$$d_{iB} = \frac{1}{k_{t,i}^2} \quad (3.6)$$

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  $\eta$ , 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 [77], 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  $\eta$ -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 [78]. 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.

### 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 [79], 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 ??

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 [80] algorithm, which takes the following input parameters:

- Track multiplicity in jet.

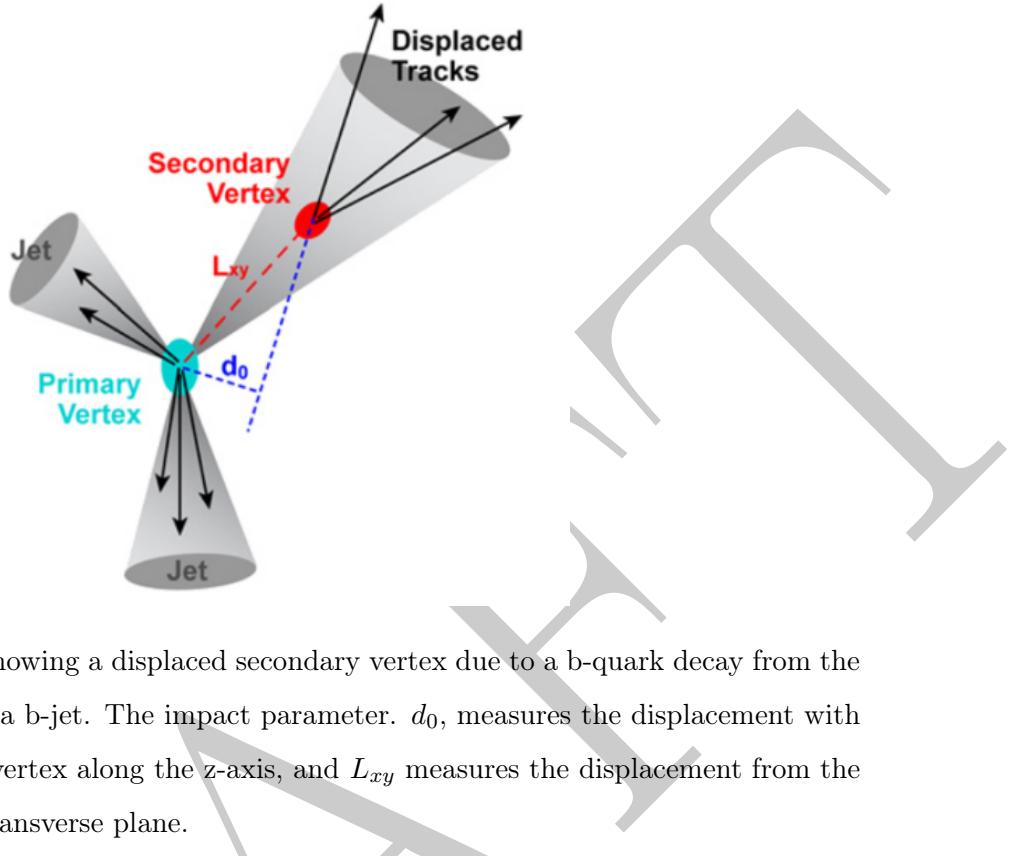


Figure 3.8: Schematic showing a displaced secondary vertex due to a b-quark decay from the primary vertex forming a b-jet. The impact parameter.  $d_0$ , measures the displacement with respect to the primary vertex along the z-axis, and  $L_{xy}$  measures the displacement from the primary vertex in the transverse plane.

- 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.
- $\eta$  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  $2.5 \text{ cm}$ .
- Impact parameter significance ( $L_{xy}/\sigma_{L_{xy}}$ ) of the first track, increasing the invariant mass above the threshold for charm production of  $1.5 \text{ GeV}$ .

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 [81] as for the

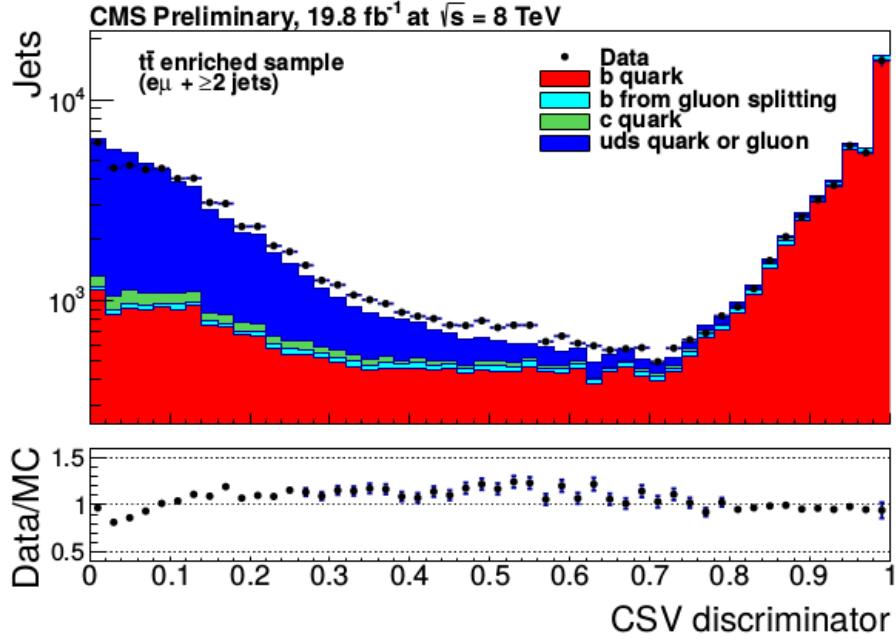


Figure 3.9: Logarithmic distribution of the partons as a function of the CSV b-tagging discriminant [14].

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 ??, showing the efficiency of correctly identifying a b-jet.

### 3.8.5 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:

$$E_T^{miss} = - \sum_i (p_x^2 + p_y^2)^{1/2} \quad (3.7)$$

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 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 [82]. 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  $\phi$  modulation in the MET.

The source of the  $\phi$  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  $\phi$ , 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

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 ( $\Delta R$  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) [83] 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.

The event selection steps for pre-selection di-muon and di-electron channels are listed as follows:

- Skim
- Event cleaning and trigger
- dilepton selection
- Z-mass veto
- $\geq 1$  jet
- $\geq 2$  jets
- Missing transverse energy selection
- $\geq 1$  CSV b-jet

For the mixed channel we select two oppositely signed leptons where one is an electron and the other a muon. Because of this feature, Drell-Yan is significantly reduced, and therefore there is no need for a Z-mass veto. We also do not require a cut on the missing transverse energy. The event selection for mixed final state event topologies are as shown below.

- Skim
- Event cleaning and trigger
- dilepton selection
- $\geq 1$  jet
- $\geq 2$  jets
- $\geq 1$  CSV b-jet

Each step will be discussed in greater detail within the following sections.

### 4.3 Skim

### 4.4 Trigger and Event Cleaning

#### 4.4.1 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  $\eta$ , 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 ??.

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\*,

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.

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  $\sqrt{s} = 8 \text{ TeV}$  data-taking period with  $19.6 \text{ fb}^{-1}$ .

#### 4.4.2 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 -  $\sim 5000$  htis per event. The HCAL laser filter was then designed and introduced for the 2012 data taking period.

## 4.5 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 4.4. 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 [83].

### 4.5.1 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 [84]. 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)  $\rho$  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  $I_\gamma$  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  $\rho$  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  $\eta_{SC}$  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  $> 10 \text{ GeV}$ , 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  $10 \text{ GeV}$
- Absolute value of pseudorapidity less than  $2.5$
- Combined relative Particle Flow (PF)  $\rho$  corrected isolation in cone  $0.3$  less than  $0.15$
- Trigger version of electron multivariate discriminator Non-Trigger MVAID greater than  $0$

#### 4.5.2 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) [85] are imposed as follows:

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

#### 4.5.3 Corrections to simulated events

## 4.6 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 [86].

**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 4.5.3. 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.

## 4.7 Missing Transverse Energy

## 4.8 Comparison of Data to MC

## 4.9 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 (loose cut-based photon ID 2012 with particle flow based isolation, [87]). In order to suppress FSR from final state particles a  $\Delta R$  cut of photons between jets and leptons is applied. All simulated data is scaled to the luminosity of the recorded data used.

The fiducial cuts on the photon are given as:

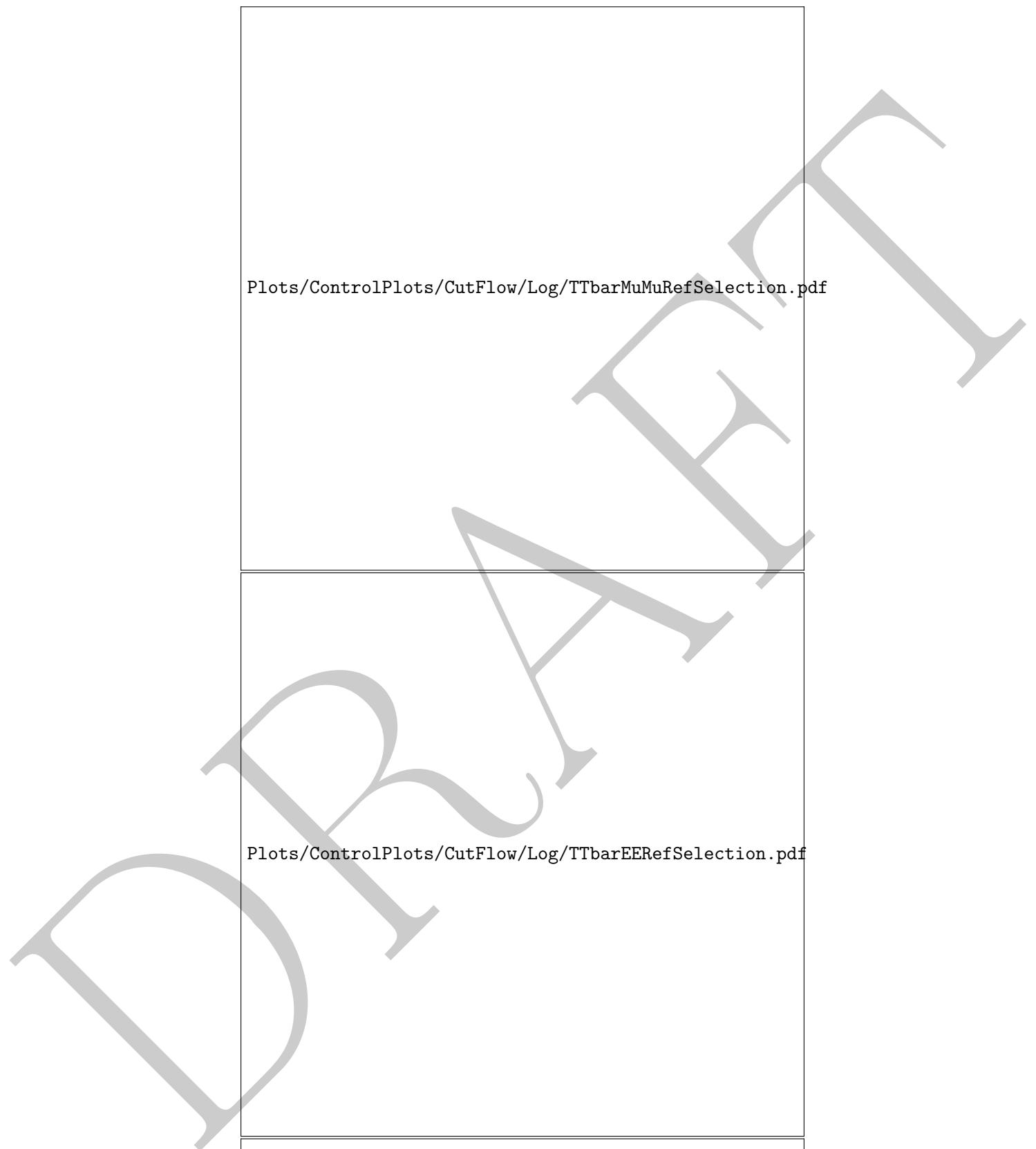


Figure 4.1

**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.

#### 4.9.1 Cut based photon ID

The cut-based photon isolation cuts are taken from the recommended values [87], with the inclusion of super-cluster footprint-removed isolation (see Section 4.9.3), and can be seen below.

