Top Quark Physics at the LHC with the CMS Detector

Sergey Senkin

University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

$\approx$ 34500 words

School of Physics

April 2014

Abstract

In this thesis, the top quark mass and top quark pair differential cross section measurements are presented, based on data collected by the CMS detector at the LHC in 2011 and 2012. The mass of the top quark is measured using a sample of top-antitop candidate events with one electron and at least four jets in the final state, obtained from approximately 5.0 fb^{-1} of data at a centre of mass energy of $\sqrt{s} = 7$ TeV. The top quark mass is extracted from a combined likelihood fit, and measured to be 172.87 ± 0.27 (stat.) $\pm\, 2.17$ (syst.) GeV. The normalised differential cross section of top quark pair production is measured with respect to a number of event-level observables, including missing transverse energy, scalar sum of jet transverse momenta, scalar sum of total event transverse momenta, and leptonically decaying W boson transverse momentum and transverse mass. The analysis is performed for semileptonic top-antitop events, with either an electron or a muon, and at least four jets in the final state. Approximately 19.7 fb^{-1} of data at a centre of mass energy of $\sqrt{s} = 8$ TeV are used. No significant deviations from the Standard Model predictions are observed.

To my parents,

Ivan Senkin and Gulzhan Senkina

Acknowledgements

The past three and a half years have been a challenging yet extremely rewarding experience. Quite a few people have made it possible, and I am incredibly grateful to all of them. First of all, I would like to thank my supervisor, Greg Heath, for his valuable guidance, advice and constructive suggestions, especially in writing up this thesis. I also thank Joel Goldstein for leading the research group and being a wise and kind host to our weekly meetings. Your advice has been greatly appreciated.

This work would not have been possible without some serious funding. I would like to thank the Faculty of Science and the selection panel for awarding me the University of Bristol Postgraduate Research Scholarship. It has been a great honour and I hope I managed to live up to it. I'm also very thankful to Dave Newbold (and Joel, again) for arranging the additional funding for my work at CERN, as otherwise my time there would have been rather miserable despite the great privilege of witnessing the Higgs boson discovery.

Many thanks to Jeson Jacob, Łukasz Kreczko and Phil Symonds for being great teammates over the last few years. I am particularly grateful to Luke who has been a nearly endless source of knowledge, and always willing to help. K-thx-bye!

A number of people helped me during my placement at CERN. In particular, I would like to thank Martijn Mulders and Enrique Palencia for their collaboration and helpful advice on the top mass analysis. My work on the high-level trigger was greatly influenced by Stéphanie Beauceron and Sam Harper – thank you for the guidance and useful advice. And special thanks to the UK Liaison Office at CERN for helping me deal with the French bureaucracy.

I would also like to express my gratitude to Alexei Safonov and Jim Pivarski for introducing me to CMS back in 2008, and giving me a chance to appreciate and get inspired by the collaboration spirit. Thank you, without your help and good will this PhD would not have been possible.

My warmest thanks go to all my friends and family, and especially to my parents. Your continuous encouragement and support throughout all this time have been immensely appreciated.

Author's Declaration

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

Signed:. . . . . Date:. . . . .

Contents

1 Introduction **1**

2 Theoretical Background **5**

2.1 The Standard Model of Particle Physics 5

2.1.1 Gauge principle 7

2.1.2 Electroweak theory 9

2.1.3 Quantum Chromodynamics 11

2.1.4 Electroweak symmetry breaking 13

2.1.5 Shortcomings of the SM and physics beyond it 16

2.2 Top Quark Physics within the Standard Model 18

2.2.1 Top quark production at the LHC 18

2.2.2 Top quark decay 20

2.2.3 Background processes to semileptonic top quark pair decay . . 21

2.2.4 Top quark mass 24

2.2.5 Top quark pair production cross section 28

2.3 Summary 31

3 The LHC and the CMS detector **33**

3.1 The Large Hadron Collider 33

3.2 The CMS Detector 36

3.2.1 Inner Tracking System 38

3.2.2 Electromagnetic Calorimeter 40

3.2.3 Hadron Calorimeter 42

3.2.4 Superconducting Magnet 44

3.2.5 Muon System 44

3.2.6 Trigger and Data Acquisition 47

3.3 Computing 48

3.3.1 Event Data Model 48

3.3.2 Analysis Software 49

3.4 Monte Carlo Simulation 50

3.4.1 Generation of hadron collision events 51

3.4.2 Monte Carlo event generators 53

3.4.3 CMS detector simulation 54

3.5 Object Reconstruction . . . 55
3.5.1 Electron reconstruction . . . 55
3.5.1.1 Electron identification . . . 57
3.5.1.2 Electron isolation . . . 61
3.5.1.3 Identification of photon conversions . . . 61
3.5.2 Muon reconstruction . . . 62
3.5.3 Jet reconstruction . . . 64
3.5.3.1 Jet energy corrections . . . 64
3.5.3.2 Particle Flow jet identification . . . 65
3.5.3.3 b-tagging . . . 66
3.5.4 Missing transverse energy . . . 67
3.6 Summary . . . 68

4 High-Level Triggers for Top Quark Physics 69
4.1 Level-1 triggers . . . 70
4.2 High-Level Triggers . . . 71
4.3 Trigger efficiency measurement . . . 75
4.4 Validation of jet energy corrections and charged hadron subtraction . 77
4.5 Summary . . . 83

5 Top Quark Mass Measurement 85
5.1 Data and Simulation . . . 85
5.1.1 Data . . . 85
5.1.2 Monte Carlo samples . . . 85
5.2 Event Selection . . . 86
5.3 Kinematic Fit . . . 93
5.4 The Ideogram Method . . . 97
5.4.1 Event likelihood . . . 97
5.4.2 Combined likelihood fit . . . 102
5.5 Mass Calibration . . . 103
5.6 Systematic Uncertainties . . . 105
5.7 Results . . . 108
5.8 Summary . . . 108

6 Top Quark Pair Differential Cross Section Measurement 109
6.1 Data and Simulation . . . 109
6.1.1 Data . . . 109
6.1.2 Monte Carlo samples . . . 109
6.1.3 Pile-up reweighting . . . 112

6.1.4 b-tagging corrections . . . 114
6.2 Event Selection . . . 117
6.2.1 Electron plus jets channel . . . 117
6.2.2 Muon plus jets channel . . . 119
6.2.3 Additional filters for missing transverse energy . . . 121
6.3 Data-driven QCD Estimation . . . 123
6.3.1 QCD multi-jet events with electrons . . . 123
6.3.2 QCD multi-jet events with muons . . . 125
6.4 Data/MC Comparison . . . 126
6.5 Differential Cross Section Measurement . . . 130
6.5.1 Primary variables . . . 130
6.5.2 Choice of binning . . . 131
6.5.3 Fitting procedure . . . 132
6.5.4 Unfolding . . . 139
6.5.5 Differential cross section calculation . . . 141
6.6 Systematic Uncertainties . . . 144
6.7 Results . . . 145
6.8 Summary . . . 150

7 Conclusions **151**

A High-level triggers in 2011 data **153**

B Choice of binning for primary variables **155**

C Fitting templates **159**

D Systematic uncertainties of the cross section measurement **167**

References **171**

List of Figures

2 Theoretical Background

2.1 The Higgs field potential . . . 14
2.2 Loop contributions to the Higgs boson mass . . . 17
2.3 Feynman diagrams for leading order $t\bar{t}$ production at the LHC . . . 19
2.4 Feynman diagrams for leading order single top production . . . 19
2.5 Semileptonic $t\bar{t}$ decay mode . . . 21
2.6 Example diagrams for leading order W and Z boson production at the LHC . . . 22
2.7 Example diagrams for leading order QCD multi-jet production at the LHC . . . 23
2.8 The Standard Model vacuum stability . . . 26
2.9 World combination of the top quark mass measurements . . . 27
2.10 CT10 NNLO parton distribution functions . . . 29
2.11 Measurement of the $t\bar{t}$ production cross section . . . 30
2.12 Normalised differential $t\bar{t}$ production cross section measurements by ATLAS and CMS collaborations . . . 31

3 The LHC and the CMS detector

3.1 CERN accelerator complex . . . 33
3.2 Production cross sections for different processes as a function of the centre of mass energy . . . 35
3.3 Sectional view of the CMS detector . . . 37
3.4 Cross section of the CMS tracker system . . . 38
3.5 Material budget as a function of pseudorapidity η for the different sub-detectors of the tracker . . . 39
3.6 Layout of the CMS electromagnetic calorimeter . . . 41
3.7 Relative ECAL energy resolution in bins of pseudorapidity η . . . 42
3.8 Longitudinal view of the CMS detector . . . 43
3.9 Layout of one quarter of the CMS muon system . . . 45
3.10 Muon transverse momentum resolution as a function of the transverse momentum . . . 46
3.11 Schematic overview of a hadron collision process . . . 51

3.12 b-tagging efficiency for the CSV algorithm 67

4 High-Level Triggers for Top Quark Physics

4.1 Efficiency of the 20 GeV e/γ trigger 70

4.2 HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_TriCentralPFJet30 trigger efficiency as a function of the third jet p_{T} and η 79

4.3 HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_TriCentralPFJet30 trigger efficiency as a function of the third jet p_{T} and η 79

4.4 HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_TriCentralPFJet30 trigger efficiency as a function of the third jet p_{T} and η 80

4.5 HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_TriCentralPFJet30 trigger efficiency as a function of the third jet p_{T} and η 80

4.6 Jet response as a function of jet p_{T} and η 82

5 Top Quark Mass Measurement

5.1 Kinematic variable distributions after all selection cuts 92

5.2 Residuals of the fitted top quark mass 95

5.3 Kinematic distributions after the fit 96

5.4 Fitted top quark mass distribution for $\mathrm{t\bar{t}}$ events with correct permutations 99

5.5 Fitted top quark mass distribution for $\mathrm{t\bar{t}}$ events with wrong permutations 101

5.6 Fitted top quark mass distribution for background events 101

5.7 Log-likelihood curve as a function of top quark mass 102

5.8 Average mass bias and width of the pull as a function of the true top quark mass before and after calibration 104

6 Top Quark Pair Differential Cross Section Measurement

6.1 Number of reconstructed vertices per event before and after pile-up reweighting 113

6.2 Number of expected vertices per event for central and $\pm 1\sigma$ variations of the total inelastic cross section for the 2012 data 114

6.3 Number of reconstructed vertices per event for $\pm 1\sigma$ variations of the pile-up reweighting procedure 115

6.4 b-tag multiplicity before and after applying the b-tagging scale factors 116

6.5 Distribution of the electron $|\eta|$ for conversion selection and non-isolated electron selection, and the shape comparison of both QCD control regions 124

6.6 Muon relative isolation with no isolation cut and muon $|\eta|$ for non-isolated selection in muon plus jets events . . . 125
6.7 Data/MC comparison plots of electron and muon p_{T} and $|\eta|$ after the final event selection . . . 127
6.8 Data/MC comparison plots of $E_{\mathrm{T}}^{\mathrm{miss}}$ after the final event selection . . . 128
6.9 Data/MC comparison plots of $\phi(E_{\mathrm{T}}^{\mathrm{miss}})$ and jet multiplicity after the final event selection . . . 129
6.10 Reconstructed versus generated $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 133
6.11 Electron $|\eta|$ templates for the fit in different bins of $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 134
6.12 Muon $|\eta|$ templates for the fit in different bins of $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 135
6.13 Data/MC comparison plots of $E_{\mathrm{T}}^{\mathrm{miss}}$ using the normalisation from the fitted results . . . 136
6.14 Data/MC comparison plots of H_{T} and S_{T} using the normalisation from the fitted results . . . 137
6.15 Data/MC comparison plots of $p_{\mathrm{T}}^{\mathrm{W}}$ and $M_{\mathrm{T}}^{\mathrm{W}}$ using the normalisation from the fitted results . . . 138
6.16 Pull distributions for the SVD unfolding . . . 142
6.17 Unfolding performance tests for $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 143
6.18 Normalised semileptonic $\mathrm{t\bar{t}}$ differential cross section with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 147
6.19 Normalised semileptonic $\mathrm{t\bar{t}}$ differential cross section with respect to H_{T} 147
6.20 Normalised semileptonic $\mathrm{t\bar{t}}$ differential cross section with respect to S_{T} 148
6.21 Normalised semileptonic $\mathrm{t\bar{t}}$ differential cross section with respect to $M_{\mathrm{T}}^{\mathrm{W}}$ . . . 148
6.22 Normalised semileptonic $\mathrm{t\bar{t}}$ differential cross section with respect to $p_{\mathrm{T}}^{\mathrm{W}}$ 149

B Choice of binning for primary variables

B.1 Reconstructed versus generated H_{T} and S_{T} . . . 155
B.2 Reconstructed versus generated $p_{\mathrm{T}}^{\mathrm{W}}$ and $M_{\mathrm{T}}^{\mathrm{W}}$ . . . 156