**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,  $\rho$ , applied. Cut given as  $I_{char.had} < 1.5GeV + 0.04 * E_T(\gamma)$ . See Fig.??.

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

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

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

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

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

It must be noted that only super-cluster footprint-removed charged hadron isolation is used in this analysis, where PF isolation is used for neutral hadron and photon isolation.

#### 4.9.2 FSR suppression

It is crucial that initial and final state radiation (ISR/FSR) is modelled correctly for this analysis as photons from initial and final state radiation are not considered as signal and we therefore implement cuts in  $\eta$  and  $\phi$  to reduce these events. The definition of isolation has been modified from the recommended values in order to make the data-driven estimate of the selection purity more robust.

$\Delta R(\gamma, leptons)$  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  $\Delta R(\gamma, leptons) > 0.3$ .

$\Delta R(\gamma, jets)$  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  $\Delta R(\gamma, jets) > 0.3$ .

$\Delta R(leptons, jets)$  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  $\Delta R(leptons, jets) > 0.3$ .

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

### 4.9.3 Super-cluster footprint-removal for photon isolation

The  $t\bar{t} + \gamma$  analysis uses a method of removing the photon shower energy deposit, or “footprint”, from the isolation cone in order to remove the energy deposits of the selected photon from the isolation sum, and thus minimise correlation between shower shape and isolation components in our signal process — super-cluster footprint removal (SCFR). With this method the footprint is, effectively, cleaned so that the isolation sum is not biased from the presence of the photon at the centre of the cone. Shower-shape variables, defined within the super-cluster, are then decoupled from the isolation computation, defined outside of the super-cluster. When the footprint of the photon has finally been removed, the isolation sum for prompt photons is due only to pileup and underlying events. The process calculates each isolation component individually, however we only consider the charged hadron component of the isolation sum using this technique, which is then used to model our background and signal processes. This method was first implemented as a technique upon measuring the diphoton cross-section with 7 TeV data [?]. The main features of this module are improvement in the agreement between data and MC for ECAL detector-based isolation, better discriminating power against the misidentification of photons, and it allows us to use a fully data-driven method for constructing our template fit.

We define photon isolation as the sum of the transverse momenta of all particles falling within the isolation cone surrounding the photon candidate. We define the isolation cone to be within a  $\Delta R < 0.3$ , where  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ . The momentum of the prompt photon should not contribute to the isolation sum, and thus particles found close to the photon footprint are not included within the sum.

SCFR is a purely geometrical procedure which is computed as follows:

**Step 1 :** Propagate the PF candidate trajectory from the primary vertex to the surface of the ECAL. We must take into account the magnetic field for charged hadron candidates. The impact parameter,  $d_z$ , is calculated in the z direction with respect to the primary vertex, neglecting the transverse distance,  $d_{xy}$ .

**Step 2 :** If the propagated PF candidate is found to hit the surface of a crystal lying within the super-cluster, then the candidate is therefore removed from the isolation sum, as shown in Figure ???. The PF candidate is allowed to fall within a volume of 25% of the crystal size around the crystal, in order to account for the fact that the PF candidate’s energy deposit has a finite extension in the ECAL, and thus has a reasonably large effect at the edges of the super-cluster.

This is such that the super-cluster shape defines the region which is excluded from the isolation cone in each event.

In the standard particle flow algorithm isolation is calculated on an individual basis for PF candidates divided into three groups: Charged hadrons, neutral hadrons, and photons. This procedure carries potential pit-falls, listed below.

- When computing the sum of isolation, any energy deposit that is not associated with a reconstructed particle is not included within the sum.
- When reconstructing a photon, its energy may be dispersed over a large radius within the detector, and thus reconstructed as several particles. This will greatly affect the isolation of the photon.

This does not have too much of an impact when using the standard cut-based particle IDs, such as used in this analysis, however as we are interested in the isolation profile shape some improvements can be made.

Super-cluster footprint-removed isolation deals with PF candidates, whereas PF isolation deals with identified particles. This reduces the probability for the first pit-fall to happen. The second pit-fall arises due to the leakage of energy from the photon into the isolation cone around the super-cluster. Therefore, a different method for defining the isolation cone is implemented to overcome this problem, whereby summing the transverse momentum of particle flow candidates that fall within the isolation cone, but which are not close to the super-cluster in the ECAL. The  $\eta$  and  $\phi$  coordinates of the super-cluster are calculated by checking the position of hits recorded, and if a PF candidate that has been propagated to the surface of the calorimeter lies within the super-cluster, enlarged by 25%, then the photon is not considered isolated and thus the candidate is not added to the isolation sum. The standard CMS implementation for energy density correction,  $\rho$ , is applied to the isolation to account for pileup.

#### 4.9.4 Signal template construction using the random cone method

We construct our prompt signal photon template using the “random cone” method. The events that pass our full photon event selection are used as the method relies on the assumption that a prompt photon is isolated, such that the activity recorded within the detector around the prompt photon, not regarding the area affected by the energy deposit from the photon, arises only from pile-up, underlying events, and electronic noise in the ECAL.

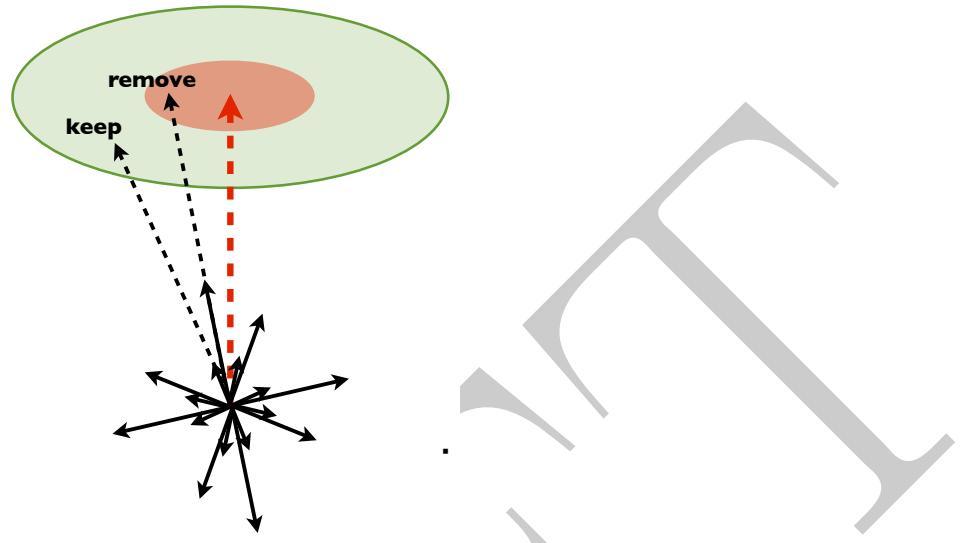


Figure 4.2: Graphic representation of the PF candidate footprint (in red) from the primary interaction vertex, within the isolation cone (green). [15]

Contributions from these effects, generally, do not vary for an isolation sum which is calculated in a region which is separate from the photon candidate, but using the same pseudorapidity range,  $\phi$ . A suitable orientation of the isolation cone is chosen by using the proceeding steps:

- We define a direction which is obtained from the photon candidate direction, and then rotating by a random angle,  $\phi_{Rand.Cone}$  between  $0.8$  and  $2\pi - 0.8$  radians, around the z-axis of the detector system (or rotating in  $\phi$  with a fixed  $\eta$ ) as shown in Figure ???. This is designed such that an isolation cone with  $\Delta R = 0.4$  centred on the random cone axis would not overlap with an isolation cone which is centred on the photon candidate, or an object that is produced back-to-back with respect to it.
- We then check for objects lying within the region, such that there is no energy originating from anything other than pile-up, underlying events, or electrical noise in the ECAL. Such that there are no AK5<sup>1</sup> PF jets with a  $p_T$  of at least 20 GeV at  $\Delta R < 0.8$ , photons with  $p_T$  less than 10 GeV at  $\Delta R < 0.8$ , or muons at  $\Delta R < 0.4$  with respect to the random cone axis. If such an object is found to fulfil our requirements, then we perform the same search again to generate another random angle,  $\phi_{Rand.Cone}$ , until a suitable direction is found. In practice, the procedure is not repeated more than twice in order to find a suitable direction.