C Fitting templates

C.1 Electron $|\eta|$ templates for the fit in different bins of H_{T} . . . 159
C.2 Muon $|\eta|$ templates for the fit in different bins of H_{T} . . . 160
C.3 Electron $|\eta|$ templates for the fit in different bins of S_{T} . . . 161
C.4 Muon $|\eta|$ templates for the fit in different bins of S_{T} . . . 162
C.5 Electron $|\eta|$ templates for the fit in different bins of $p_{\mathrm{T}}^{\mathrm{W}}$ . . . 163
C.6 Muon $|\eta|$ templates for the fit in different bins of $p_{\mathrm{T}}^{\mathrm{W}}$ . . . 164
C.7 Electron $|\eta|$ templates for the fit in different bins of $M_{\mathrm{T}}^{\mathrm{W}}$ . . . 165
C.8 Muon $|\eta|$ templates for the fit in different bins of $M_{\mathrm{T}}^{\mathrm{W}}$ . . . 166

List of Tables

2 Theoretical Background

2.1 Fundamental fermions of the Standard Model 6

2.2 Fundamental forces and corresponding gauge bosons 6

2.3 Theoretical predictions for $\mathrm{t\bar{t}}$ production cross section at different LHC centre of mass energies 29

3 The LHC and the CMS detector

3.1 LHC beam parameters 35

3.2 ECAL crystal characteristics 41

3.3 Parameters of the CMS superconducting solenoid 44

3.4 Variables used in electron identification algorithms 58

4 High-Level Triggers for Top Quark Physics

4.1 Naming conventions for electron trigger working points 73

5 Top Quark Mass Measurement

5.1 Signal and background Monte Carlo samples at $\sqrt{s} = 7$ TeV 87

5.2 QCD multi-jet background and γ + jets MC samples at $\sqrt{s} = 7$ TeV 87

5.3 Systematic Monte Carlo samples at $\sqrt{s} = 7$ TeV 88

5.4 Central values and variations of factorisation/renormalisation scale (Q^2) and matrix element to parton shower (ME-PS) matching 88

5.5 Number of expected and observed events from MC and data in the electron channel 91

5.6 Overview of the systematic uncertainties and their effects on the top quark mass measurement 105

6 Top Quark Pair Differential Cross Section Measurement

6.1 Signal and background Monte Carlo samples at $\sqrt{s} = 8$ TeV 110

6.2 QCD multi-jet background and γ + jets MC samples used in the electron plus jets channel at $\sqrt{s} = 8$ TeV 111

6.3 QCD multi-jet background MC samples used in the muon plus jets channel at $\sqrt{s} = 8$ TeV 111

6.4 Systematic Monte Carlo samples at $\sqrt{s} = 8$ TeV 112

6.5 Number of expected and observed events from MC and data in the electron channel . . . 126
6.6 Number of expected and observed events from MC and data in the muon channel . . . 130
6.7 Stability and purity of $E_{\mathrm{T}}^{\mathrm{miss}}$ bins in the electron channel . . . 132
6.8 Stability and purity of $E_{\mathrm{T}}^{\mathrm{miss}}$ bins in the muon channel . . . 132
6.9 Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$ . . . 146

B Choice of binning for primary variables

B.1 Stability and purity of H_{T} bins in the electron channel . . . 157
B.2 Stability and purity of H_{T} bins in the muon channel . . . 157
B.3 Stability and purity of S_{T} bins in the electron channel . . . 157
B.4 Stability and purity of S_{T} bins in the muon channel . . . 158
B.5 Stability and purity of $p_{\mathrm{T}}^{\mathrm{W}}$ bins in the electron channel . . . 158
B.6 Stability and purity of $p_{\mathrm{T}}^{\mathrm{W}}$ bins in the muon channel . . . 158
B.7 Stability and purity of $M_{\mathrm{T}}^{\mathrm{W}}$ bins in the electron channel . . . 158
B.8 Stability and purity of $M_{\mathrm{T}}^{\mathrm{W}}$ bins in the muon channel . . . 158

D Systematic uncertainties of the cross section measurement

D.1 Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to H_{T} . . . 167
D.2 Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to S_{T} . . . 168
D.3 Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to $M_{\mathrm{T}}^{\mathrm{W}}$ . . . 169
D.4 Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to $p_{\mathrm{T}}^{\mathrm{W}}$ . . . 170

1. Introduction

The beginning of the 21st century has been a truly exciting time in particle physics. First LHC collisions in 2009 marked the start of a new era which only three years later brought about the discovery of the Higgs boson, the last missing piece of the Standard Model. The hopes are high that it is not the last discovery of this era, and new physics is hiding around the corner at the energies accessible to the LHC.

While searches for new physics are very exciting, one should not underestimate the importance of precision measurements of the Standard Model. Even after the Higgs boson discovery, top quark physics has remained in the priorities of the LHC physics programme. One of the main reasons is its importance as a primary background to many new physics scenarios beyond the Standard Model. Moreover, it is still not understood why the top quark Yukawa coupling is so close to unity, which implies that severe fine tuning of the Higgs mass happens mostly due to the top quark. Many extensions of the Standard Model offer solutions to this hierarchy problem by extending the top quark sector and introducing more degrees of freedom, which are expected to cause deviations from the Standard Model predictions in top quark-related observables. Therefore, precise measurements of the top quark properties are of high importance.

The top quark mass is a crucial fundamental parameter of the Standard Model. The mass analysis presented in this thesis contributes to the most precise single measurement of the top quark mass to date. While the systematic uncertainty due to the jet energy scale (JES) in the author's work is significantly larger than that of the published CMS measurement, the analysis serves as an important cross-check of the mass extraction technique and the kinematic fitting procedure. The mentioned JES systematic uncertainty is mitigated in the published measurement via the *in situ* measurement of the JES and the top quark mass in a joint likelihood fit. This thesis shows the consistency of these measurements within uncertainties, as well as with the first world combination of the top quark mass measurements.

The LHC has proven to be a top quark factory, with an abundance of statistics pushing the limits of precision calculations to the level of uncertainties of theoretical predictions. All these data need to be carefully selected, as recording the top quark events is significantly complicated by background processes with similar signatures occurring at much higher rates. Therefore, an efficient and manageable trigger is a crucial part of any top quark analysis. In this thesis, work on the high-level triggers used for selection of top quark events with semileptonic signature, particularly with an electron and jets in the final state, is presented.

For the first time in the history of particle physics, the abundance of top quark pair ($\mathrm{t\bar{t}}$) events at the LHC gives an opportunity to measure the $\mathrm{t\bar{t}}$ differential cross section with respect to various quantities. Both the CMS and ATLAS collaborations have published results on such measurements with respect to top quark-related variables. The $\mathrm{t\bar{t}}$ differential cross section presented here is measured with respect to event-level distributions, which do not require a kinematic reconstruction of the $\mathrm{t\bar{t}}$ decay. The absence of systematic uncertainties associated with the kinematic reconstruction is the main advantage of these measurements, which allows more detailed comparison of the data and predictions by different Monte Carlo generators. Furthermore, new physics could reveal itself in the tails of these event-level distributions. For instance, an associated production of a $\mathrm{t\bar{t}}$ pair with a new resonance ($\mathrm{t\bar{t}+X}$) decaying invisibly may show up in the tail of the missing transverse energy distribution.

The work on the top quark mass analysis using 7 TeV LHC data was performed at CERN in collaboration with Martijn Mulders and Enrique Palencia. The author's main contribution was to the electron side of the analysis, particularly all the analysis steps up to setting up the kinematic fit and obtaining the fitted information which was used in the mass extraction technique. The high-level triggers used in the analysis were developed by the author in collaboration with Łukasz Kreczko and Stéphanie Beauceron.

The differential cross section analysis was done in close collaboration with Łukasz Kreczko, Jeson Jacob and Phil Symonds under the supervision of Greg Heath and Joel Goldstein. On the technical side of the analysis, the author mainly focused on developing the C++/Python-based analysis framework (Bristol Analysis Software) and ensuring the latest and most precise physics object definitions were used; producing the local n-tuples; programming Python scripts in conjunction with ROOT/PyROOT software to extract the differential cross section and produce final results. On the physics side, the main contribution was also to the electron channel, particularly the missing transverse energy measurement, data-driven QCD estimation, evaluation of systematic uncertainties and implementation of the Singular Value Decomposition (SVD) unfolding.

The thesis is structured as follows. Some theoretical background on the Standard Model and top quark physics is given in Chapter 2. Then the LHC and the CMS detector, as well as the relevant aspects of Monte Carlo simulation and object reconstruction are discussed in Chapter 3. The author's contribution to the CMS high-level trigger development is detailed in Chapter 4. Afterwards, two major analysis chapters are presented: the top quark mass measurement in Chapter 5, and the top quark pair differential cross section measurement in Chapter 6. Finally, conclusions and outlook are given in the last chapter.

Conventions

In this thesis, the natural units convention is adopted:

$$\hbar = c = 1,$$

therefore, the same units (electronvolts, eV) are used to denote momentum, energy and mass quantities. Typically, these would be GeV units (1 GeV = 10^9 eV), due to the energies accessible at the LHC. The coordinate system used to describe the geometry of the detector and event kinematics is explained in Section 3.2.

2. Theoretical Background

2.1 The Standard Model of Particle Physics

The Standard Model (SM) is a quantum field theory which is currently the most accurate description of matter and all known fundamental interactions, with the exception of gravity. It was developed in the 20th century as a combination of two complementary field theories, the Glashow-Weinberg-Salam (GWS) [1, 2, 3] theory of the electroweak interaction, and the Quantum Chromodynamics (QCD) [4, 5, 6] theory of the strong interaction.

The fundamental particles in the Standard Model are the three generations of fermions, shown in Table 2.1, interacting via gauge bosons of three fundamental interactions, presented in Table 2.2. Fermions, subdivided into quarks and leptons, are the building blocks of matter in the observable universe. Their interaction via exchanging the gauge bosons allows formation of hadrons and atoms. The difference between quarks and leptons mainly lies in the way they interact: quarks can interact via strong, electromagnetic and weak forces, whereas leptons are only subject to electromagnetic and weak interactions. Each of these particles also has a corresponding antiparticle with the same mass, but opposite charge.

There are six flavours of leptons in the SM, making up three generations: electron (e), muon (μ) and tau (τ) leptons with their corresponding neutrinos. Electrons, muons and taus are massive and charged particles, interacting via electromagnetic and weak forces, described in Section 2.1.2. Neutrinos, on the other hand, do not carry an electric charge, and therefore are only subject to the weak force, which makes them extremely difficult to detect as they barely interact with matter. In the classic version of the Standard Model neutrinos are considered to be massless, however, the observation of neutrino oscillations between different flavours [7] implies that neutrinos are in fact massive. This is one of the deficiencies of the Standard Model, discussed in Section 2.1.5.

The quarks also exist in three generations, and are represented by up and down (u, d), charm and strange (c, s), and bottom and top (b, t) flavours. Together with the mediators of the strong interaction (gluons), these particles possess another physical property called colour charge. Due to the phenomenon of colour confinement, they can only exist as constituents of composite particles – hadrons. More detail on the QCD theory of the strong interaction is presented in Section 2.1.3. The top quark, however, is too heavy to hadronise as it has an extremely short lifetime of $\approx 5 \times 10^{-25}$ s [8], which is an order of magnitude shorter than the timescale of strong interactions. This and other properties of the top quark will be discussed in Section 2.2.

The last essential constituent of the Standard Model is the recently discovered Higgs boson [9, 10]. The Higgs field is responsible for electroweak symmetry breaking of the $SU(2)_L \times U(1)_Y$ group and therefore the acquisition of mass by vector bosons. This mechanism is described in more detail in Section 2.1.4.

Table 2.1: Three generations of fundamental fermions of the Standard Model with charge and mass properties quoted from the 2012 edition of Review of Particle Physics with 2013 partial update for the 2014 edition [8, 11].

Generation	*Leptons*			*Quarks*		
	Flavour	*Charge*	*Mass* [MeV]	*Flavour*	*Charge*	*Mass* [MeV]
1	e	-1	0.511	u	2/3	$2.3^{+0.7}_{-0.5}$
	ν_e	0	$< 2 \times 10^{-6}$	d	$-1/3$	$4.8^{+0.5}_{-0.3}$
2	μ	-1	105.66	c	2/3	$(1.29^{+0.05}_{-0.11}) \times 10^3$
	ν_μ	0	< 0.19	s	$-1/3$	95 ± 5
3	τ	-1	1776.82 ± 0.16	t	2/3	$(173.34 \pm 0.76) \times 10^3$
	ν_τ	0	< 18.2	b	$-1/3$	$(4.18 \pm 0.03) \times 10^3$

Table 2.2: Fundamental forces and corresponding gauge bosons with their properties [8]. The gravitational force is the only fundamental interaction not described by the Standard Model (the graviton has not been discovered yet).