<sup>1</sup>Anti- $k_T$  jet clustering algorithm with cone size of  $\Delta R = 0.5$  (see Section 3.8.1)

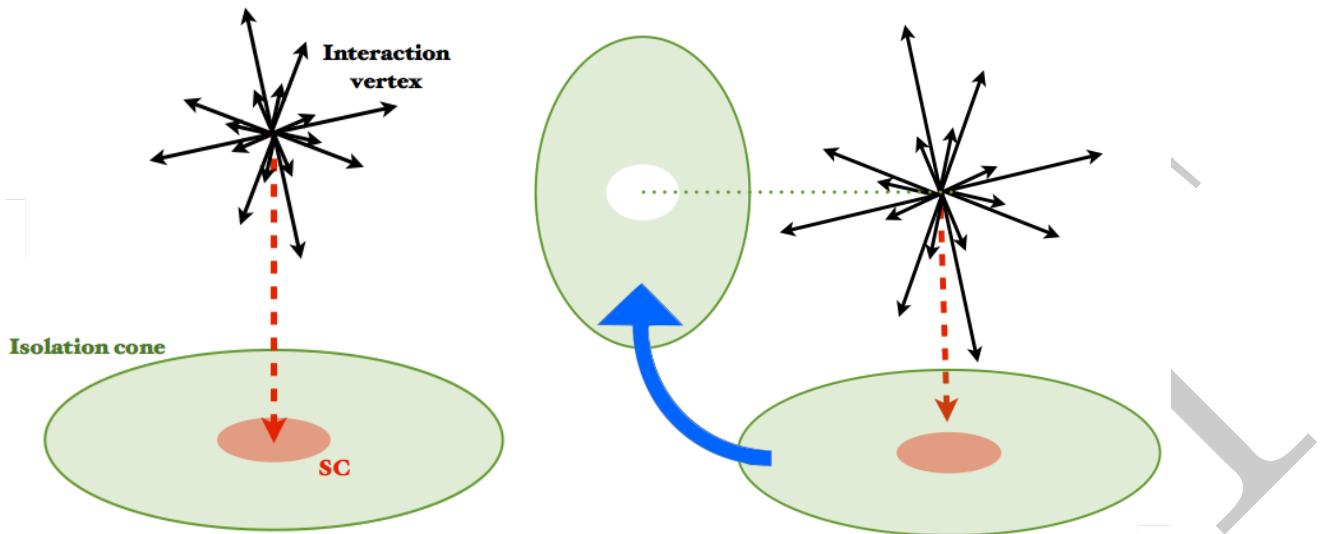


Figure 4.3: Illustration of the random cone technique. Under the assumption that the footprint of the photon (red) is completely removed from the isolation sum, the energy deposited in the random-cone area (green in right image) predicts the energy deposited in the isolation cone (green area in left image). [15]

We then calculate the isolation sum within a cone centred on our chosen direction. The difference when performing footprint removal now is that there is no super-cluster lying along the cone axis, and therefore it must be performed in a different way: the area corresponding to that in which the photon super-cluster lies, which is rotated in  $\phi_{Rand.Cone}$  in order to align with the random cone axis, is not included within the calculation process. In this manner, both the area which we consider for the isolation sum, and the area throughout the centre of the cone replacing the photon candidate footprint, has exactly the same shape and extension in the photon and random cone directions.

## 4.10 Background Estimation

### 4.10.1 Background estimation using the $\sigma_{inj}$ side-band method

Background events are identified as non-prompt or misidentified photons. We define a template to model our background by selecting photon candidates that fail one selection requirement, which we call the “side-band” region. The logic behind this method is such that by inverting a particular cut on the photon selection, then we have a sample that is rich in “photon-like” jets (whereby the hadronisation process resolves to a neutral boosted meson

with very little activity around it). In order to model the background correctly, we must select the requirements on the side-band region such that they do not differ too greatly from the full selection, and thus we do not incorrectly reject a genuine jet.

For the  $t\bar{t}+\gamma$  analysis, we create a side-band region by inverting the shower-shape selection cut, thus giving a cut of  $0.12 < \sigma_{in\eta} < 0.20$  where all other photon ID selection requirements still remain the same. When considering a side-band variable, we choose a variable with little to no correlation with our template variable in order to optimise the performance of the method. The shape of the side-band template will then closely resemble the isolation distribution for photons that are not prompt but pass the photon ID requirements.

We calculate the shower shape of the photon by measuring energy deposits from within a  $5 \times 5$  matrix of crystals, centred around the crystal with the largest energy in the super-cluster. As the super-cluster footprint-removal technique removes the super-cluster from the isolation sum, then we have no shared input between the two variables (shower shape and isolation) effectively removing any correlation between the two. However, no two variables are ever truly uncorrelated, and any residual correlation can be attributed to the properties of hadronisation of jets. As our main source of fake photons arises in the form of jets misidentified as photons, we use the super-cluster footprint-removed charged hadron isolation as our background template variable to discriminate against the fake photons.

Talk about agreement between data and simulation.