Force	*Gauge Boson*	*Charge*	*Spin*	*Mass* [GeV]	*Range*
electromagnetic	photon (γ)	0	1	0	∞
weak	$W^\pm$	± 1	1	80.385 ± 0.015	10^{-18} m
	Z^0	0		91.1876 ± 0.0021	
strong	8 gluons (g)	0	1	0	10^{-15} m
gravitational	graviton (G)	0	2	0	∞

2.1.1 Gauge principle

As previously mentioned, the Standard Model is a unification of two gauge theories describing electroweak and strong interactions. The GWS model, a unified theory of electromagnetic and weak interactions, is based on a gauge group of $SU(2)_L \times U(1)_Y$, whereas the QCD theory has a gauge symmetry of $SU(3)_C$. Therefore, the Standard Model is also a gauge theory with the symmetry group of $SU(3)_C \times SU(2)_L \times U(1)_Y$. A gauge theory is a theory that is invariant under a set of local transformations, i.e. transformations with space-time dependent parameters. To explain this concept, let us consider a Lagrangian density of a free Dirac field ψ, describing free-moving fermions with mass m:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{2.1}$$

where γ^μ are Dirac matrices and μ the Lorentz indices. Clearly, it is invariant under the following phase transformation:

$$\psi \rightarrow \psi' = e^{i\theta}\psi, \ \bar{\psi} \rightarrow \bar{\psi}' = e^{-i\theta}\bar{\psi}, \tag{2.2}$$

since in the combination $\bar{\psi}\psi$ the exponential factors cancel out ($e^{i\theta}e^{-i\theta} = 1$). This transformation is called a *global* phase transformation, since the phase θ is space-time independent. The *local* phase transformation, on the other hand, corresponds to the phase $\theta(x_\mu)$ which is different at each space-time point:

$$\psi \rightarrow \psi' = e^{i\theta(x)}\psi, \ \bar{\psi} \rightarrow \bar{\psi}' = e^{-i\theta(x)}\bar{\psi}. \tag{2.3}$$

Here the space-time indices are dropped on $x_\mu \equiv x$ for clarity. One can see that the Lagrangian is no longer invariant under such transformation, since

$$\partial_\mu(e^{i\theta(x)}\psi) = i(\partial_\mu\theta(x))e^{i\theta(x)}\psi + e^{i\theta(x)}\partial_\mu\psi, \tag{2.4}$$

and therefore

$$\mathcal{L} \rightarrow \mathcal{L} - [\partial_\mu\theta(x)]\bar{\psi}\gamma^\mu\psi. \tag{2.5}$$

The space-time dependent local transformations are called gauge transformations. In order to restore gauge invariance under such transformations, we need to replace the ordinary derivative ∂_μ with the *covariant* derivative D_μ:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu, \tag{2.6}$$

where A_μ is a new vector field which transforms under a local gauge transformation as

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{q}\partial_\mu \theta(x). \tag{2.7}$$

Together with the kinetic energy term of the field-strength tensor defined as

$$F_{\mu\nu} \equiv [\partial_\mu A_\nu] - [\partial_\nu A_\mu], \tag{2.8}$$

we obtain the full gauge invariant QED Lagrangian:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu(\partial^\mu + iqA_\mu) - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.9}$$

The requirement of local gauge invariance implies the gauge field A_μ must be massless, since the term proportional to $A_\mu A^\mu$ is not invariant under local gauge transformations and therefore has to be excluded from the Lagrangian. In the case of QED, the gauge field A_μ represents the photon field, and the Lagrangian describes the interaction of Dirac fields (electrons and positrons) with Maxwell fields (photons).

The idea of requiring local gauge invariance by introducing additional fields in the Lagrangian to make it covariant with respect to an extended group of local transformations is an important procedure in particle physics, referred to as the gauge principle. In the example above, the local phase transformation can be thought of as multiplication of the Dirac field ψ by a 1×1 unitary matrix ($\psi \to U\psi$, where $U^\dagger U = 1$). Here $U = e^{i\theta(x)}$, and all such transformations form the Lie group $U(1)$. This is an Abelian group, since any two of its elements commute:

$$[e^{i\theta(x)},\ e^{i\theta(x')}] = 0. \tag{2.10}$$

The same strategy of requiring a global invariance to hold locally can be generalised to any $SU(N)$ group: in the 1950s Yang and Mills [12] produced a theory of $SU(2)$ gauge fields, and later on the idea was extended to $SU(3)$ group giving rise to Quantum Chromodynamics. In fact, all of the fundamental interactions in the Standard Model are introduced using the same gauge principle.

2.1.2 Electroweak theory

In 1961, a partially unified theory combining the electromagnetic and weak interactions was proposed by Sheldon Glashow [1]. A few years later the model was independently revised by Weinberg [2] and Salam [3] to introduce massive vector bosons, acquiring mass through spontaneous symmetry breaking. The resulting electroweak theory, usually referred to as GWS model, is based on a non-Abelian (i.e. not commuting) symmetry group of $SU(2)_L \times U(1)_Y$. Here $U(1)_Y$ is no longer the QED gauge group mentioned previously, but the group associated with the weak hypercharge Y, which relates to the electric charge Q and the third component of weak isospin I_3 as

$$Q = I_3 + \frac{Y}{2} \tag{2.11}$$

This equation is known as the Gell-Mann-Nishijima relation, and the coefficient in front of the weak hypercharge is purely conventional. The weak force is the only interaction that violates parity, meaning that it can distinguish between left-handed and right-handed systems – this fact was famously confirmed by Madame Wu's experiment in 1957 [13]. To accommodate for this observation, fermion fields are divided into left-handed and right-handed components as follows:

$$\begin{pmatrix} \nu_{l,L} \\ l_L \end{pmatrix}, l_R \tag{2.12}$$

for leptons ($l = e, \mu, \tau$), and

$$\begin{pmatrix} \mathrm{u}_L \\ \mathrm{d}_L \end{pmatrix}, \mathrm{u}_R, \mathrm{d}_R; \quad \begin{pmatrix} \mathrm{c}_L \\ \mathrm{s}_L \end{pmatrix}, \mathrm{c}_R, \mathrm{s}_R; \quad \begin{pmatrix} \mathrm{t}_L \\ \mathrm{b}_L \end{pmatrix}, \mathrm{t}_R, \mathrm{b}_R \tag{2.13}$$

for all three generations of quarks. The left-handed particles are doublets with the weak isospin of $I = \frac{1}{2}$, whereas the right-handed particles are singlets with $I = 0$. These components are obtained by applying the projection operators $\frac{1}{2}(1 \pm \gamma^5)$ to fermionic fields:

$$\psi = \frac{1}{2}(1 - \gamma^5)\psi + \frac{1}{2}(1 + \gamma^5)\psi = \psi_L + \psi_R, \tag{2.14}$$

where

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.15}$$

By requiring local gauge invariance under the group $SU(2)_L \times U(1)_Y$, the covariant derivative is written as

$$\partial_\mu \to D_\mu = \partial_\mu + ig\mathbf{T} \cdot \mathbf{W}_\mu + i\frac{g'}{2}YB_\mu. \tag{2.16}$$

Here g and g' are the coupling constants for the $SU(2)_L$ and $U(1)_Y$ groups, respectively; $\mathbf{T} = \hat{\sigma}/2$, where $\hat{\sigma}$ are the Pauli matrices which are the generators for the $SU(2)$ group; $\mathbf{W}_\mu$ is the triplet of new gauge fields $W_\mu^{1,2,3}$ introduced for $SU(2)_L$ invariance; B_μ is the new gauge field for $U(1)_Y$ invariance. For right-handed particles (singlets), $\mathbf{T} = 0$ and the second term in the covariant derivative vanishes. The resulting gauge-invariant Lagrangian for electroweak theory with kinetic energy terms takes the form:

$$\begin{aligned}\mathcal{L}_{\mathrm{EWK}} = \bar{\psi}_L\gamma^\mu(i\partial_\mu - g\mathbf{T}\cdot\mathbf{W}_\mu - \frac{g'}{2}YB_\mu)\psi_L + \\ + \bar{\psi}_R\gamma^\mu(i\partial_\mu - \frac{g'}{2}YB_\mu)\psi_R - \frac{1}{4}\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu},\end{aligned} \tag{2.17}$$

where

$$\psi_L = \begin{pmatrix}\psi_L^1 \\ \psi_L^2\end{pmatrix},$$

and ψ_L and ψ_R are summed over all possibilities shown in Equations 2.12 and 2.13.

The newly introduced gauge fields $W_\mu^{1,2,3}$ and B_μ do not directly correspond to the physical fields of gauge bosons, but form them via the following linear combinations:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2), \tag{2.18a}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W, \tag{2.18b}$$

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W. \tag{2.18c}$$

Here the physical fields of the charged $W^\pm$ bosons, as well as a neutral Z^0 boson and a photon were produced by mixing the neutral fields W_μ^3 and B_μ with a rotation by the weak mixing (or *Weinberg*) angle θ_W. The electric charge is then given as

$$e = g'\cos\theta_W = g\sin\theta_W. \tag{2.19}$$

Up to this point, there are no mass terms in the Lagrangian as their introduction would violate gauge invariance. However, the existence of massive W and Z bosons clearly contradicts the notion of massless gauge fields in electroweak theory. This problem is solved by the Higgs mechanism of spontaneous gauge symmetry breaking, which is a subject of Section 2.1.4.

2.1.3 Quantum Chromodynamics

Quantum Chromodynamics is the quantum field theory of strong interactions between quarks and gluons. It is a non-Abelian gauge theory based on the symmetry group $SU(3)_C$, where C denotes colour charge. This is motivated mainly by experimental evidence, which suggests that quarks are multiplets of fields in colour space:

$$q = \begin{pmatrix} r \\ g \\ b \end{pmatrix}. \tag{2.20}$$

After demanding the invariance under local $SU(3)$ gauge transformations, the QCD Lagrangian is given as

$$\mathcal{L}_{\mathrm{QCD}} = \bar{q}(\gamma^{\mu}\partial_{\mu} - m)q + g_S(\bar{q}\gamma^{\mu}T_a q)G^a_{\mu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a. \tag{2.21}$$

Here $T_a = \lambda_a/2$, where λ_a are the Gell-Mann matrices, i.e. eight generators of the $SU(3)$ group, satisfying the Lie algebra:

$$[T_a, T_b] = if_{abc}T_c. \tag{2.22}$$

This gives rise to eight gauge fields – massless gluons. The field strength $G^a_{\mu\nu}$ is a tensor analogous to that of QED (Equation 2.8), but contains an extra term proportional to the structure constants f^{abc}, responsible for self-interaction of gluons:

$$G^a_{\mu\nu} = [\partial_{\mu}G^a_{\nu}] - [\partial_{\nu}G^a_{\mu}] + g_S f^{abc}G^b_{\mu}G^c_{\nu}. \tag{2.23}$$

Notably, leptons are singlets under $SU(3)$ symmetry, and therefore do not participate in strong interactions with gluons and quarks. Self-interaction of gluons leads to two distinctive fundamental properties of QCD: asymptotic freedom and colour confinement.

Asymptotic freedom was firstly described by Frank Wilczek, David Gross [5], and independently by David Politzer [6] in 1973. This phenomenon is closely connected to the behaviour of QCD gauge coupling g_S, which is usually presented in terms of α_S defined as

$$\alpha_S(\mu) = \frac{g_S(\mu)^2}{4\pi}, \tag{2.24}$$

where μ is the energy scale of the process. It was discovered that α_S *decreases* with the increase of energy scale. This can be shown by evaluating contributions from quark-antiquark and (crucially) gluon-gluon loops [14], leading to first order to:

$$\alpha_S(\mu) = \frac{12\pi}{(33 - 2n_f)\ln(\mu^2/\Lambda^2_{\mathrm{QCD}})}, \tag{2.25}$$

where n_f is the number of flavours available at the energy scale μ, and Λ_{QCD} is the QCD scale, i.e. the value of renormalisation scale at which the QCD coupling diverges. Since the number of quark flavours is known to be less than 16, the coefficient of the logarithm in Equation 2.25 is positive, meaning that α_S in fact decreases with the increase of μ (or decrease of the probe distance). An important consequence of this phenomenon is the fact that at very short distances ($\sim$ 0.1 fm), the strong force is relatively weak and quarks become effectively free (hence the name "asymptotic freedom").

Colour confinement is another important consequence – since the strong force increases with the distance between quarks, if the quarks move apart the energy rises until it becomes sufficient to form a quark-antiquark pair. This process is called hadronisation, and it continues until the kinetic energy of the quarks is completely transformed into the energy of creation of new quark pairs. All quarks are therefore confined in colourless bound states, called hadrons. Hadrons formed by a quark-antiquark pair are called mesons, whereas combinations of three quarks or three antiquarks are referred to as baryons. Recently a tetraquark (c $\bar{\mathrm{c}}$ d $\bar{\mathrm{u}}$) state has been confirmed by the LHCb collaboration with an overwhelming significance of at least 13.9 σ [15]. Bound states with higher number of quarks (e.g. pentaquarks), or containing gluons only (glueballs) remain unobserved, despite being predicted by the Standard Model.

In Section 2.1.2 we introduced weak isospin doublets for three generations of left-handed quarks (Equation 2.13). These states, however, do not exactly correspond to the strong force eigenstates. Experimental evidence (e.g. $\Lambda \to p{+}\pi^-$ decay, involving the conversion of a strange quark into an up quark via s $\to$ W$^-$ + u) suggests that the weak interaction is capable of changing the quark flavour, meaning that the weak interaction eigenstates are in fact mixtures of mass eigenstates. The relation between the two sets of eigenstates is given by the Cabbibo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}, \tag{2.26}$$

where d′, s′, b′ are the weak interaction eigenstates, and d, s, b are the mass eigenstates. The CKM matrix describes flavour mixing, i.e. the probability of transitions between different quark flavours through weak interactions. The origin of such mixing lies in the Yukawa interaction between quarks and the Higgs field, described in the next section.

2.1.4 Electroweak symmetry breaking

The mechanism of spontaneous electroweak symmetry breaking, often called the Higgs mechanism, was developed in the 1960s by Peter Higgs [16] and two other groups independently: Robert Brout and François Englert [17]; Gerald Guralnik, Carl Richard Hagen, and Tom Kibble [18]. As it was already mentioned, this mechanism is responsible for acquisition of mass by the vector bosons, since direct inclusion of mass terms in the electroweak Lagrangian (Equation 2.17) would violate gauge symmetry. Therefore, we need to spontaneously break the $SU(2)$ symmetry by adding an external field with a non-zero vacuum expectation value. This is done by introducing an $SU(2)$ doublet of complex scalar fields, called the Higgs fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.27}$$

with the following additional term in the Lagrangian:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{2.28}$$

where D_μ is the electroweak covariant derivative from Equation 2.16. A simple possible scalar potential $V(\phi)$ is given by

$$V(\phi) = -\mu^2(\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2. \tag{2.29}$$

With the choice of $\mu^2 > 0$ and $\lambda > 0$, the potential has the shape shown in Figure 2.1. Clearly, the minimum of this potential is not at $\phi = 0$, but forms a circle in $SU(2)$ space specified by

$$(\phi^\dagger \phi)_{\text{min}} = \frac{\mu^2}{2\lambda} = \frac{v^2}{2}. \tag{2.30}$$

The choice of minimum corresponding to the lowest energy state (or vacuum) is completely arbitrary. In fact, any point at which the potential is minimum loses the invariance under $SU(2)_L \times U(1)_Y$ gauge transformations. Therefore, nature spontaneously breaks the symmetry by picking the vacuum from the set of minima

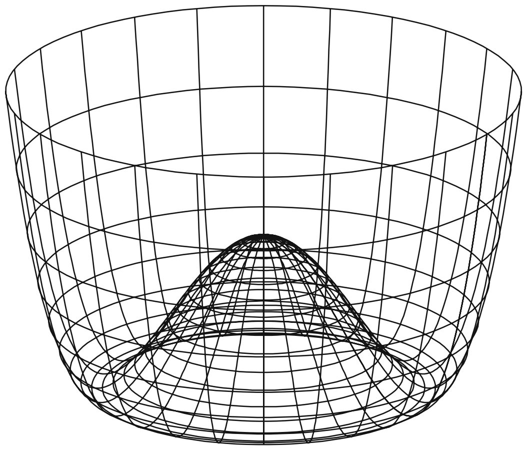


Figure 2.1: The Higgs field potential in the complex plane, commonly referred to as a "Mexican hat" potential [19]

of the Higgs potential. Conventionally, we can choose

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}. \tag{2.31}$$

Expanding around the chosen minimum, ϕ is then given by

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+H\end{pmatrix}, \tag{2.32}$$

where H is the neutral scalar Higgs field. By substituting this field into the Lagrangian in Equation 2.28, one can obtain:

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} = {}& \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{4}g^2(H^2 + 2vH + v^2)W_\mu^+ W^{-\mu} + \\ &+ \frac{1}{8}(g^2 + g'^2)(H^2 + 2vH + v^2)Z_\mu Z^\mu - \\ &- \mu^2 H^2 - \frac{\lambda}{4}(H^4 + 4vH^3),\end{aligned} \tag{2.33}$$

where physical fields $W_\mu^\pm$ and Z_μ are given by Equations 2.18a and 2.18b, respectively. Notably, we can see that these fields have acquired mass terms in the Lagrangian, with the masses of vector bosons given by

$$M_W = \frac{1}{2}gv, \tag{2.34a}$$

$$M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} = \frac{1}{2}\frac{gv}{\cos\theta_W}, \tag{2.34b}$$

whereas the mass of the Higgs boson itself is given by

$$M_H = \sqrt{2}\mu = v\sqrt{2\lambda}. \tag{2.35}$$

Naturally, the photon field A_μ acquires no mass terms in the Lagrangian. The Higgs mechanism can also be used in a similar way to generate fermion masses by introducing an $SU(2)_L \times U(1)_Y$ gauge invariant term responsible for interaction between the Higgs and fermion fields. This additional term in the Standard Model Lagrangian is called the *Yukawa term*, which for the first generation of fermions is given by

$$\mathcal{L}_{\text{Yukawa}} = -Y_e^{ij}\bar{l_L}^i \phi e_R^j - Y_u^{ij}\bar{q_L}^i \epsilon\phi^\dagger u_R^j - Y_d^{ij}\bar{q_L}^i \phi d_R^j + \text{h.c.}, \tag{2.36}$$

where coefficients $Y_{e,u,d}^{ij}$ are 3×3 complex matrices (Yukawa couplings) and ϵ is the 2×2 antisymmetric tensor. When the Higgs field ϕ acquires the vacuum expectation value given by Equation 2.31, the Yukawa Lagrangian yields mass terms for fermions, generating the masses:

$$M_f = Y_f \frac{v}{\sqrt{2}}. \tag{2.37}$$

The Yukawa couplings are not diagonal in general, which results in mixing between different generations described for the quarks by the CKM matrix (Equation 2.26). Another curious observation is that the fermion masses are proportional to Yukawa couplings, which essentially represent the interaction strength with the Higgs field. Due to the large mass of the top quark of approximately 173 GeV, and the Higgs field vacuum expectation value $v \approx 246$ GeV, the Yukawa coupling to the top quark is very close to unity.

The existence of the Higgs boson and therefore the nature of electroweak symmetry breaking had long remained a mystery. In the summer of 2012 both ATLAS and CMS experiments at the LHC observed a Higgs boson with an approximate mass of 125 GeV [9, 10], with couplings consistent with theoretical predictions [20, 21], which proved to be yet another triumph of the Standard Model.

2.1.5 Shortcomings of the SM and physics beyond it

Despite being the most successful theory to date, the Standard Model is not perfect and has its shortcomings. There is a range of physical phenomena not explained by the SM. Let us mention the most prominent ones.

Gravity is not included in the Standard Model, as a consistent theory of quantum gravity is yet to be derived. General relativity, the only accepted theory of gravity, is in fact incompatible with quantum mechanics and therefore the Standard Model.

Massive neutrinos are also not explained by the SM. The evidence for neutrino oscillations was confirmed by the Super-Kamiokande collaboration [7], which necessarily implies that neutrinos are massive particles. It is possible to include massive neutrinos in some extensions of the Standard Model, whilst also keeping the local symmetry of weak interactions [22]. However, this leads to new theoretical problems (e.g. unnaturally small neutrino Yukawa couplings).

Matter/antimatter asymmetry, which is apparent in the Universe, is not fully accounted for by the observed CP violation in the Standard Model [23, 24]. Therefore, another mechanism inducing the matter/antimatter asymmetry must exist, most likely requiring new physics models beyond the Standard Model.

Dark matter and dark energy constitute approximately 95 % of the mass and energy content in our Universe, according to the latest Planck data [25, 26]. The Standard Model only describes the remaining 5 % (i.e. ordinary matter), whereas the origin of dark matter and dark energy remains unknown.

Furthermore, there are a number of fundamental theoretical issues within the Standard Model, implying the intrinsic incompleteness of the theory. The SM does not explain why it has only three generations of fermions. Moreover, there is no explanation for the vast difference between masses of fermions (known as the Flavour problem), hence the origin of Yukawa couplings (and therefore the CKM matrix) remains an open problem in particle physics.

Additionally, the *ad hoc* nature of the Higgs mechanism gives rise to the hierarchy problem, which is related to quadratic divergence of the loop corrections to the Higgs boson mass. The Higgs mass is renormalised by loop diagrams (e.g. shown in Figure 2.2), making it extremely large – of the order of Planck scale ($\sim 10^{19}$ GeV) – unless there is some unnatural fine tuning involved.

All these problems motivate the need for new models, referred to as physics beyond the Standard Model (BSM). Most notable BSM models are briefly mentioned below.

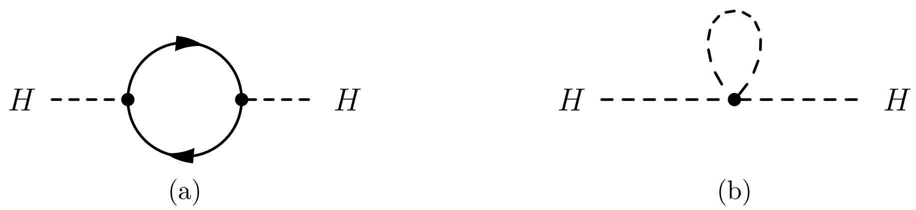


Figure 2.2: Loop contributions to the Higgs boson mass: (a) fermion loops, (b) boson loops.

Supersymmetry (SUSY) [27] is an extension of the SM, introducing a supersymmetric partner to each ordinary particle by varying its spin by 1/2. Thus, each fermion has a bosonic partner, and all bosons have fermionic partners. Despite the greater number of free parameters, SUSY offers attractive solutions to most of the mentioned problems of the SM. For example, the hierarchy problem is naturally solved in SUSY since the loop corrections are automatically cancelled by corresponding contributions from supersymmetric partners. Most SUSY models also provide good candidates for Dark Matter.

Extra dimensions refer to theories containing additional space-time dimensions, in which gravity is allowed to propagate. This leads to a reduction of the hierarchy between the Planck and electroweak scales, thus resolving the hierarchy problem [28, 29].

Composite models, such as topcolour [30] or composite Higgs [31] models, attempt to resolve some of the issues of the Standard Model by either introducing new composite particles or suggesting a composite nature for the existing particles in the SM.

An important and desirable feature of any physics model is its falsifiability in the accessible energy range. A variety of BSM models provide particle candidates that may exhibit themselves at the energy scale accessible at the LHC, therefore a rich physics programme at the LHC must ensure an active search for all possible signs of new physics.

2.2 Top Quark Physics within the Standard Model

The top quark is the heaviest elementary particle in the Standard Model, and the heaviest observed particle, discovered in 1995 at the Tevatron collider by the CDF and DØ collaborations [32, 33]. Its large mass of approximately 173 GeV suggests that the top quark may have a special role in nature. It is substantially more massive than any other fermion. As mentioned in Section 2.1.5, the reason for such vast discrepancy between fermion masses is unknown. It is also not yet understood why the top quark's Yukawa coupling to the Higgs boson is so close to unity. Naturally, studying the most massive fermion is a logical starting point in search for answers to these questions.

Due to its large mass, the top quark has an extremely short lifetime on the order of 10^{-25} s [8], meaning that it decays before the top-flavoured hadrons or $\mathrm{t\bar{t}}$ bound states can form. Therefore, the spin information is passed from the top quark on to its decay products, providing a unique possibility to study a "bare" quark. High-precision measurements of the top quark properties, such as mass (Chapter 5) and cross section (Chapter 6) are important tools to test the Standard Model and probe for new physics beyond it.

2.2.1 Top quark production at the LHC

There are two particular mechanisms of top quark production in hadron collisions: top-antitop ($\mathrm{t\bar{t}}$) pair production via the strong interaction, and single top quark production via the electroweak interaction. The $\mathrm{t\bar{t}}$ production dominates over single top production, being the main source of top quarks at the LHC, and therefore is the main focus of this thesis.

Feynman diagrams for leading order $\mathrm{t\bar{t}}$ production are presented in Figure 2.3. Since the LHC is a proton-proton collider with a high centre of mass energy, the only source of antiquarks are virtual quarks (so-called sea quarks), and the gluon contribution becomes dominating. Therefore, the major process for $\mathrm{t\bar{t}}$ production is gluon fusion, constituting approximately 90 % (80 %) at $\sqrt{s} = 14$ TeV (7 TeV) [8]. Quark-antiquark annihilation and higher-order processes contribute the rest of the Standard Model $\mathrm{t\bar{t}}$ production. At the proton-antiproton Tevatron, on the contrary, the quark-antiquark annihilation was the main $\mathrm{t\bar{t}}$ production mode.

Figure 2.4 shows the Feynman diagrams of single top production modes. Three major production mechanisms include s-channel and t-channel W boson exchange, and associated production with a W boson (tW-channel). Measurement of single top production is of significant interest since it allows the direct measurement of the Wtb vertex and therefore the magnitude of the $|V_{\mathrm{tb}}|$ element of the CKM matrix.