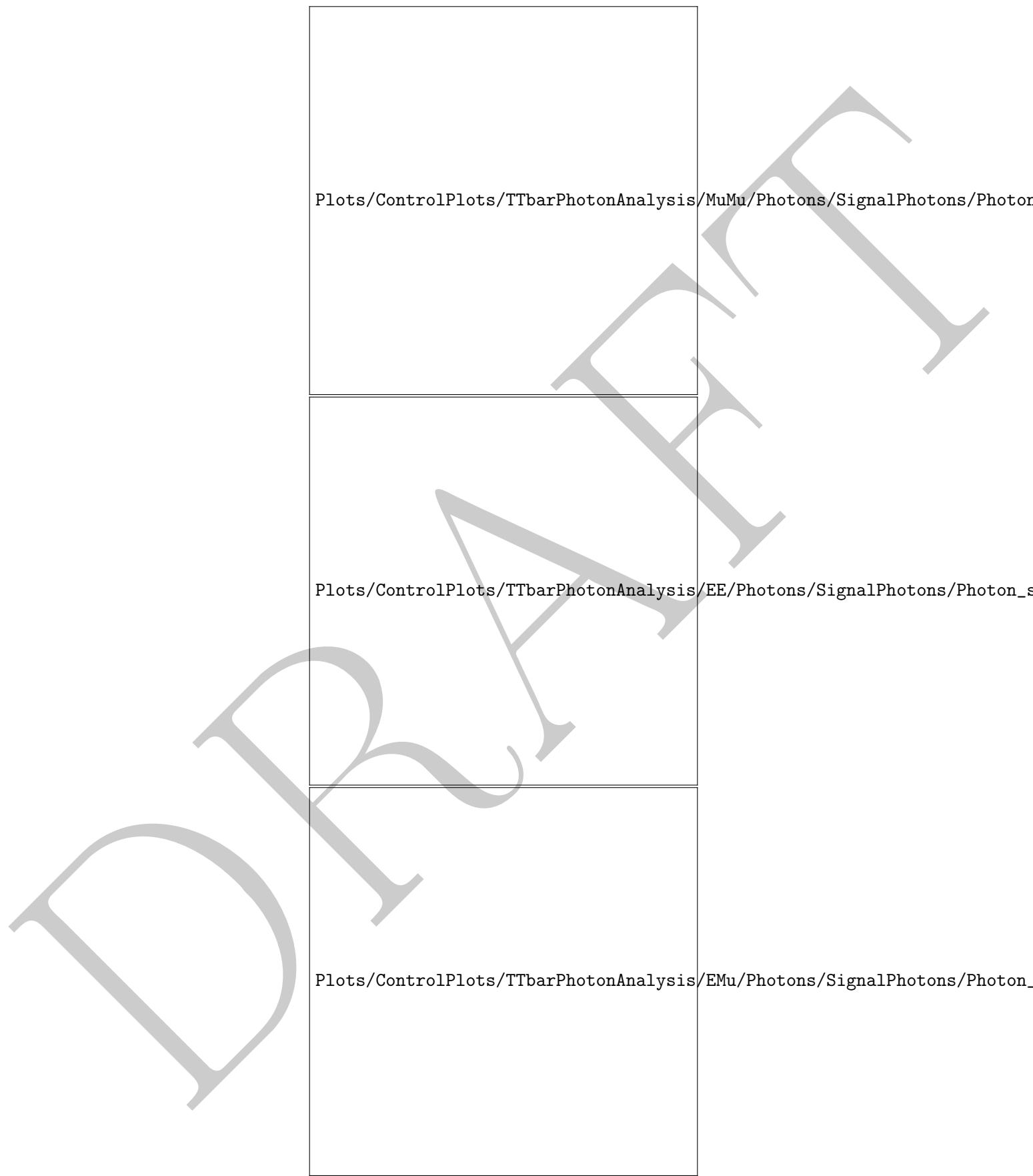


Figure 4.4

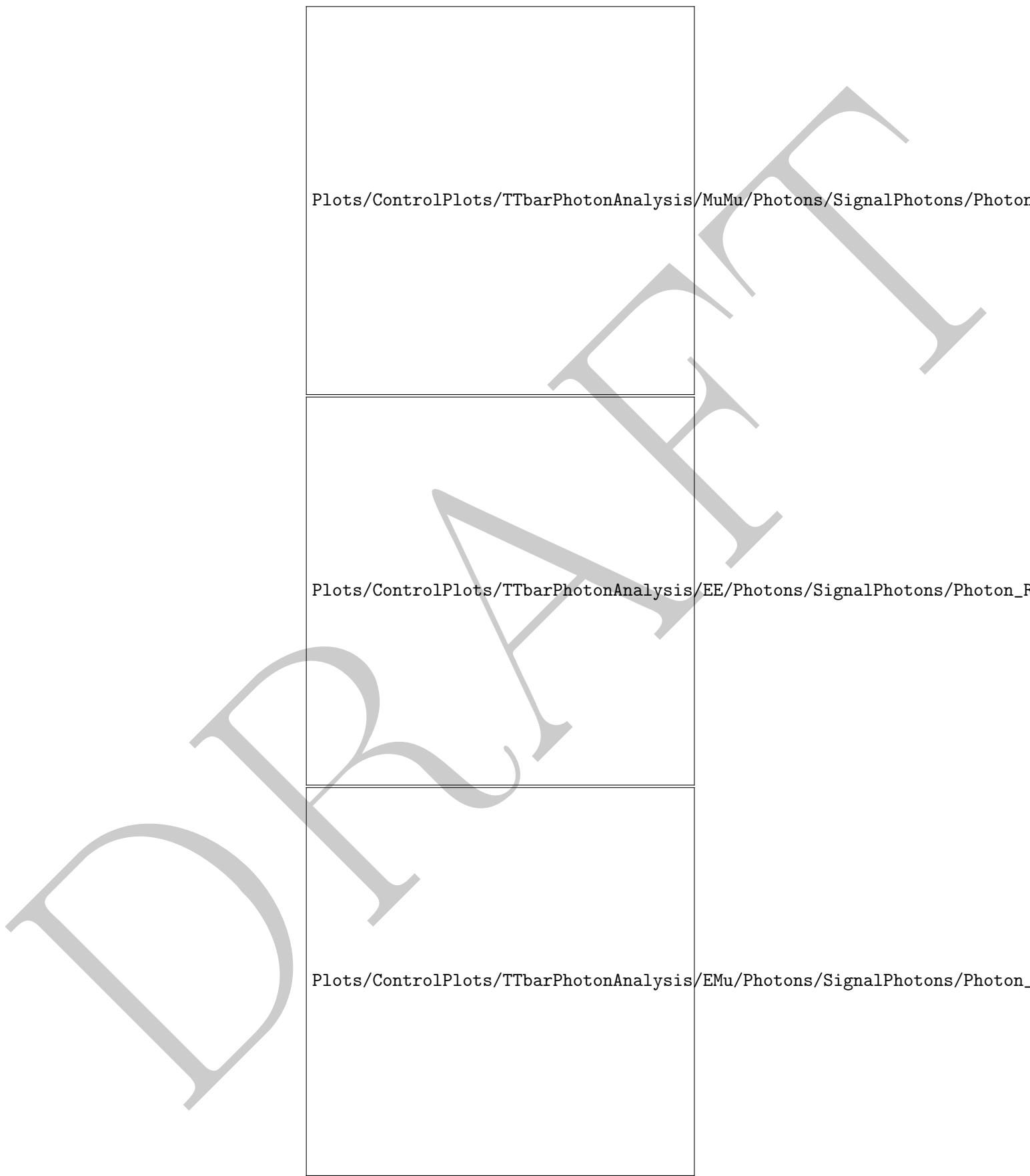


Figure 4.5

Cut Flow table for the $\mu^+\mu^-$ selection											
	ttgamma	ttbar 01	ttbar 11	tbar 21	wjets	z_jets	diboson	single t	qcd	all MC	data
Skim	3650 ± 12	18015 ± 38	141668 ± 117	147272 ± 84	213643 ± 1763	1318333 ± 1868	16881 ± 33	33315 ± 365	21035073 ± 153027	22928349 ± 153049	2413504 ± 1554
Cleaning and HLT	1018 ± 6	3117 ± 16	36105 ± 59	46211 ± 47	10348 ± 387	624329 ± 1273	5977 ± 17	8049 ± 176	1737634 ± 21436	2472789 ± 21478	1887786 ± 1374
Di-lepton Sel	292 ± 3	0 ± 0	149 ± 4	21477 ± 31	47 ± 25	422524 ± 1027	4348 ± 14	1186 ± 24	2437 ± 605	452461 ± 1193	459842 ± 678
m(Z) veto	262 ± 3	0 ± 0	126 ± 3	19169 ± 30	47 ± 25	82812 ± 489	1047 ± 8	1062 ± 23	2001 ± 559	106525 ± 744	114102 ± 338
≥ 1 jet	241 ± 3	0 ± 0	104 ± 3	18574 ± 29	32 ± 20	61503 ± 431	915 ± 8	1001 ± 22	368 ± 261	82738 ± 506	8892 ± 298
≥ 2 jets	210 ± 3	0 ± 0	64 ± 2	17585 ± 28	32 ± 20	47551 ± 379	772 ± 7	878 ± 21	368 ± 261	67462 ± 462	71800 ± 268
/ $E_T$ cut	196 ± 3	0 ± 0	59 ± 2	16468 ± 27	32 ± 20	21441 ± 253	483 ± 6	825 ± 20	0 ± 0	339505 ± 256	42953 ± 206
≥ 1 CSV b-tag	179 ± 3	0 ± 0	48 ± 2	15127 ± 26	32 ± 20	8246 ± 165	207 ± 4	721 ± 19	0 ± 0	24559 ± 169	26478 ± 163
≥ 1 Photon Presel	173 ± 3	0 ± 0	37 ± 2	11099 ± 22	23 ± 17	6217 ± 150	151 ± 4	497 ± 15	0 ± 0	18198 ± 154	19105 ± 138
≥ 1 Good Photon Postsel	17 ± 1	0 ± 0	0 ± 0	12 ± 1	0 ± 0	12 ± 6	1 ± 0	0 ± 0	0 ± 0	41 ± 6	46 ± 7
1 Good Photon Postsel	17 ± 1	0 ± 0	0 ± 0	11 ± 1	0 ± 0	13 ± 7	1 ± 0	0 ± 0	0 ± 0	43 ± 7	46 ± 7

Table 4.2: The number of expected events in MC and events observed in data for the  $\mu^+\mu^-$  channel, before the fitting process, including statistical uncertainties.

	Cut Flow table for the $e^+e^-$ selection										
	ttgamma	ttbar 01	ttbar 11	tbar 21	wjets	z_jets	diboson	single t	qcd	all MC	data
Skim	3632 ± 12	18015 ± 38	141660 ± 117	146469 ± 83	213610 ± 1763	1295804 ± 1838	16644 ± 33	33257 ± 365	21035139 ± 153027	22904230 ± 153048	1137895 ± 1067
Cleaning and HLT	533 ± 5	142 ± 3	4302 ± 20	25215 ± 34	10504 ± 390	438516 ± 1038	4631 ± 15	2072 ± 66	692073 ± 19655	692073 ± 19655	817968 ± 904
Di-lepton Sel	275 ± 3	0 ± 0	109 ± 3	17942 ± 28	299 ± 59	303836 ± 837	3389 ± 12	974 ± 22	327080 ± 860	327080 ± 860	384642 ± 620
m(Z) veto	247 ± 3	0 ± 0	93 ± 3	16018 ± 26	248 ± 54	66531 ± 418	852 ± 7	874 ± 21	257 ± 182	85120 ± 460	106244 ± 326
≥ 1 jet	220 ± 3	0 ± 0	85 ± 3	15296 ± 26	166 ± 44	51367 ± 375	757 ± 7	838 ± 21	257 ± 182	68987 ± 420	83739 ± 289
≥ 2 jets	188 ± 3	0 ± 0	72 ± 2	14501 ± 25	97 ± 34	40075 ± 332	654 ± 6	754 ± 19	137 ± 137	56478 ± 362	67769 ± 260
/ $E_T$ cut	175 ± 3	0 ± 0	65 ± 2	13596 ± 24	81 ± 31	18112 ± 220	407 ± 5	701 ± 18	137 ± 137	33274 ± 263	39949 ± 200
≥ 1 CSV b-tag	161 ± 2	0 ± 0	55 ± 2	12487 ± 23	29 ± 21	6826 ± 142	170 ± 3	601 ± 17	134 ± 134	20463 ± 198	24449 ± 156
≥ 1 Photon Presel	157 ± 2	0 ± 0	53 ± 2	12213 ± 23	34 ± 24	7433 ± 156	182 ± 4	585 ± 16	131 ± 131	20790 ± 207	24449 ± 156
≥ 1 Good Photon Postsel	15 ± 1	0 ± 0	0 ± 0	9 ± 1	0 ± 0	12 ± 5	0 ± 0	1 ± 1	0 ± 0	37 ± 5	34 ± 6
1 Good Photon Postsel	15 ± 1	0 ± 0	0 ± 0	9 ± 1	0 ± 0	12 ± 5	1 ± 0	1 ± 0	0 ± 0	37 ± 5	34 ± 6

Table 4.3: The number of expected events in MC and events observed in data for the  $e^+e^-$  channel, before the fitting process, including statistical uncertainties.

Cut Flow table for the $e\mu$ selection										
	ttgamma	ttbar 01	ttbar 11	ttbar 21	wjets	z_jets	diboson	single t	qcd	all MC
Skim	3610 ± 12	18015 ± 38	141648 ± 117	144829 ± 82	213614 ± 1763	1334209 ± 1889	16963 ± 33	33168 ± 365	21035027 ± 153027	22941082 ± 153049
Cleaning and HLT	1407 ± 8	1147 ± 10	34466 ± 58	63666 ± 54	22308 ± 570	28157 ± 277	2018 ± 14	8624 ± 167	1410492 ± 34863	1572284 ± 34869
Di-lepton Sel	560 ± 5	0 ± 0	261 ± 5	38902 ± 41	313 ± 62	2776 ± 80	852 ± 9	2114 ± 35	1301 ± 406	47079 ± 422
≥ 1 jet	508 ± 4	0 ± 0	226 ± 4	37416 ± 41	230 ± 53	2253 ± 73	713 ± 8	2009 ± 35	145 ± 131	43499 ± 168
≥ 2 jets	440 ± 4	0 ± 0	172 ± 4	35445 ± 40	157 ± 43	1676 ± 63	572 ± 8	1767 ± 29	0 ± 0	40230 ± 91
≥ 1 CSV b-tag	404 ± 4	0 ± 0	142 ± 3	32369 ± 38	90 ± 34	622 ± 39	219 ± 5	1529 ± 27	0 ± 0	38657 ± 197
≥ 1 Photon Presel	395 ± 4	0 ± 0	139 ± 3	31855 ± 37	98 ± 37	676 ± 43	239 ± 5	1490 ± 26	0 ± 0	34891 ± 73
≥ 1 Good Photon Postsel	40 ± 1	0 ± 0	0 ± 0	23 ± 1	0 ± 0	0 ± 0	0 ± 0	1 ± 1	0 ± 0	38657 ± 197
1 Good Photon Postsel	39 ± 1	0 ± 0	0 ± 0	22 ± 1	0 ± 0	0 ± 0	0 ± 0	1 ± 1	0 ± 0	42 ± 6

Table 4.4: The number of expected events in MC and events observed in data for the  $e\mu$  channel, before the fitting process, including statistical uncertainties.

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 4.5: Dataset information for each run in each respective decay channel.

# 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

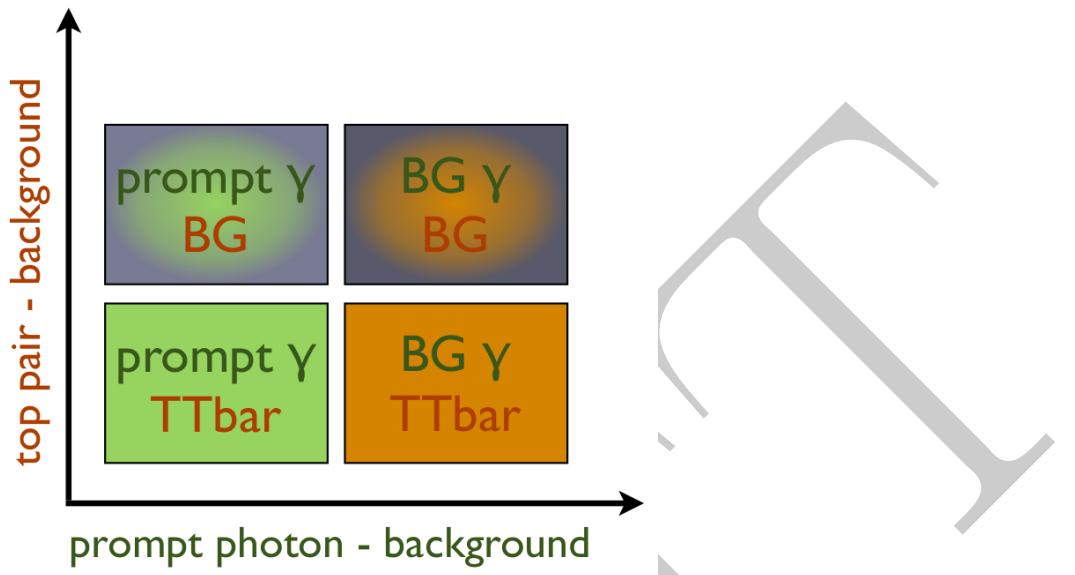


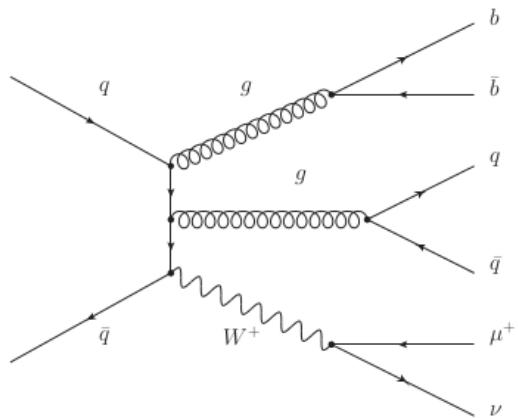
Figure 5.1: Graphic representation of the signal and background definitions [?].

least one generator-level photon with the following properties:

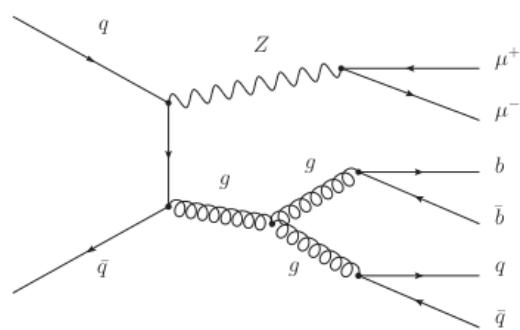
- $p_T(\gamma) > 13$  GeV
- $|\eta| < 3.0$
- Only gluons, bosons, or leptons are in the parents list. This ensures that photons from  $\pi^0$  decays are not considered as signal
- $\Delta R(\gamma, \text{other}) > 0.2$  where other particles include leptons, b quarks and final state particles (hadrons,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

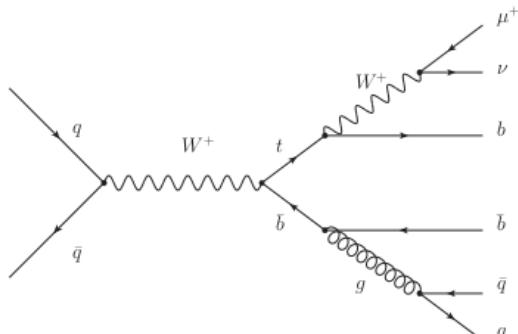
Similarly, we also observe an overlap between Z+Jets and ZGamma processes, and between W+Jets and WGamma samples, for the same reasons as described above. The phase space overlap removal procedure is applied on Z+Jets and W+Jets samples to remove events containing generator-level photons. Events containing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



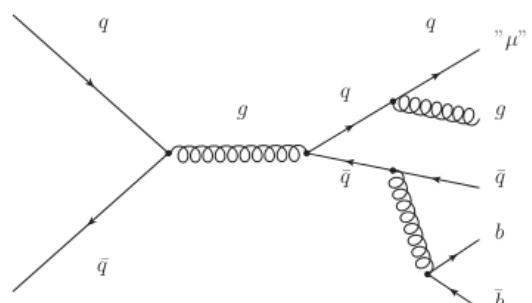
(a)  $W + \text{Jets}$



(b)  $DY + \text{Jets}$



(c) Single Top (s-channel)



(d) QCD Multijet

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 [16].

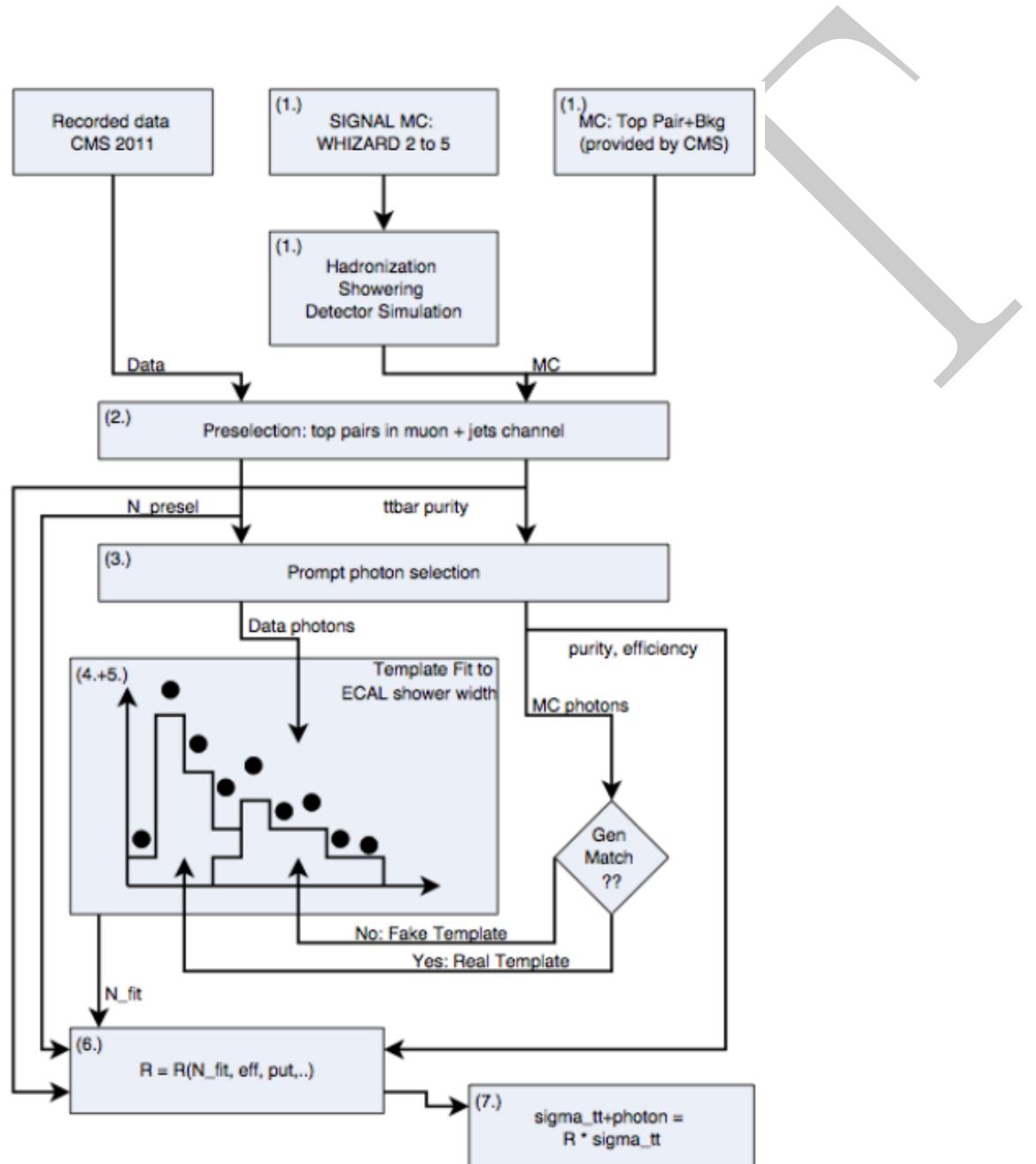


Figure 5.3: Flow chart showing each stage of the analysis. The box numbers represent the outlined analysis steps.

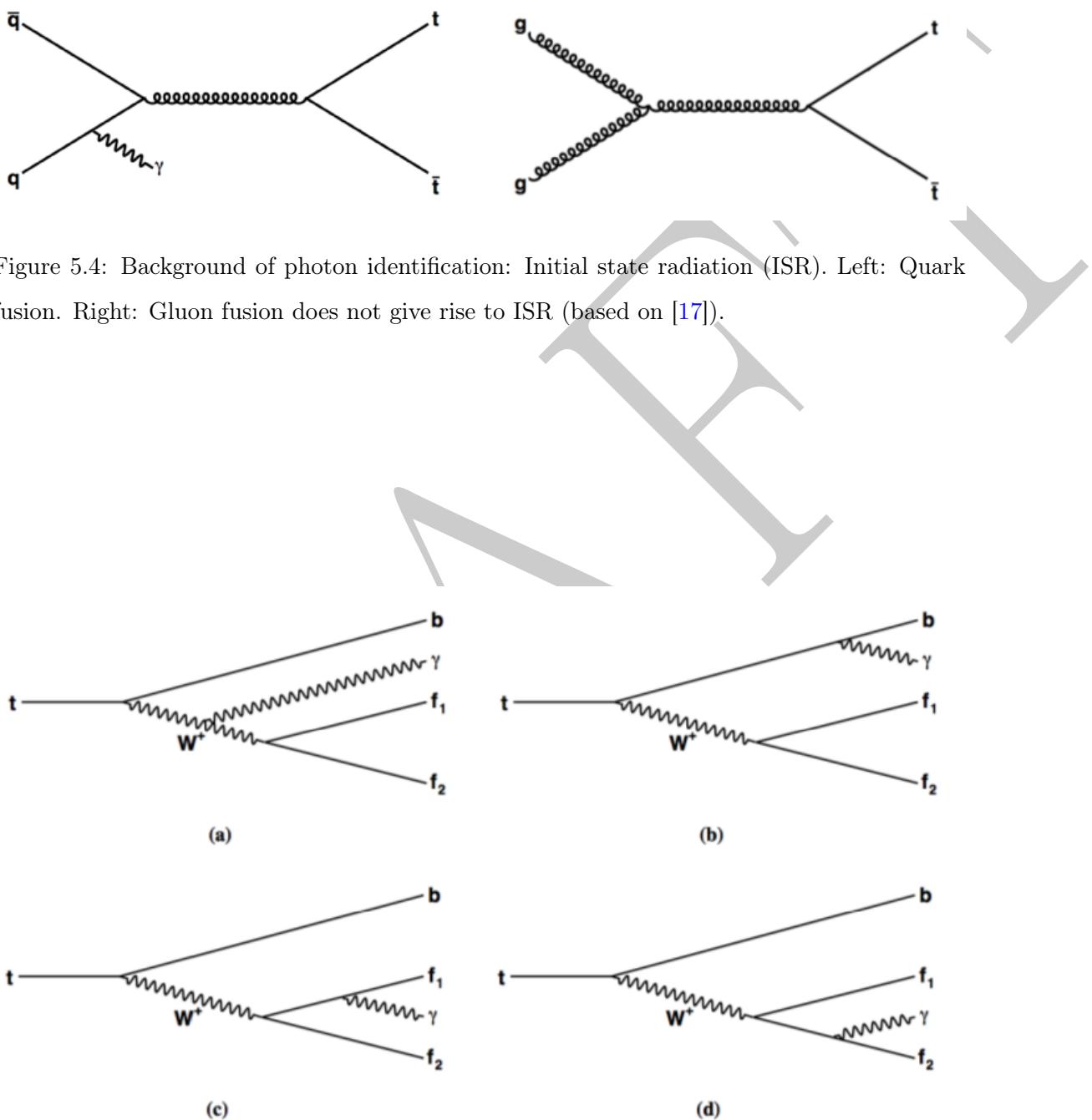
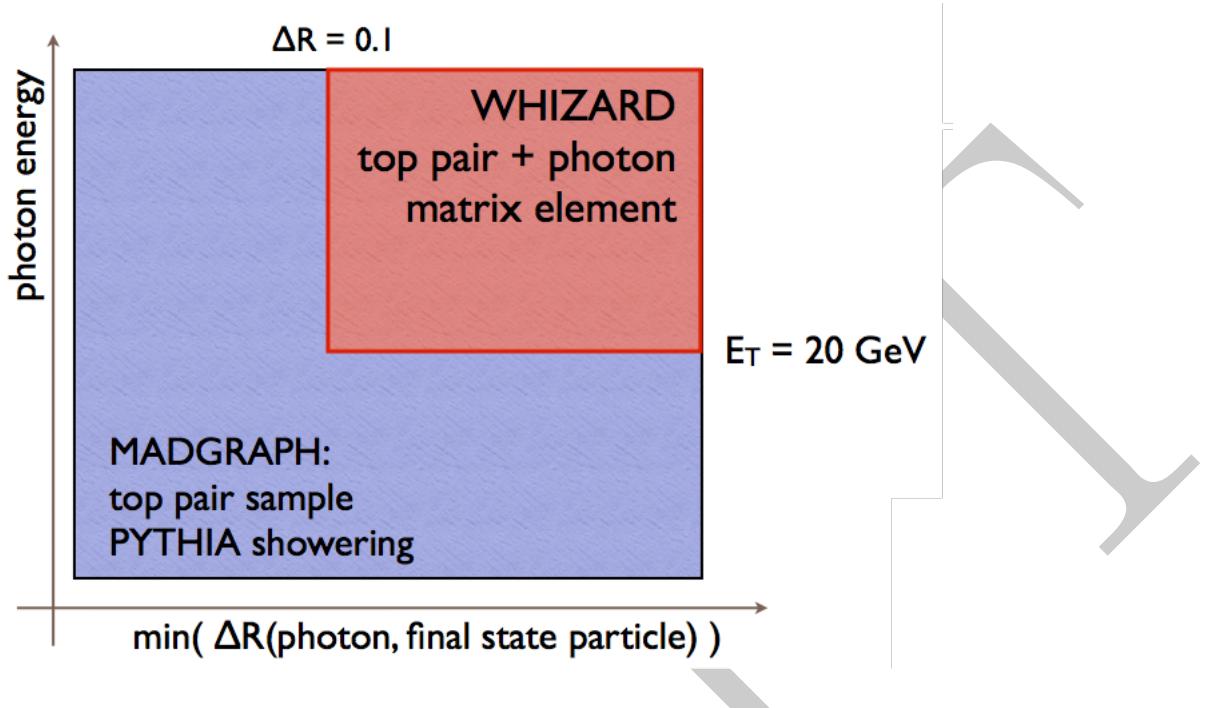