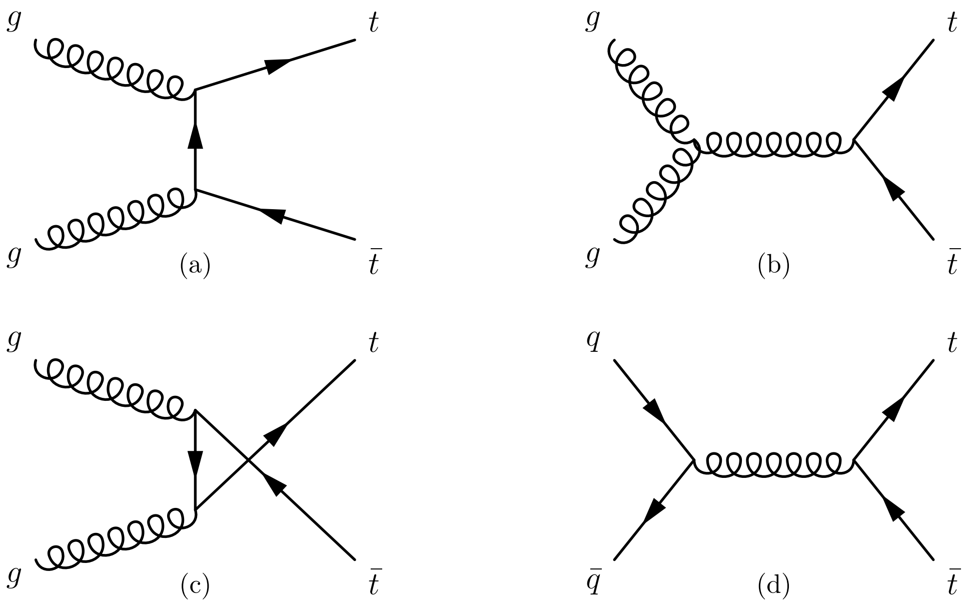


Figure 2.3: Feynman diagrams for leading order $t\bar{t}$ production at the LHC: (a), (b) and (c) show the gluon fusion, the dominant production mechanism, (d) represents quark-antiquark annihilation.

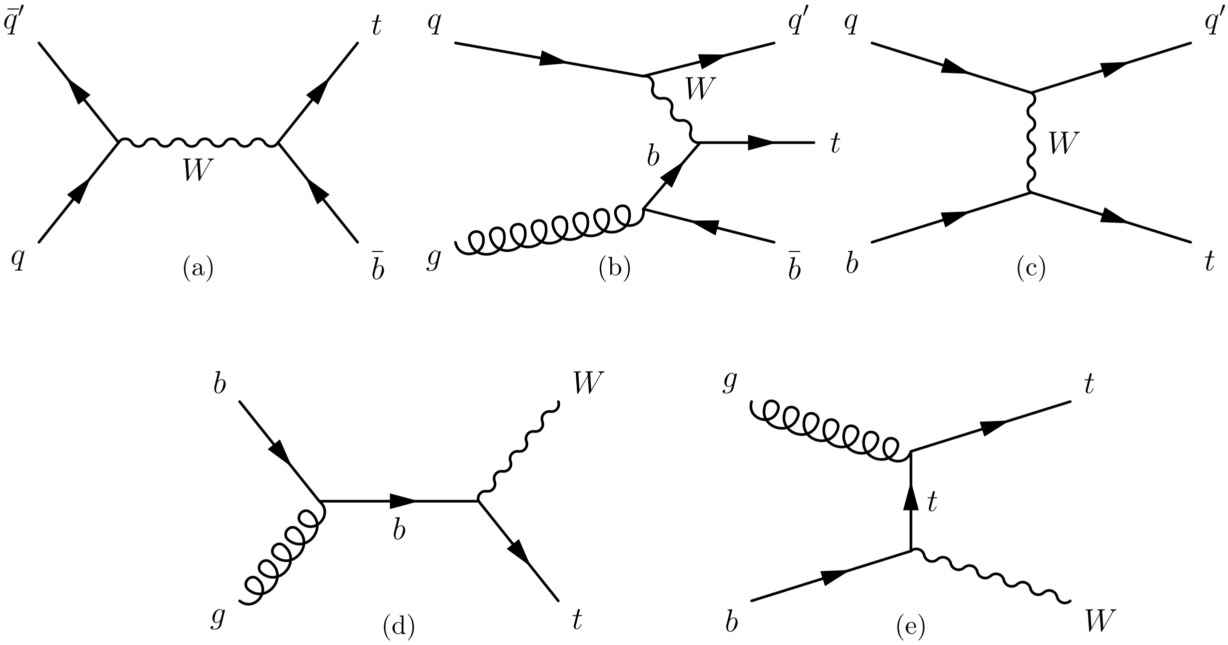


Figure 2.4: Feynman diagrams for leading order single top production: (a) s-channel, (b) and (c) t-channel, (d) and (e) tW-channel.

2.2.2 Top quark decay

The top quark predominantly decays to a W boson and a b-quark: $\mathrm{t} \to \mathrm{Wb}$. Other possible decay modes ($\mathrm{t} \to \mathrm{Ws}$ and $\mathrm{t} \to \mathrm{Wd}$) are suppressed in the Standard Model by the unitarity requirement of the CKM matrix, which yields [8] the $|V_{\mathrm{tb}}|$ value of

$$|V_{\mathrm{tb}}| = 0.999146^{+0.000021}_{-0.000046}. \tag{2.38}$$

As already mentioned, the direct measurement of $|V_{\mathrm{tb}}|$ (without assuming unitarity) can be performed by measuring the single top production cross section. The recent CMS result is consistent with the Standard Model [34]:

$$|V_{\mathrm{tb}}| = 0.998 \pm 0.038 \text{ (experimental)} \pm 0.016 \text{ (theoretical)}. \tag{2.39}$$

The branching fraction of $\mathrm{t} \to \mathrm{Wb}$ decay is assumed to be 100 % hereafter. Therefore, the $\mathrm{t\bar{t}}$ decay modes completely depend on subsequent W boson decays:

- fully hadronic: $\mathrm{t\bar{t}} \to \mathrm{bW^+\ \bar{b}W^-} \to \mathrm{b\ q\bar{q}'\ \bar{b}\ q''\bar{q}'''}$ (45.7 %);
- semileptonic: $\mathrm{t\bar{t}} \to \mathrm{bW^+\ \bar{b}W^-} \to \mathrm{b\ q\bar{q}'\ \bar{b}}\ l^-\bar{\nu}_l + \mathrm{b}\ l^+\nu_l\ \mathrm{\bar{b}\ q''\bar{q}'''}$ (43.8 %);
- dileptonic: $\mathrm{t\bar{t}} \to \mathrm{bW^+\ \bar{b}W^-} \to \mathrm{b}\ \bar{l}\nu_l\ \mathrm{\bar{b}}\ l'\bar{\nu}_{l'}$ (10.5 %).

Percentages in parentheses show the branching fractions corresponding to the decay modes [8]. The quarks in all final states hadronise into jets. The number of jets is not limited to the number of initial quarks due to the contribution from extra QCD radiation (i.e. gluon radiation) either before or after $\mathrm{t\bar{t}}$ production.

The fully hadronic $\mathrm{t\bar{t}}$ decay signature implies at least six jets in the final state and no leptons, hence it is heavily contaminated by the background from QCD and W/Z + jets processes. Dileptonic $\mathrm{t\bar{t}}$ decay has a clean signature, but due to the lowest branching fraction can suffer from limited statistics. The semileptonic decay mode, shown in Figure 2.5, has exactly one highly energetic lepton and at least four jets in the final state, which is therefore called the lepton plus jets final state. The lepton can be an electron, a muon or a τ-lepton; however, since the τ-leptons are difficult to reconstruct, they are usually excluded from top quark analyses studying semileptonic decays. Analyses in this thesis study the electron plus jets (e+jets) and muon plus jets (μ+jets) decay channels, which have branching fractions of about 14 % each.

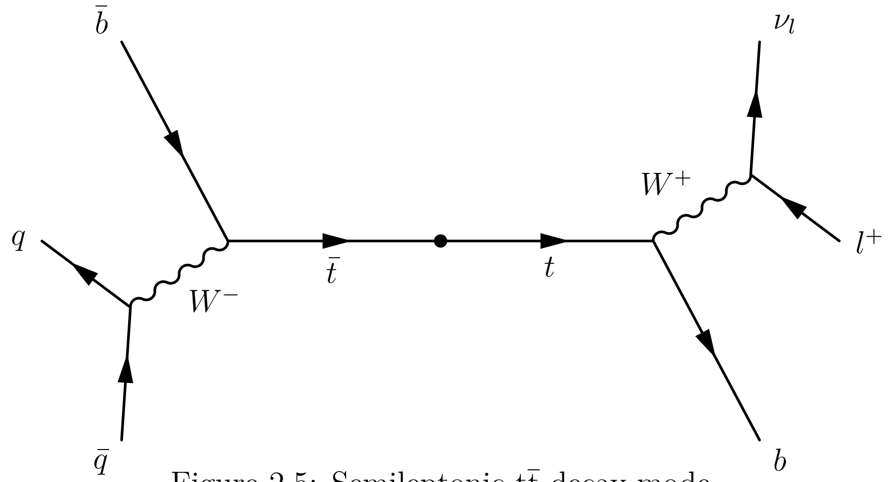


Figure 2.5: Semileptonic $\mathrm{t\bar{t}}$ decay mode

2.2.3 Background processes to semileptonic top quark pair decay

Both top quark mass and differential cross section analyses presented in this thesis are based on a selection of events with semileptonic (e/μ+jets) $\mathrm{t\bar{t}}$ decay signature. Being relatively clean yet still providing sufficient event yields, this decay mode is used in many measurements of top quark properties. Good understanding of background contributions is very important for accuracy and reliability of measurements. The most significant background processes to semileptonic $\mathrm{t\bar{t}}$ decay are described below.

Single top quark background

One of the less significant backgrounds to lepton plus jets analyses is the single top quark production. Major production modes for this process (mentioned in Section 2.2.1) have a lower number of jets in the final state than that of the $\mathrm{t\bar{t}}$ production, however, additional jets can emerge from initial and final state radiation (ISR/FSR), i.e. soft gluon radiation before and after the hard scattering process. This contribution is not negligible, and is modelled using Monte Carlo simulation.

$\mathrm{W}+$ jets background

Production of W bosons with additional jets (called $\mathrm{W}+$ jets) is one of the most significant backgrounds to semileptonic $\mathrm{t\bar{t}}$ decays. Example Feynman diagrams for leading order W boson production are shown in Figure 2.6 (a, b). Leptonic decays in the $\mathrm{W}+$ jets process together with extra jets can successfully mimic the $\mathrm{t\bar{t}}$ decay since it also includes two W bosons. However, leptons and jets in $\mathrm{W}+$ jets processes are usually much softer than those coming from $\mathrm{t\bar{t}}$ decay, due to the large mass of

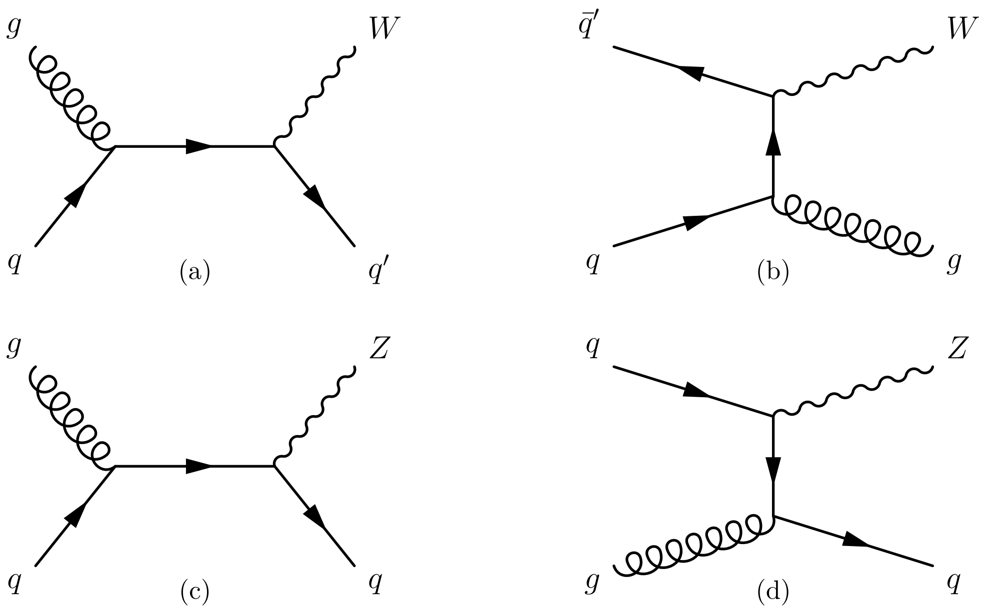


Figure 2.6: Example diagrams for leading order W boson (a, b) and Z boson (c, d) production at the LHC.

top quarks. Moreover, additional jets are less likely to be b-jets which are necessarily present in $\mathrm{t\bar{t}}$ decay. Therefore, tagging of b-jets and high kinematic cuts on lepton and jets transverse momenta can significantly reduce the W + jets background. Besides, the production cross section of W + jets events reduces exponentially with increasing number of jets [35], therefore jet multiplicity cuts also help to mitigate this background. These methods are detailed in the selection procedures of top quark mass and differential cross section analyses in Sections 5.2 and 6.2, respectively.