Figure 5.5: Background of photon identification: Final state radiation (FSR). All charged particles in the top-pair decay tree contribute to FSR [17].



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 Measurement of the cross-section ratio $R = \sigma_{t\bar{t}+\gamma}/\sigma_{t\bar{t}}$

A common form of cross-section calculation for a “counting” analysis can be written as

$$\sigma = \frac{N_{\text{signal}}}{\mathcal{L}\epsilon A} \quad (5.1)$$

where  $N_{\text{signal}}$  is the number of signal events found in data passing the full event selection requirements,  $\mathcal{L}$  is the integrated luminosity of the data,  $A$  is the signal acceptance, defined as the fraction of signal events that fall into the region of phase-space that is chosen for our event selection, and  $\epsilon$  is the efficiency of correctly selecting a signal event (the fraction of signal events passing event selection after acceptance cuts are implemented). The number of signal events,  $N_{\text{signal}}$ , must be modified to account for the presence of background events in our data, and thus becomes the number of observed events minus the number of background events,  $N_{\text{observed}} - N_{\text{background}}$ , where the number of background events is estimated in a

different manner.

Due to the nature of the analysis, event selection is performed in two separate steps (see Section 4); di-leptonic top pair selection and photon selection. This feature changes the way in which we must compute the cross-section in Equation 5.1. The changed cross-section is written as

$$\sigma_{t\bar{t}+\gamma} = \frac{N_{signal}}{\mathcal{L}\epsilon_{top}A_{top}\epsilon_{\gamma}A_{\gamma}} \quad (5.2)$$

where  $\epsilon_{top}$  and  $A_{top}$  are the efficiency and acceptance of top selection for our signal top pair events. Thus,  $\epsilon_{\gamma}$  and  $A_{\gamma}$  are the efficiency and acceptance of the photon selection, introduced after the top quark pair event selection. Here we consider our signal as a top quark pair, decaying to leptons and b-jets, plus a prompt radiated photon as our signal. We take the number of observed,  $N_{signal}$ , as the number of events counted from data after top pair and photon selections have been applied.

In order to simplify the calculation, we are able to take the ratio of the  $t\bar{t} + \gamma$  cross-section to the  $t\bar{t}$  cross-section, as previously measured by the CMS experiment [88]. In doing this, we simplify the treatment of systematic uncertainties and cancel shared terms from each measurement, such as luminosity. The new ratio is calculated using the following equation

$$R = \frac{\sigma_{t\bar{t}+\gamma}}{\sigma_{t\bar{t}}} = \frac{N_{signal}}{\epsilon_{top}^{t\bar{t}+\gamma} A_{top}^{t\bar{t}+\gamma} \epsilon_{\gamma}^{t\bar{t}+\gamma} A_{\gamma}^{t\bar{t}+\gamma}} \cdot \frac{\epsilon_{top}^{t\bar{t}} A_{top}^{t\bar{t}}}{N_{t\bar{t}}} \quad (5.3)$$

We now introduce a second term into the cross-section calculation which includes three new variables that stem from the  $t\bar{t}$  inclusive cross-section: where  $\epsilon_{top}^{t\bar{t}}$  and  $A_{top}^{t\bar{t}}$   $N_{t\bar{t}}$  are the efficiency and acceptance of the  $t\bar{t}$  process, and  $N_{t\bar{t}}$  is the number of top pair events passing the full top selection, prior to the photon selection. The efficiencies for both top pair selection and top pair with a photon selection is very similar due to the ordering of the event selection - the selection of leptons, jets, b-jets, and MET are the same in each scenario. As the acceptance calculation depends on individual samples generator cuts and even selection, we must calculate this variable separately for each stage of the selection process. By taking the ratio of each cross-section we cancel out the luminosity parameter.

The prominent feature of this analysis is the addition of a radiated photon from one of the decay products of a di-leptonic  $t\bar{t}$  decay. In order to correctly measure the number of signal  $t\bar{t} + \gamma$  events we must make sure that we have correctly identified our signal events. Even though an object may pass the full photon ID selection requirements, it may not be our signal photon. Electrons and jets are likely to pass photon ID selection and be recorded

as real signal photons, thus we must remove these ‘fakes’ from our signal events.

For jets that pass the photon ID requirements and are thus misidentified, we must note that they do not contain a genuine photon. Electrons that have been misidentified as photons are more difficult to remove as they hold the same properties as a photon, including isolated electromagnetic shower etc. Electrons differ from our signal photons because of the addition of a charged track pointing in the direction of the energy deposit. As a result, we treat each source of ‘fakes’ in a different manner, described in Section ??.

## 5.5 Number of top pair events

## 5.6 Photon Purity Estimation

We define the photon purity,  $\pi$ , as the number of photons that are prompt and pass the photon ID selection over the total number of photon candidates. Prompt photons are created in the hard-scattering process and are emitted from the high- $p_T$  hard-scattering particles. The main sources of non-prompt photons come from charged particles interacting with the detector material and from quark hadronisation, and are generally not isolated and have a low  $p_T$ . We also obtain a source of photons from  $\pi^0$  decays, whereby two very collinear photons are produced.

One method to find the photon purity is through simulation, where we have access to information of all the particles simulated by the MC event generator. We match reconstructed photons to generator photons by using information of their mother particles in order to determine the origin of the photons. From this, we can divide our events into three categories based on the event generator matching to a reconstructed photon.

- Signal: reconstructed photon is matched to a generated photon
- Misidentified electron: reconstructed photon is matched to a generated electron
- Misidentified jet: photon is not matched to either a generated photon or electron

In order to find if our photon is prompt (signal), we use our matching process to find a generator-level photon that matches its properties. We use the transverse momentum along with the  $\eta$  and  $\phi$  coordinates of the reconstructed and generator-level photon to implement a  $p_T$  and  $\Delta R$  requirement on the matching process. Events are iterated over in the order in which they were created by the generator, and if we find the  $\Delta R$  between the reconstructed

photon and generated photon is **less than 0.2** and  $|p_T^{reco} - p_T^{gen}|/p_T^{gen}$  is **less than 1.0** then we consider it a match and stop the iteration process. The parentage of the matched photon is then inspected to verify its origin, such that if it descends from a quark, gluon, charged lepton, or boson then we determine it to be prompt. We must, however, also verify that the photon is isolated. A photon maybe produced in the hard-scattering process, however we may observe hadronisation and showering in close by, such that our signal photon will not be isolated. In order to ensure that these photons are not lost to the nearby non-signal activity, we must introduce additional selection cuts to differentiate between our prompt signal photons and non-isolated photons.

- $(p_T^{reco} - p_T^{gen}) / p_T^{gen} < 0.1$
- $\Delta R(\gamma^{gen}, \text{other}) > 0.2$  where other particles include leptons, photons and final state particles (hadrons, leptons, photons) with transverse momenta above 5 GeV
- $|\Delta\eta(\gamma^{reco}, \gamma^{gen})| < 0.005$
- $\Delta R(\gamma^{reco}, \gamma^{gen}) < 0.01$

The listed additional cuts help to select photons that are well isolated and do not have any undesired activity nearby.

As described above, electrons leave a very similar trace to photons as it showers within the electromagnetic calorimeter, except for an additional charged track pointing towards the energy deposit. We implement specific selection criteria in order to minimise the fake-rate of electrons misidentified as photons within the event selection. However, it is observed that a large number of electrons still pass photon selection requirements and are recorded as signal. We thus include a criteria to find well isolated generator-level electrons in the same manner as for photons, however in this scenario we consider electrons from W and Z decays only. The cuts are shown below.

- $(p_T^{reco} - p_T^{gen}) / p_T^{gen} < 0.1$
- $\Delta R(e, \text{other}) > 0.2$  where other particles include leptons, photons and final state particles (hadrons, leptons, photons) with transverse momenta above 5 GeV
- $|\Delta\eta(\gamma^{reco}, e)| < 0.005$
- $\Delta R(\gamma^{reco}, e) < 0.04$

The cuts have been cross-checked with the di-leptonic  $t\bar{t}$  sample as it contains a significant amount of electrons identified as photons.

All reconstructed photons that pass the cuts described previously and is matched with a generator-level photon is then defined as a real photon. If it is instead matched to a generator-level electron it is then defined as a misidentified electron. Any other event that is not matched to a generator-level electron or photon is then classified as a misidentified jet.

As the photon purity directly affects the cross-section ratio measurement, we do not want to rely explicitly on simulation, and therefore it would be preferable to compute the photon purity via a data-driven method. We make use of the fact that our signal photon, whether a real signal photon or misidentified electron, are isolated objects, while misidentified jets and non-prompt photons close to or within a jet.

[89]

## 5.7 Number of signal events in data

## 5.8 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:

$$\epsilon_{top}^{t\bar{t}} \cdot A_{top}^{t\bar{t}} = \frac{N_{t\bar{t}.preselection}}{N_{t\bar{t}.generated}} \quad (5.4)$$

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 [90]. 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:

$$\epsilon_{top}^{t\bar{t}+\gamma} \cdot A_{top}^{t\bar{t}+\gamma} = \frac{N_{t\bar{t}+\gamma.\text{preselection}(1\text{lor}2l)}}{N_{t\bar{t}+\gamma.\text{generated}(1\text{lor}2l)}} \quad (5.5)$$

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

$$\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \frac{N_{t\bar{t}+\gamma.\text{photonselection}(1\text{lor}2l)}}{N_{t\bar{t}+\gamma.\text{preselection}(1\text{lor}2l)}} \quad (5.6)$$

Thus, yields of the acceptance times efficiency of the inclusive  $t\bar{t} + \gamma$  process to be  $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \pm(\text{stat.})$  in the di-muon channel,  $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \pm(\text{stat.})$  in the di-electron channel, and  $\epsilon_{\gamma}^{t\bar{t}+\gamma} \cdot A_{\gamma}^{t\bar{t}+\gamma} = \pm(\text{stat.})$  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  $\eta$  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  $|\eta| < 1.4442$ )

$$\epsilon_{top}^{t\bar{t}+\gamma Vis} = \frac{N_{t\bar{t}+\gamma.\text{preselection}(1\text{lor}2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma^{gen})| < 1.4442)}{N_{t\bar{t}+\gamma.\text{generated}(1\text{lor}2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma^{gen})| < 1.4442)} \quad (5.7)$$

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.

$$\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = \frac{N_{t\bar{t}+\gamma.\text{photonselection}(1l\text{or}2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma_{gen})| < 1.4442)}{N_{t\bar{t}+\gamma.\text{preselection}(1l\text{or}2l)}(p_T(\gamma_{gen}) > 25\text{GeV}, |\eta(\gamma_{gen})| < 1.4442)} \quad (5.8)$$

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,  $\epsilon_{\gamma}^{t\bar{t}+\gamma Vis} = 0.2860.004(\text{stat.})$  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 [91] [92]

## 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 [93].

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 [94]. The total calculated integrated luminosity for the entire 2012 dataset was measured to be  $23.27 \text{ fb}^{-1}$ .

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  $19.7 \text{ fb}^{-1}$  for the CMS experiment with the full 2012 dataset at, where a flat rate associated uncertainty of 2.6% is assigned [95]. 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

We calculate lepton efficiencies and associated uncertainties in order to correct for the number of leptons observed in data and those in simulation. In order to calculate lepton efficiencies, the tag-and-probe method is used [?]. The method analyses events from a Drell-Yan  $Z \rightarrow l^+l^-$  sample as it contains a large number of unbiased lepton-pair events with a high purity. Using this sample, the tag-and-probe technique selects lepton-pair events such that one of the leptons is defined as the “tag” lepton, which is selected with under much tighter requirements, and the second “probe” lepton is selected under much looser constraints relative to the “tag”. The “tag” lepton candidate satisfies the trigger criteria, tight identification, and isolation requirements. The “probe” lepton candidate is required to pass specific criteria depending on the efficiency under study. We thus create two subsets of the data, such that one contains events that pass the probe selection, and one collection that contains events that failed probe selection. We then take the efficiency of the selection to be the fraction of events that pass the probe selection criteria, defined as

$$\epsilon_{all} = \epsilon_{reco.} \epsilon_{tight} \epsilon_{trig.} \quad (6.1)$$

The tag-and-probe method is applied to electrons and muons separately, where it is applied to electrons in the barrel and endcap regions individually. The purity of the di-lepton sample for the tag-and-probe is held by requiring the invariant mass of the lepton pair to fall within the mass window of the  $Z$  boson,  $70 < M_{ll} < 130$  GeV/ $c^2$ . We then divide the total lepton efficiency into three sub-divisions: the trigger efficiency of identifying a lepton candidate, the efficiency for the reconstruction algorithms to reconstruct leptons, and the efficiency for the identification and isolation selection requirements to correctly select leptons.

In order to measure the trigger efficiencies for both electrons and muons, selected probes are required to pass normal kinematic cuts such that it must pass the HLT to be considered. We find that the trigger efficiency for muons is greater than 99% and for electrons is greater than 95%, with an associated uncertainty of the order of 4% and varies depending on the used trigger. The reconstruction efficiency,  $\epsilon_{reco.}$ , is defined as the efficiency that an ECAL supercluster seeds an ECAL-driven electron candidate that passes the probe selection requirement, and is relative to ECAL clusters within the ECAL acceptance. We define the probe to have a reconstructed energy greater than 10 GeV, such that the supercluster lies within range of the tracker system. At this energy, we find the reconstruction efficiency to be greater than

85%, and more than 99% with an energy threshold of 20 GeV [96].

For the case of muons, an initial “preselection” of Z events for the tag-and-probe method is obtained by selecting two oppositely charged tracks measured in the central tracker that each have a  $p_T > 25$  GeV,  $|\eta| < 2.1$ , and that when combined lie within the mass window of  $60 < m_{\mu^+\mu^-} < 120$  GeV. When defining a “tag” muon we require that it is matched to a preselected track, is a global and tracker muon, passes the selection described in Section 4, and corresponds to a HLT muon. All other preselected tracks are considered as probes and are used in order to measure the efficiency. An efficiency of approximately 95-99% is observed in data for all muon systems [97].

The tag-and-probe method is applied to both data and simulated samples, and thus we compute the efficiency for MC simulation ( $\epsilon_{sim.}$ ) and data ( $\epsilon_{data}$ ). We then compute the ratio of efficiencies along with the associated statistical and systematic uncertainties, given as:

$$\rho = \frac{\epsilon_{data}}{\epsilon_{sim.}} \quad (6.2)$$

where the efficiencies and the ratios of the efficiencies are estimated in bins of  $E_T$  and  $\eta$  of the electron. The efficiencies and associated statistical and systematic uncertainties are derived by the EGamma and Muon Physics Object Groups (POGs) for electrons and muons, respectively.

## 6.2 Shape Uncertainties

### 6.2.1 Parton Distribution Function

Parton Distribution 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 and thus 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 [98].

Event generators assign momentum fraction and energy to a parton based on PDFs which are calculated by taking data from various experiments, where each experiment has an associated uncertainty. The uncertainties must be propagated throughout the PDFs, therefore we

must further propagate into our final physics analysis. PDFs are updated by the collaborations who perform the fits, such as CTEQ [99], each time new data or theoretical predictions become available. The set of PDFs used in this analysis are taken from the CT10 [100] set. CT10 provides the nominal PDF weight along with 25 free parameters to describe the parton distributions, and thus 25 eigenvalues, providing 50 alternative weights per event. In order to access the weights, we use the Les Houches Accord Parton Distribution Function (LHAPDF) library [59].

We take the difference between each of the weights and the nominal and add in quadrature, where the final result is then used to calculate the systematic uncertainty associated with the PDFs.

### 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.

We take the number of primary vertices directly from the minimum bias data obtained over the running period in question. We then vary the minimum bias cross-section by  $\pm 5\%$  which then recalculates the primary vertex distributions which we then use to measure the total impact of the pile-up on the analysis when more, or less, pile-up is observed within the data.

### 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  $\eta$  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 [78, 101].

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,  $\sigma$ , 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 [102]. 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  $\eta$ . 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 ( $\rho$ -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  $\mu\mu$  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 multi-jet samples [103]. 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 Re-weighting

## 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
Statistical	~1	~1
Systematic	~1	~1
Pileup (PU)	~1	~1
Out-of-time Pileup (OOT)	~1	~1
Top $P_T$	~1	~1
b-tag	~1	~1
Photon $E_T$	~1	~1
JEC	~1	~1
JER	~1	~1
Electron Efficiency	~1	~1
Electron $P_T$	~1	~1
PDF	~1	~1
Total	~1	~1

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-	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/ tauola/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	1985529 2112387 2170270 1934901
	Single Top tW scale up Single Top tW scale down Single TopBar $\bar{t}W$ scale up Single TopBar $\bar{t}W$ scale down W+Jets matching up W+Jets matching down W+Jets scale up	/TToDilepton_tW-channel-DR_scaleup_8TeV-powheg- tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM /TToDilepton_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-	1492816 497658 1492534 1493101 21364637 21364637 20784770

# Chapter 7

## Results

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

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  $\eta$  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:

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

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

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

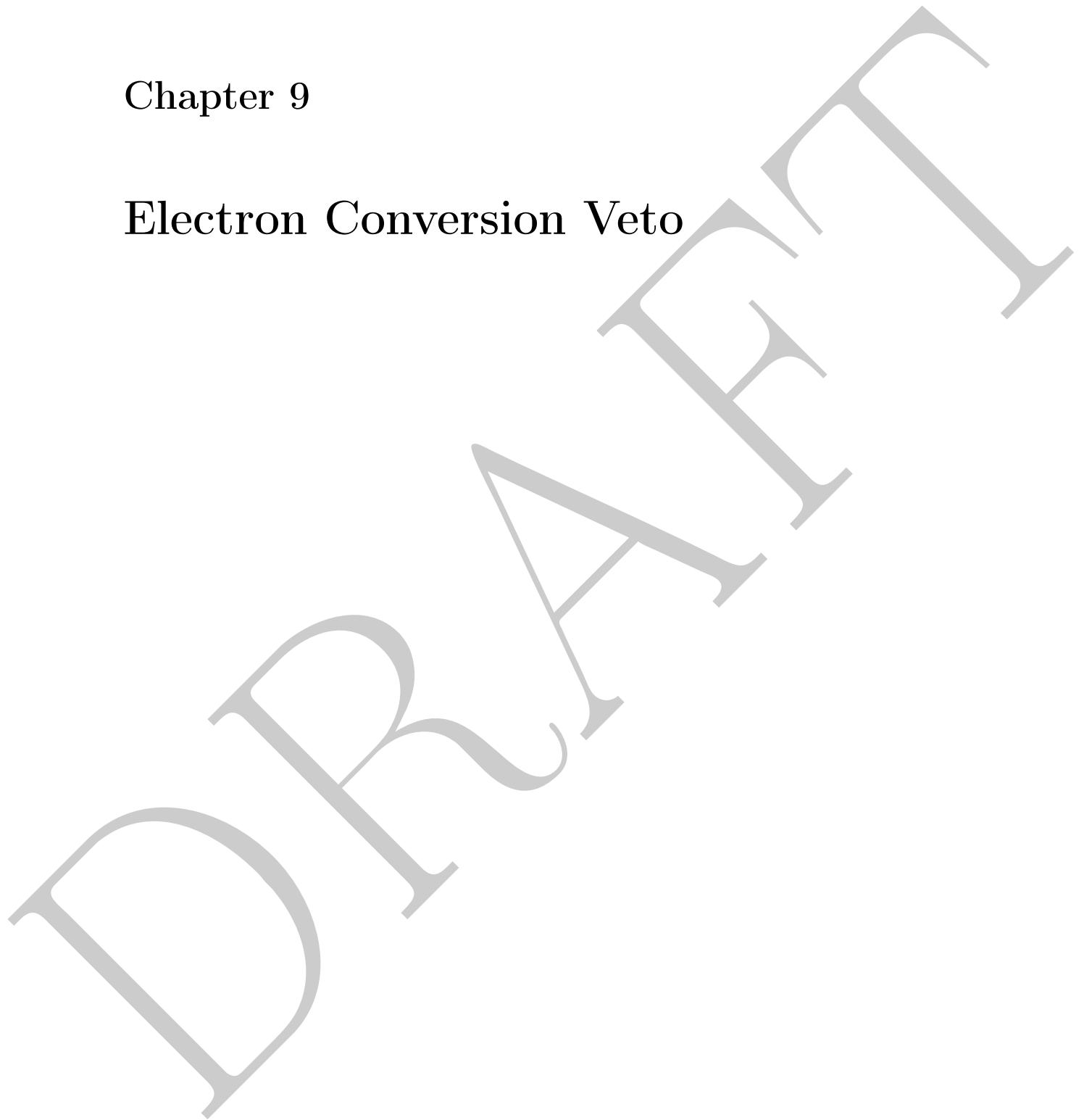
## Chapter 8

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

[90] [104]

## 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$	$0.969 \pm 0.007$	$0.926 \pm 0.003$	$0.969 \pm 0.002$	$0.975 \pm 0.000$
$0.8 <  \eta  < 1.4442$	$0.935 \pm 0.017$	$0.945 \pm 0.004$	$0.964 \pm 0.002$	$0.974 \pm 0.002$
$1.4442 <  \eta  < 1.5660$	$1.032 \pm 0.039$	$0.907 \pm 0.015$	$0.957 \pm 0.022$	$0.877 \pm 0.020$
$1.5660 <  \eta  < 2.5$	$0.919 \pm 0.014$	$0.926 \pm 0.005$	$0.952 \pm 0.003$	$0.950 \pm 0.005$

Table A.1: Electron ID & isolation efficiencies [22].