Z + jets background

Z boson production processes with additional jets (Z + jets) are largely similar to W + jets processes, as can be seen in Figure 2.6 (c, d). The difference lies in leptonic decays of the Z boson, as it produces two opposite-sign leptons. Vetoing the second lepton in the selection is therefore particularly helpful in reducing this background. Contamination can still occur if one of the leptons is misreconstructed as a jet or does not pass identification criteria. This contribution is estimated to be small, and together with the jet multiplicity cuts the Z + jets background is usually mitigated significantly.

QCD multi-jet background

Standard QCD processes in hadron collisions, such as gluon fusion and quark-antiquark annihilation, usually result in the production of highly energetic jets. Examples of these processes in the lowest order with two jets in the final state are

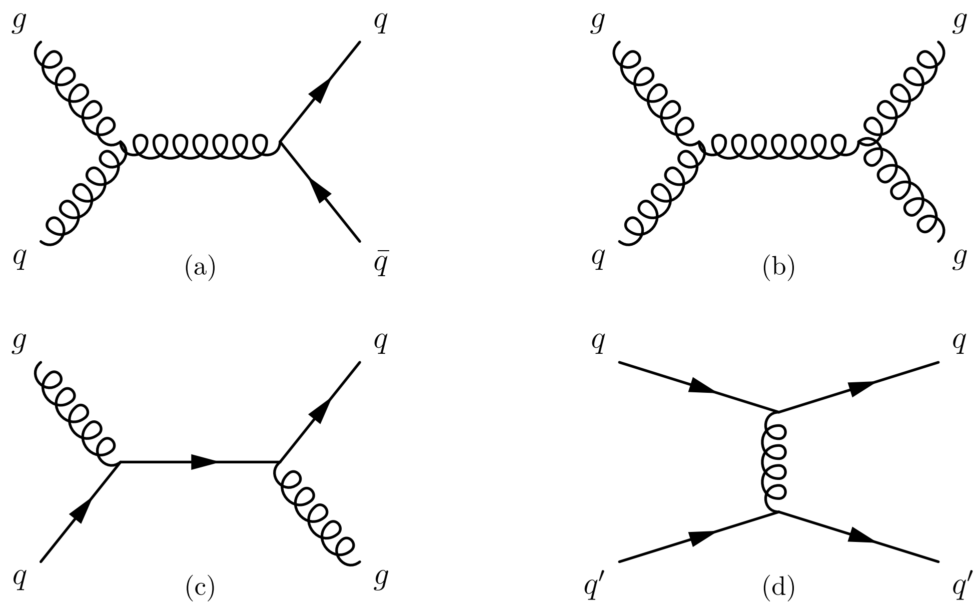


Figure 2.7: Example diagrams for leading order QCD multi-jet production at the LHC

shown in Figure 2.7. Additional jets arise at higher orders and can potentially fake $\mathrm{t\bar{t}}$ signal. Although QCD events do not produce real prompt leptons, this can happen when jets are mis-identified as leptons. These objects are referred to as fake leptons. Due to strict identification criteria of real electrons and muons (described in Sections 3.5.1 and 3.5.2), such mis-identifications are extremely rare. However, since the production cross section of QCD events is very large, this background becomes non-negligible.

QCD multi-jet background is usually less significant in the muon plus jets channel, since jets faking muons are easier to recognise. For instance, when a highly energetic jet punches through the calorimetry system into the muon chambers and thus becomes a fake muon, it can be removed by isolation criteria as it also leaves a significant energy deposit in the calorimeters. Muons (and electrons) can also arise from heavy flavour decays of b and c quarks – these are in fact real leptons, but as they don't originate at the interaction point, they can be removed by track quality cuts most of the time.

In the electron plus jets channel, the QCD multi-jet background is more prominent. This is mainly due to photon conversions, which occur when photons convert into electron-positron pairs. Photons can arise from prompt production (γ+jets), particle decays or bremsstrahlung radiation. Different methods for identifying such conversions have been developed, which are discussed in detail in Section 3.5.1.3.

Due to the large contribution from higher-order processes in the signal region, the QCD background is very hard to model. Large uncertainties in theoretical cross section and potential mismodelling of higher-order effects may lead to biases in kinematic distributions based on Monte Carlo simulation, as well as in overall acceptance of QCD events passing the signal selection. Therefore, data-driven methods of QCD background estimation are often preferable (see Section 6.3).

Additional backgrounds

Other background processes potentially mimicking semileptonic $\mathrm{t\bar{t}}$ decay include the Drell-Yan process with dilepton production ($q\bar{q} \to \mathrm{Z}/\gamma^* \to l\bar{l}$), and diboson production. The background contamination from these processes can be reduced by vetoing the second lepton, and is generally considered negligible due to the relatively small production cross sections of these processes.

2.2.4 Top quark mass

The top quark mass is a fundamental parameter of the Standard Model. However, its precise definition is a very non-trivial problem, ultimately subject to convention. One of the most popular definitions is the pole mass, corresponding to the real part of a pole in the perturbative propagator of a particle [36]. For the top quark, the perturbative propagator with a four-momentum p has a pole at [37]:

$$\sqrt{p^2} = m_\mathrm{t} - \frac{i}{2}\Gamma_\mathrm{t}, \tag{2.40}$$

where m_t is the pole mass and $\Gamma_\mathrm{t} \approx 1.5$ GeV is the top quark decay width. This definition yields a peak in the invariant mass distribution of the top quark decay products (i.e. W boson and b-quark), which is accessible experimentally. As $m_\mathrm{t} \gg \Gamma_\mathrm{t}/2$, the peak approximately corresponds to the pole mass, therefore it is usually assumed that experimentally measured top quark mass represents the pole mass. As will be discussed later, there are a few problems with this approach.

It has been shown [38] that the pole mass is infrared finite[1] and gauge-invariant to all orders of perturbative QCD. Despite being well-defined in perturbation theory, an all-order resummation of a particular class of diagrams known as "infrared renormalons" [39, 40] leads to an intrinsic ambiguity of the top quark pole mass proportional to the energy scale of the strong interaction $\Lambda_\mathrm{QCD} \approx 200$ MeV [41].

Another frequently used mass definition which is free from renormalon ambiguity is the renormalised mass $\overline{m}_\mathrm{t}(\mu)$ in the minimal subtraction ($\overline{\mathrm{MS}}$) renormalisation scheme. This mass definition is often called a running mass, since it depends on the

[1]Infrared finiteness corresponds to the absence of divergences due to contributions of terms with energy apporaching zero.

renormalisation scale μ which is in principle arbitrary. For a choice of $\mu = \overline{m}_t$, the $\overline{\text{MS}}$ mass $\overline{m}_t$ is related to the pole mass m_t as follows [37]:

$$m_t = \overline{m}_t(\overline{m}_t)\left(1 + \frac{4}{3}\frac{\overline{\alpha}_S(\overline{m}_t)}{\pi} + 8.28\left(\frac{\overline{\alpha}_S(\overline{m}_t)}{\pi}\right)^2 + ...\right) + \mathcal{O}(\Lambda_{\text{QCD}}). \tag{2.41}$$

The difference between the pole mass and the $\overline{\text{MS}}$ mass is then estimated to be on the order of 10 GeV (with $m_t > \overline{m}_t$). However, it has recently been shown [41] that including electroweak corrections for a Higgs boson mass of ~ 125 GeV mostly cancels the corresponding QCD corrections, reducing the $m_t - \overline{m}_t$ difference down to ~ 1 GeV. Theoretical research such as studying higher-order corrections is actively ongoing in this field.

It has been argued that the top quark mass measured at hadron colliders is in fact a parameter of Monte Carlo generators, which does not necessarily correspond to the pole mass. This parameter is determined by (next to) leading order calculations of the matrix elements of hard processes (see Section 3.4), and this mass definition does not absorb any corrections from parton showers and hadronisation. Therefore, the actual mass definition measured experimentally remains not fully understood, with a conceptual uncertainty on the order of 1 GeV [42].

Nevertheless, experimental measurements of the top quark mass are of paramount importance. Continuously increasing precision of the measurements is challenging the current definition of the measured mass, undoubtedly triggering more effort in reaching a more precise theoretical specification of the mass parameter used in Monte Carlo simulation. As the top quark is a major background to many new physics searches beyond the Standard Model, deep understanding of its mass parameter in simulation as well as pushing down the corresponding theoretical and experimental uncertainties are extremely important.

Another strong motivation for precise top quark mass measurement is its importance in the determination of the electroweak vacuum stability. Measurements of both the Higgs boson and top quark masses allow extrapolating the Standard Model Higgs potential (Equation 2.1) up to Planck-scale energies. By performing a NNLO renormalisation procedure and using the latest experimental values of m_t and M_H, it has been shown [43, 44] that there is a significant preference for meta-stability of the SM electroweak vacuum, as illustrated in Figure 2.8. The meta-stability corresponds to the regime where the true minimum of the scalar potential is lower than the standard electroweak minimum, which has a lifetime exceeding the age of the Universe. Currently the dominant uncertainty in evaluating the vacuum stability is experimental and mostly comes from the top quark mass measurement, which provides further motivation for improved m_t measurements.

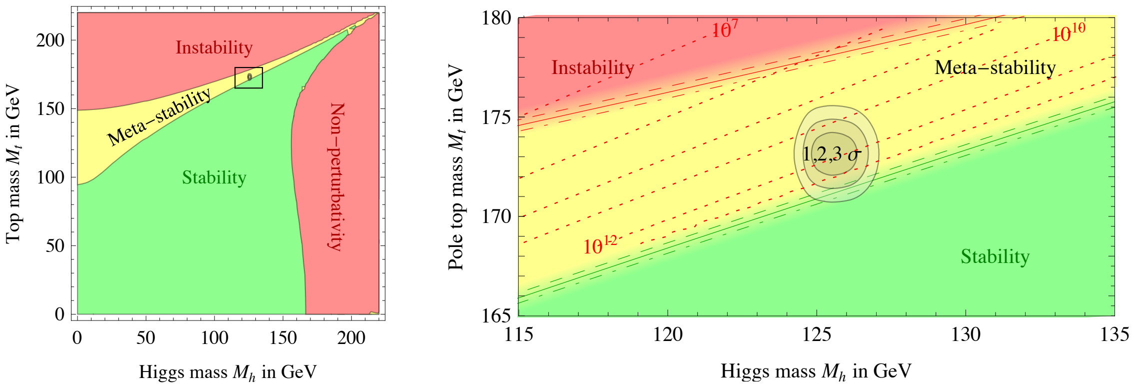


Figure 2.8: Regions of absolute stability, meta-stability and instability of the SM vacuum in the m_t/M_H plane (left) and zoom into the region of the preferred experimental mass ranges. Gray areas show the allowed region at 1, 2 and 3 σ [43].

Nearly 20 years of the Tevatron operation resulted in a very good understanding of the machine and detectors, leading to an impressively small systematic uncertainty on the top quark mass measurement. The latest combination result from the Tevatron gives [45]:

$$m_t = 173.20 \pm 0.51\ (\text{stat.}) \pm 0.71\ (\text{syst.})\ \text{GeV}. \tag{2.42}$$

The LHC experiments have reached nearly as good precision, with a statistical uncertainty being expectedly smaller than that of the Tevatron, however, with a larger systematic uncertainty due to still improving understanding of detector effects. The latest LHC combination result yields [46]:

$$m_t = 173.29 \pm 0.23\ (\text{stat.}) \pm 0.92\ (\text{syst.})\ \text{GeV}. \tag{2.43}$$

Recently, the first world combination result of both LHC and Tevatron measurements has become available [11]:

$$m_t = 173.34 \pm 0.27\ (\text{stat.}) \pm 0.71\ (\text{syst.})\ \text{GeV}, \tag{2.44}$$

which corresponds to a total uncertainty on the top quark mass of 0.76 GeV, or 0.44 %. All these results compared with contributions of individual mass measurements are shown in Figure 2.9.

At a proposed linear electron-positron collider (International Linear Collider, ILC), the uncertainty on the top quark mass is expected to fall below 200 MeV [47], i.e. the energy scale of the strong interaction Λ_{QCD}. It will clearly challenge the theoretical understanding of different mass schemes, ultimately leading to more vigorous tests of the SM and potentially providing sensitivity to new physics models.