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

Table A.1: Muon ID & isolation efficiencies [23].

## A.1 Electron & Muon Efficiencies

# Bibliography

- [1] Sergey Senkin. Top quark physics at the lhc with the cms detector. *PhD Thesis*, page 204, 04/2014.
- [2] Thomas Mueller. New results from the top quark. 2012.
- [3] Alemany-Fernandez et al. Operation and configuration of the lhc in run 1. *CERN Internal Note*, 25 Nov 2013.
- [4] CMS Collaboration. The cms experiment at the cern lhc. *JINST*, 3:S08004, 2008.
- [5] CMS Collaboration. Tracking results. (2012).
- [6] Cristina Biino. The cms electromagnetic calorimeter: overview, lessons learned during run 1 and future projections. *Journal of Physics: Conference Series*, 587(1):012001, 2015.
- [7] Eunhyang Kwon Jiwoong Seo Junghwan Goh, Min Suk Kim and Intae Yu. Muon reconstruction and identification in the rpc detector with tracker tracks. *CMS Internal Note*, v2, 2013.
- [8] CMS Collaboration. The cms trigger. *CMS PAPER*, v10, 2015.
- [9] Stefan HÃ¡jche. Introduction to monte carlo event generators. *Slides from 2012 CTEQ-Fermilab School on QCD and Electroweak Phenomenology*, 2012.
- [10] Andrea Giannanco. The fast simulation of the cms experiment. *Journal of Physics: Conference Series*, 513(2):022012, 2014.
- [11] 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.

- [12] Mikhail Makouski Nik Berry Till Arndt Heiner Tholen Heiko Geenen, Yurii Maravin. Monte carlo simulation of the top quark pair plus photon production. *CMS Note*, 2013/382, 2013.
- [13] 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.
- [14] 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.
- [15] Marco Peruzzi. Measurement of differential cross sections for diphoton production in pp collisions with the cms experiment. *CMS Thesis*, 2015.
- [16] Thomas Steeger. Background estimation in a search for  $t\bar{t} + \gamma$  events. *Bachelor Thesis*, August, 2012.
- [17] 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.
- [18] MichaÅĆ Czakon, Paul Fiedler, and Alexander Mitov. Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through  $O(\frac{4}{S})$ . *Phys. Rev. Lett.*, 110:252004, 2013.
- [19] Oliver Sim et al. BrÃijning. Lhc design report. *Geneva : CERN, v.1 : the LHC Main Ring:548*, 2004.
- [20] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.
- [21] CMS Collaboration. Precise mapping of the magnetic field in the cms barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021, 2010.
- [22] CMS Collaboration. Electron efficiencies for top quark analysis. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/KoPFAElectronTagAndProbe>.
- [23] CMS Collaboration. Reference muon id and isolation efficiencies. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs>.

- [24] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.
- [25] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.
- [26] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [27] Abdus Salam and John Clive Ward. Weak and electromagnetic interactions. *Nuovo Cim.*, 11:568–577, 1959.
- [28] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.
- [29] David J. Gross and Frank Wilczek. Asymptotically free gauge theories. i. *Phys. Rev. D*, 8:3633–3652, Nov 1973.
- [30] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973.
- [31] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [32] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [33] John P. Blewett. 200-gev intersecting storage accelerators. *Proceedings of The 8th International Conference on High-Energy Accelerators. CERN. Geneva, Switzerland*, 1971.
- [34] 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.
- [35] 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.
- [36] L. Borrello et al. Sensor design for the cms silicon strip tracker. *CMS-NOTE-2003-020*.
- [37] 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).

- [38] V. Veszpreni. Operation and performance of the cms tracker. *Journal of Instrumentation*, 9(03):C03005, 2014.
- [39] L. Rossi et al. Pixel detectors. *Springer-Verlag, Heidelberg Germany*, (2006).
- [40] CMS Collaboration. Letter of intent. *CERN-LHCC-92-003*, 1992.
- [41] CMS Collaboration. The e.m. calorimeter project: Technical design report. *CERN-LHCC-97-33*, 1997.
- [42] 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.
- [43] CMS Collaboration. CMS technical design report, volume II: Physics performance. *J. Phys. G*, 34:995–1579, 2007.
- [44] Serguei Chatrchyan et al. The performance of the CMS muon detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV at the LHC. *JINST*, 8:P11002, 2013.
- [45] Serguei Chatrchyan et al. Measurement of the  $t\bar{t}$  production cross section in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with lepton + jets final states. *Phys. Lett.*, B720:83–104, 2013.
- [46] CMS Collaboration. Top pair cross section in e/mu+jets at 8 tev. *CMS Physics Analysis Summary*, CMS-PAS-TOP-12-006, 2012.
- [47] 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.
- [48] 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.
- [49] CMS Collaboration. Cmssw software framework.
- [50] GEANT4 Collaboration. Geant4: A simulation toolkit. *Nucl.Instrum.Meth.*, A506:250, 2003.
- [51] et al. D. Acosta, CMS Collaboration. Cms physics: Technical design report, volume 1: Detector performance and software. *CERN/LHCC, Geneva, Switzerland*, 2, 2006.

- [52] T. Ohl W. Kilian and J. Reuter. Whizard: Simulating multi-particle processes at lhc and ilc. *Eur.Phys.J.*, C71:1742, 2011.
- [53] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), 2011.
- [54] 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.
- [55] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.
- [56] Stefano Frixione and Bryan R. Webber. Matching nlo qcd computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029, 2002.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [61] Z. Was. Precision simulations with TAUOLA and PHOTOS. *Nucl. Phys. Proc. Suppl.*, 169:16–21, 2007. [,16(2006)].
- [62] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009, 2014.
- [63] Pierre Billoir. Progressive track recognition with a kalman-like fitting procedure. *Computer Physics Communications*, 57(1âS3):390 – 394, 1989.

- [64] W Erdmann. Vertex reconstruction at the cms experiment. *Journal of Physics: Conference Series*, 110(9):092009, 2008.
- [65] W Adam, R FrÃijhwirth, 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.
- [66] 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.
- [67] 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(2â€¢3):577 – 584, 2005.
- [68] CMS Collaboration. Cut-based electron identification algorithm. *CMS TWiki*, 2015.
- [69] CMS Collaboration. Cuts in categories (cic) electron identification. *CMS TWiki*, 2015.
- [70] CMS Collaboration. Particle-flow event reconstruction in cms and performance for jets, taus, and met. *Technical Report*, CMS-PAS-PFT-09-001, CERN, Geneva, Switzerland, 2009.
- [71] 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.
- [72] The CMS collaboration. Performance of cms muon reconstruction in pp collision events at  $\sqrt{s} = 7$  tev. *Journal of Instrumentation*, 7(10):P10002, 2012.
- [73] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.
- [74] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.
- [75] 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.

- [76] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- $k_t$  jet clustering algorithm. *JHEP*, 04:063, 2008.
- [77] 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.
- [78] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011.
- [79] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
- [80] Christian Weiser. A combined secondary vertex based b-tagging algorithm in cms. *Technical Report*, CMS-NOTE-2006-014, CERN, Geneva, Jan 2006.
- [81] 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.
- [82] The CMS collaboration. Missing transverse energy performance of the cms detector. *Journal of Instrumentation*, 6(09):P09001, 2011.
- [83] CMS Collaboration. Top reference selections and recommendations (run1) twiki. *CMS TWiki*.
- [84] CMS Collaboration. Cms top egamma coordination (run1). *CMS TWiki*.
- [85] CMS Collaboration. Top muon information for analysis (run1). *CMS TWiki*.
- [86] CMS Collaboration. Jet performance in pp collisions at  $\sqrt{s}=7$  TeV. CMS Physics Analysis Summary CMS-PAS-JME-10-003, 2010.
- [87] CMS Collaboration. Simple cut based id 2012. *CMS Twiki*.
- [88] CMS Colaboration. Top pair cross section in dileptons. *CMS Physics Analysis Summary*, TOP-12-007, 2012.
- [89] CMS Collaboration. Measurement of the  $w\gamma$  and  $z\gamma$  inclusive cross sections in  $pp$  collisions at  $\sqrt{s} = 7$  TeV and limits on anomalous triple gauge boson couplings. *Phys. Rev. D*, 89:092005, May 2014.

- [90] 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.
- [91] Simon White. Luminosity scans at the lhc. luminosity scans at lhc. 2011.
- [92] Cms luminosity based on pixel cluster counting - summer 2013 update. Technical Report CMS-PAS-LUM-13-00,, CERN, Geneva, 2013.
- [93] CMS Collaboration. Absolute calibration of the luminosity measurement at cms: Winter 2012 update. 2012.
- [94] Simon White. Luminosity Scans at the LHC. Luminosity Scans at LHC. 2011.
- [95] CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.
- [96] Vardan Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV. *JINST*, 10(06):P06005, 2015.
- [97] The CMS collaboration. The performance of the cms muon detector in proton-proton collisions at  $\sqrt{s} = 7$  tev at the lhc. *Journal of Instrumentation*, 8(11):P11002, 2013.
- [98] 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.
- [99] 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.
- [100] 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.
- [101] 8 TeV Jet Energy Corrections and Uncertainties based on  $19.8 \text{ fb}^{-1}$  of data in CMS. Oct 2013.
- [102] Jet Energy Resolution in CMS at  $\sqrt{s}=7$  TeV. 2011.
- [103] Performance of b tagging at  $\sqrt{s} = 8$  tev in multijet,  $t\bar{t}$  and boosted topology events. Oct 2013.

- [104] Valentin Hirschi, Rikkert Frederix, Stefano Frixione, Maria Vittoria Garzelli, Fabio Maltoni, and Roberto Pittau. Automation of one-loop QCD corrections. *JHEP*, 05:044, 2011.