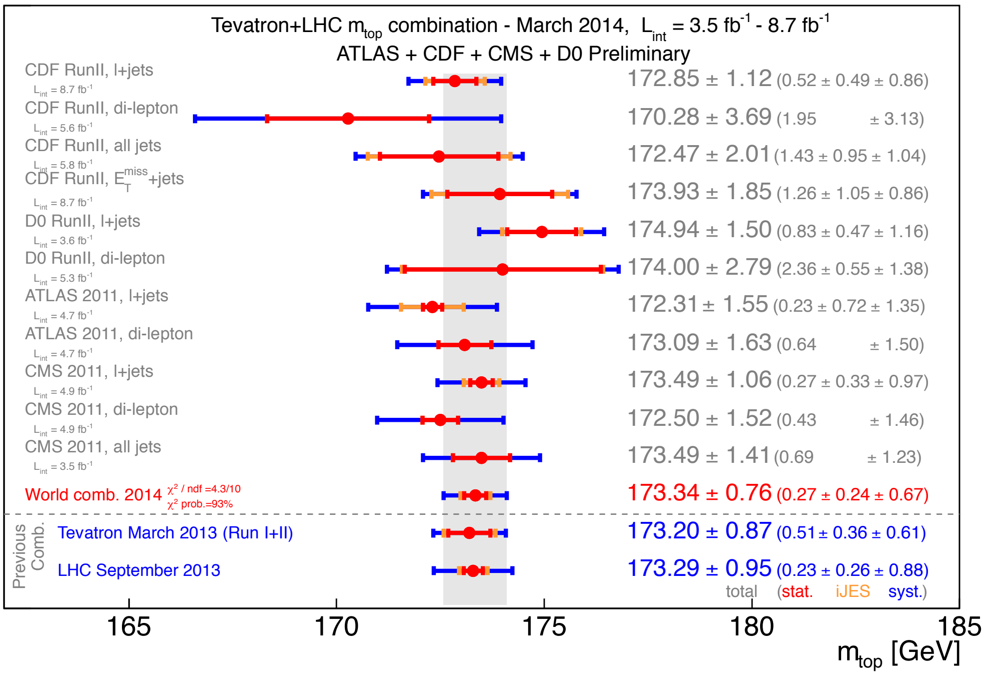


Figure 2.9: Top quark mass measurements and the result of their combination, compared with the Tevatron and LHC combinations. Each measurement is shown with the total uncertainty, the statistical, the jet energy scale (iJES, when applicable), and the systematic uncertainty. The iJES contribution is statistical in nature and applies only to analyses performing *in situ* jet energy calibration procedures. The grey vertical band shows the total uncertainty on the combined value of the top quark mass [11].

2.2.5 Top quark pair production cross section

The $\mathrm{t\bar{t}}$ production is a strong interaction process, and therefore is described by perturbative QCD. Hadron collisions at the LHC (or other hadron colliders) are best viewed as interactions between their constituent quarks and gluons, also referred to as partons. Since collisions occur at high energies, these interactions result in hard scattering processes between incoming partons, meaning those involving a significant momentum transfer comparing to the proton mass, and potentially giving rise to highly energetic final states like top quarks. Each incoming parton carries only a fraction x of the total momentum of a parent hadron. The distribution of momentum fractions for all flavours of partons are described by parton distribution functions (PDFs).

A PDF $f_i(x_i, \mu_f^2)$ is defined as a probability density of finding a parton with flavour i and a longitudinal momentum fraction x_i when probed at momentum scale μ_f^2, which is known as factorisation scale. This parameter separates the hard scattering process into the hard partonic interaction and the soft (or long-range) interaction. The former interaction happens at a short distance, and therefore only involves high-momentum transfer which is calculable in perturbative QCD. The soft long-range part of the interaction, on the contrary, can not be calculated in QCD and instead is parametrised by the PDFs, which have to be obtained from experimental data. Figure 2.10 shows one of the latest PDFs obtained by the CTEQ-TEA collaboration [48].

The total $\mathrm{t\bar{t}}$ production cross section for hard scattering processes in hadron collisions can be calculated to a fixed order as [49, 50]:

$$\sigma^{\mathrm{t\bar{t}}}(s, m_{\mathrm{t}}^2) = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \hat{\sigma}_{i,j\to\mathrm{t\bar{t}}}(\hat{s}, m_{\mathrm{t}}^2, \alpha_S(\mu_r^2)), \tag{2.45}$$

where the indices i, j run over the incoming partons (gluons and quark-antiquark pairs), $x_{i,j}$ are the momentum fractions of the incoming partons, $f_{i,j}$ are their parton distribution functions, and $\hat{\sigma}_{i,j\to\mathrm{t\bar{t}}}$ is the cross section of interacting partons, which depends on the parton centre of mass energy $\hat{s} \sim x_i x_j s$, the top quark mass m_{t}, and the QCD strong coupling constant α_S. The latter parameter is evaluated at renormalisation scale μ_r. The top quark mass m_{t}, as discussed in Section 2.2.4, also depends on the choice of renormalisation scheme and scale μ_r.

The partonic cross section $\hat{\sigma}_{i,j\to\mathrm{t\bar{t}}}$ can be calculated as a power series expansion in α_S to different orders. If the calculation is performed in the lowest order of α_S, it is referred to as the leading order calculation (LO). Such calculations provide a very limited prediction power, since they do not include any additional final-state

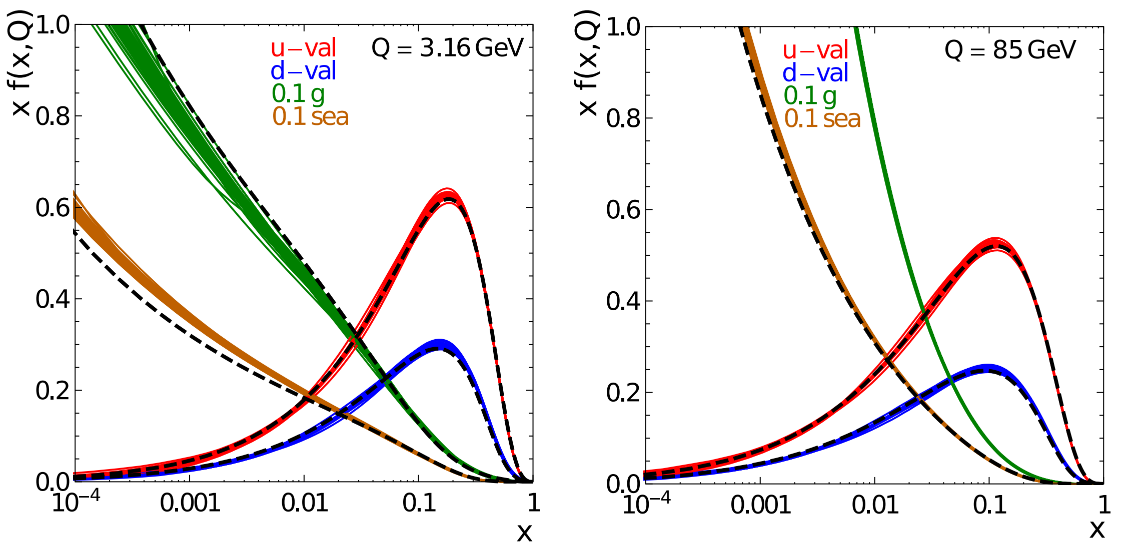


Figure 2.10: CT10 NNLO parton distribution functions for up and down quarks, gluons and sea quarks, given at two factorisation scales: (a) Q = 3.16 GeV, (b) Q = 85 GeV [48].

partons produced in the hard scattering process. Moreover, LO calculations have a large uncertainty from the renormalisation scale ambiguity. Inclusion of higher-order terms in the calculation decreases this uncertainty, and a substantially better level of precision can be obtained with next-to-leading-order (NLO) and next-to-next-to-leading-order (NNLO) calculations, which have recently become available for the $t\bar{t}$ production process.

The results of the most recent theoretical calculation of $t\bar{t}$ cross section performed with NNLO QCD corrections [51] are given in Table 2.3. The cross section has also been measured by CMS and ATLAS experiments at the LHC – the comparison of these measurements with the theoretical predictions at different centre of mass energies is shown in Figure 2.11.

Table 2.3: Theoretical predictions for $t\bar{t}$ production cross section at different LHC centre of mass energies, calculated at NNLO [51]. The scales uncertainty corresponds to the choice of factorisation and renormalisation scales.

$\sqrt{s}$	σ_{total} [pb]	scales [pb]	PDF [pb]
7 TeV	172.0	+4.4 (2.6%) −5.8 (3.4%)	+4.7 (2.7%) −4.8 (2.8%)
8 TeV	245.8	+6.2 (2.5%) −8.4 (3.4%)	+6.2 (2.5%) −6.4 (2.6%)
14 TeV	953.6	+22.7 (2.4%) −33.9 (3.6%)	+16.2 (1.7%) −17.8 (1.9%)

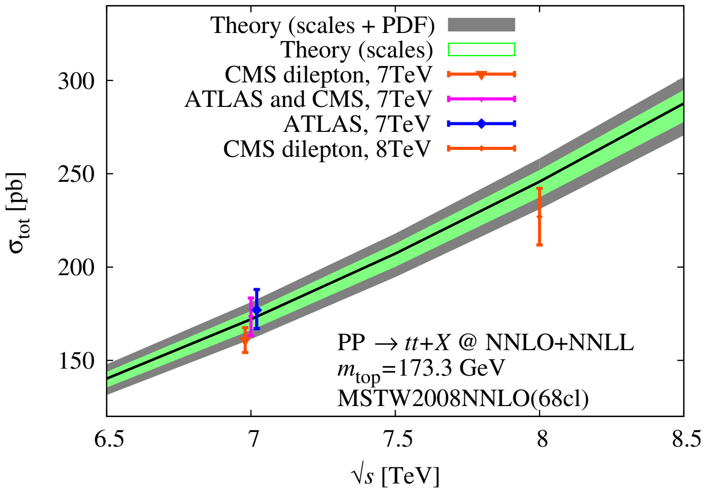


Figure 2.11: Measurement of the $\mathrm{t\bar{t}}$ production cross section at centre of mass energies of 7 TeV and 8 TeV by ATLAS and CMS experiments, compared with the theoretical predictions [51].

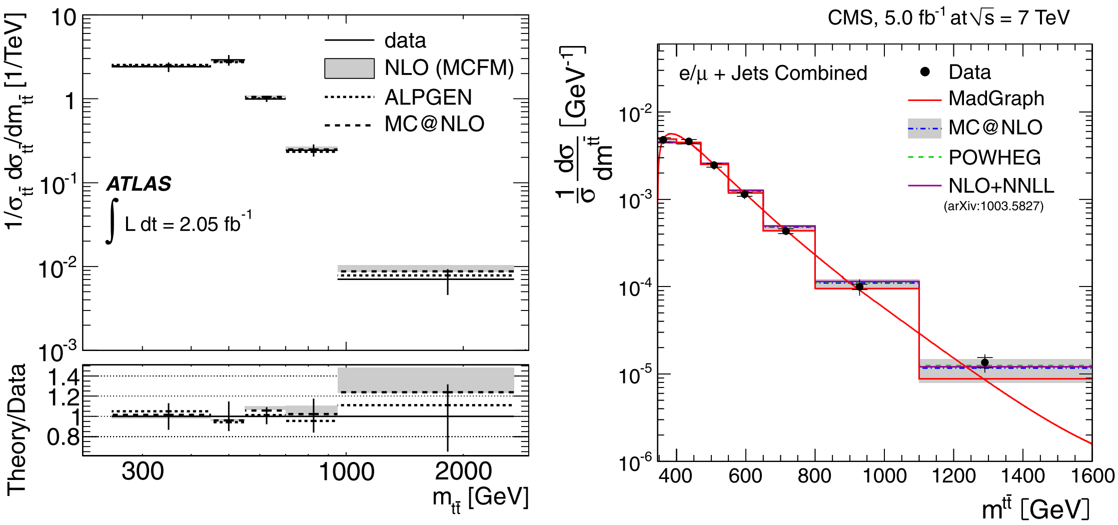
An abundance of top quark events at the LHC allows measurements of differential $\mathrm{t\bar{t}}$ production cross section with respect to many different variables. These measurements provide important tools for validation of various MC models and higher-order QCD calculations of $\mathrm{t\bar{t}}$ production. Since the very same models are used in BSM searches where $\mathrm{t\bar{t}}$ events represent a major background, such validation is of great importance. Moreover, these measurements can themselves be sensitive to signal contributions from new physics.

Normalised differential cross section in each bin i of an observed variable X is given by:

$$\frac{1}{\sigma_{\mathrm{t\bar{t}}}^{\mathrm{tot}}}\frac{\mathrm{d}\sigma_i}{\mathrm{dX}} = \frac{1}{\sigma_{\mathrm{t\bar{t}}}^{\mathrm{tot}}}\frac{x_i}{\Delta\mathrm{X}_i\mathcal{L}}, \tag{2.46}$$

where x_i is the number of signal events in data after subtraction of background, corrected for detector efficiency, acceptance and migration between the bins of the variable X; $\mathcal{L}$ is the total integrated luminosity; $\Delta\mathrm{X}_i$ is the variable bin width; and $\sigma_{\mathrm{t\bar{t}}}^{\mathrm{tot}}$ is the total $\mathrm{t\bar{t}}$ production cross section. The normalisation is performed in order to cancel the systematic uncertainties that are correlated between the bins.

Both ATLAS and CMS collaborations have delivered a range of differential $\mathrm{t\bar{t}}$ production cross section measurements [52, 53]. For instance, Figure 2.12 shows the normalised differential cross section with respect to the invariant mass of $\mathrm{t\bar{t}}$ pair ($m_{\mathrm{t\bar{t}}}$), measured by ATLAS and CMS collaborations using the $\sqrt{s} = 7$ TeV LHC data. These measurements require a full kinematic reconstruction of the semileptonic $\mathrm{t\bar{t}}$ decay in order to accurately calculate the $m_{\mathrm{t\bar{t}}}$ variable. Other variables requiring the kinematic reconstruction include the transverse momenta of top quarks decaying either leptonically or hadronically, their rapidities, etc.



(a)

(b)

Figure 2.12: Normalised differential $t\bar{t}$ production cross section measurements in the l+jets channel with respect to the invariant mass of a top quark pair. (a) ATLAS measurement [52], (b) CMS measurement [53].

In this thesis, the $t\bar{t}$ differential cross section is measured with respect to so-called event-level (or global) variables, not requiring a kinematic reconstruction of the $t\bar{t}$ decay – for example, missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) or total hadronic activity (H_{T}) in the event (see Section 6.5.1). The main advantage of these measurements is the absence of the systematic uncertainty associated with the kinematic reconstruction. Additionally, signs of new physics scenarios could be revealed in the tails of these distributions. For example, an associated production of a $t\bar{t}$ pair with some new resonance ($t\bar{t}$ + X) decaying invisibly can show up in the tail of the missing transverse energy distribution.

2.3 Summary

In this chapter, an overview of the Standard Model has been presented. The concept of the gauge principle and its applications in the electroweak theory and the Quantum Chromodynamics, as well as the mechanism of the electroweak symmetry breaking have been briefly described. Shortcomings of the Standard Model and some models proposing the solutions have been mentioned. Finally, an introduction to top quark physics has been given, pointing out the relevance and importance of the top quark mass and differential cross section measurements presented in this thesis.

3. The LHC and the CMS detector

3.1 The Large Hadron Collider

The LHC [54] is currently the largest and the most powerful particle accelerator ever built. It is installed in the 26.7 km tunnel that was originally constructed for the LEP accelerator in the 1980s. The tunnel lies at a depth of 45 m to 170 m underground between the Jura mountain and Lake Geneva, being the main part of the CERN accelerator complex.

The machine is designed to accelerate proton beams and provide collisions at a centre of mass energy of $\sqrt{s} = 14$ TeV. Unlike particle-antiparticle colliders, the LHC requires two rings with opposite magnetic dipole fields in order to maintain and collide two counter-rotating proton beams. Since the tunnel was originally designed for the electron-positron LEP, it has an internal diameter of 3.7 m which is not enough to install two separate independent rings. Therefore, a twin-bore magnet design was adopted [55], which resulted in substantial cost savings.

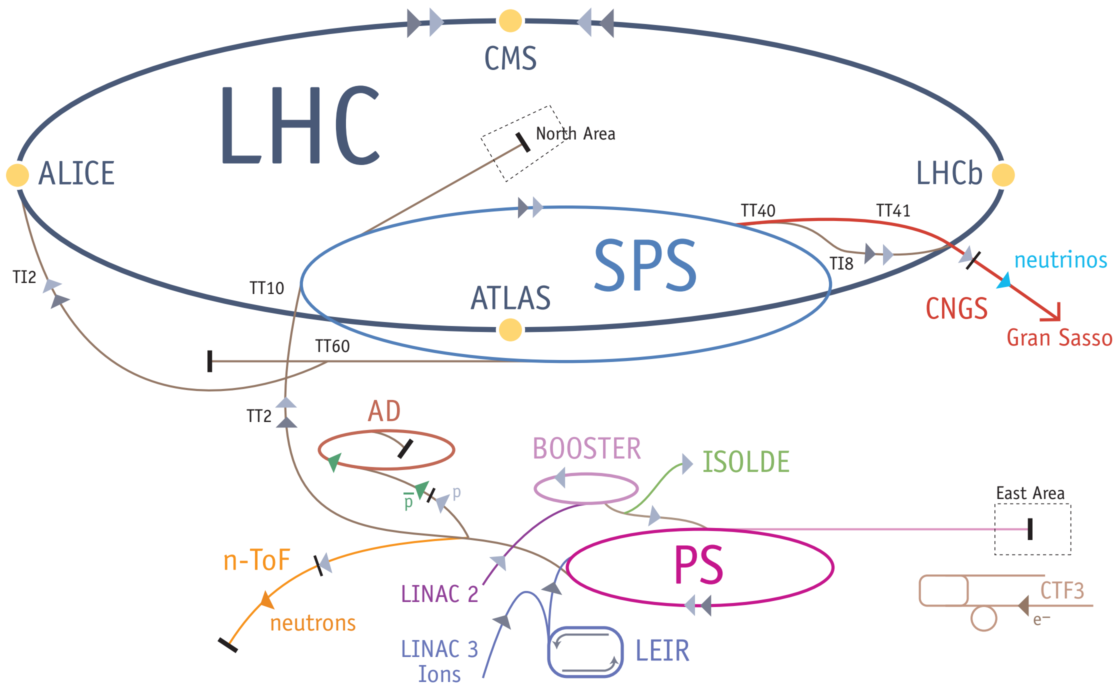


Figure 3.1: CERN accelerator complex

A schematic view of the LHC accelerator chain is shown in Figure 3.1. Initially, the protons are obtained by stripping orbiting electrons from hydrogen atoms. Then they are injected into the linear accelerator LINAC2 to reach the energy of 50 MeV and enter the Proton Synchrotron Booster (PSB). The booster accelerates them to 1.4 GeV and passes the beam to the Proton Synchrotron (PS) where the energy rises to 25 GeV. In the next step, protons enter the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Finally, the beam is transferred to the LHC in both clockwise and anti-clockwise directions where it takes about 20 minutes to reach the design 7 TeV energy (per beam).

The LHC has four interaction points, providing collisions to four major experiments. Two of them, CMS and ATLAS, are multi-purpose high-luminosity experiments with a peak luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}. The other two experiments operate at low luminosities and have more specific physics goals: LHCb studies b-meson decays, and Alice is a dedicated heavy ion experiment.

The instantaneous luminosity of a collider can be calculated as

$$\mathcal{L} = \frac{n_1 n_2 n_b f}{4\pi \sigma_x \sigma_y}, \tag{3.1}$$

where n_1 and n_2 are the numbers of particles in each of the colliding bunches, n_b is the number of bunches per beam, f is the revolution frequency, σ_x and σ_y are the horizontal and vertical beam sizes, assuming the two beams have the same size. These and other performance-related parameters of the LHC for different running conditions are shown in Table 3.1.

The number of events generated in the collisions per second is given by

$$N_{events} = \mathcal{L} \times \sigma, \tag{3.2}$$

where σ is the cross section of the process under study. Cross sections and production rates for several different processes as a function of the centre of mass energy are shown in Figure 3.2.

Table 3.1: LHC beam parameters [56, 57]. Transverse beam size and peak luminosity quoted at interaction points 1 and 5 (ATLAS and CMS detectors).

Parameter	2011 run	2012 run	Design value
Beam energy (TeV)	3.5	4	7
Maximum number of bunches	1380	1380	2808
Number of particles per bunch	1.45×10^{11}	1.7×10^{11}	1.15×10^{11}
Bunch spacing (ns)	75/50	50	25
Revolution frequency (kHz)	11.245	11.245	11.245
Transverse beam size (µm)	25.9	18.8	16.7
Peak luminosity ($cm^{-2}\,s^{-1}$)	3.7×10^{33}	7.7×10^{33}	10^{34}

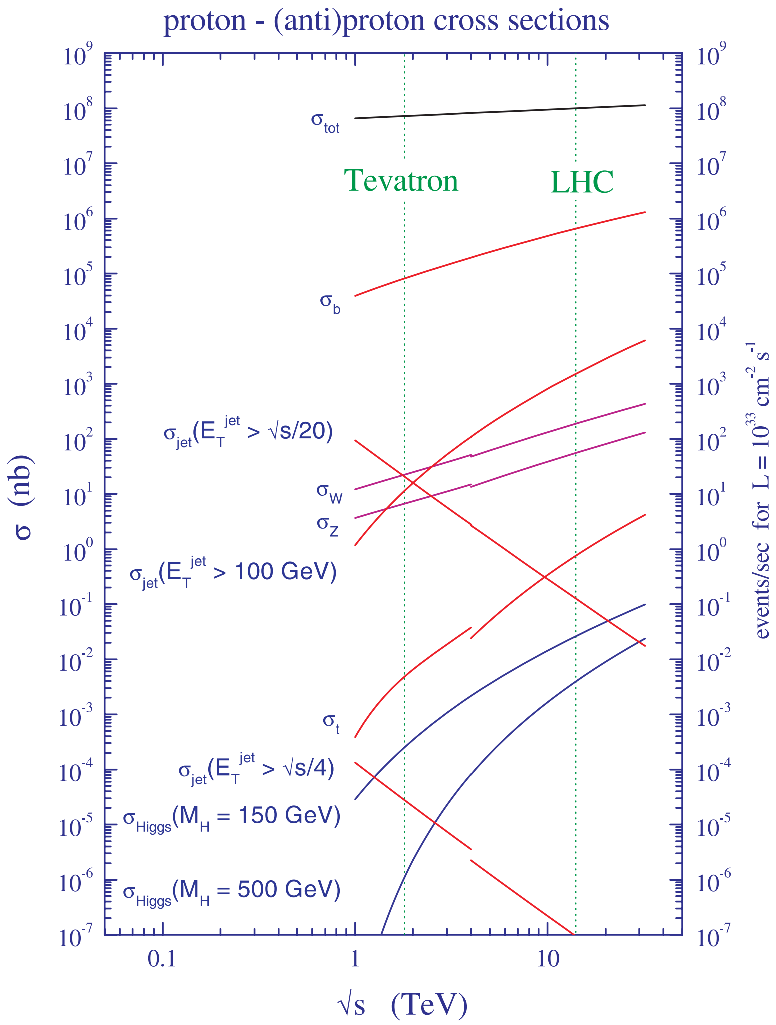


Figure 3.2: Production cross sections for different processes as a function of the centre of mass energy $\sqrt{s}$. The discontinuity is due to extrapolation between proton-antiproton Tevatron and proton-proton LHC. The $t\bar{t}$ production cross section is denoted by σ_t [50].

The LHC started operating on 10 September 2008 with the first beams fully circulating in both rings. However, only nine days later a magnet quench occurred in two sectors of the tunnel, which was caused by an electrical fault due to a bad connection between two magnets. A consequent liquid helium explosion damaged a total of 53 superconducting magnets. Over a year was spent on repairs and tests, and the first collisions were recorded on 23 November 2009 at a centre of mass energy of 0.9 TeV. The following few months showed the continuous ramp up of the beam energies up to 3.5 TeV per beam which was achieved on 30 March 2010 when the LHC physics programme started.

Throughout the rest of 2010, the two general-purpose LHC experiments (CMS and ATLAS) recorded approximately 40 pb^{-1} of data, which resulted in the first measurements of various physics processes at the LHC. The following year became the main 7 TeV data-taking period, with about 5 fb^{-1} of data recorded by ATLAS and CMS. On 5 April 2012 the centre of mass energy was increased to 8 TeV, and July of 2012 marked the first major discovery of a new boson which was later shown to be consistent with the Standard Model Higgs boson, according to approximately 21.8 fb^{-1} of data recorded until early 2013. A long shut-down is planned for the following two years with various upgrades scheduled. The next physics run is expected in 2015 with the beam energy increased up to 6 or 7 TeV.

3.2 The CMS Detector

The Compact Muon Solenoid [58] is a general-purpose detector designed to carry out precise measurements of the Standard Model and searches for physics beyond it. The primary design requirement was the ability to discover the nature of electroweak symmetry breaking, and the first observation of a Higgs boson was obtained in the Summer of 2012 [10].

The detector is installed at one of the LHC interaction points (Point 5) at about 100 m underground near the French village of Cessy, between the Jura mountains and Lake Geneva. The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12 500 t.

The sectional view of CMS is shown in Figure 3.3. In the centre of the detector, tracking and calorimetry systems are surrounded by the superconducting solenoid. On the outermost part of it the magnetic flux is returned through the iron yoke in which the muon system is also integrated. All the subsystems are discussed in the following sections in more detail.

The cylindrical shape of the CMS detector dictates using a cylindrical coordinate system, with the origin centred at the interaction point, the x-axis pointing towards the centre of the LHC ring, the y-axis pointing upwards and the z-axis pointing along

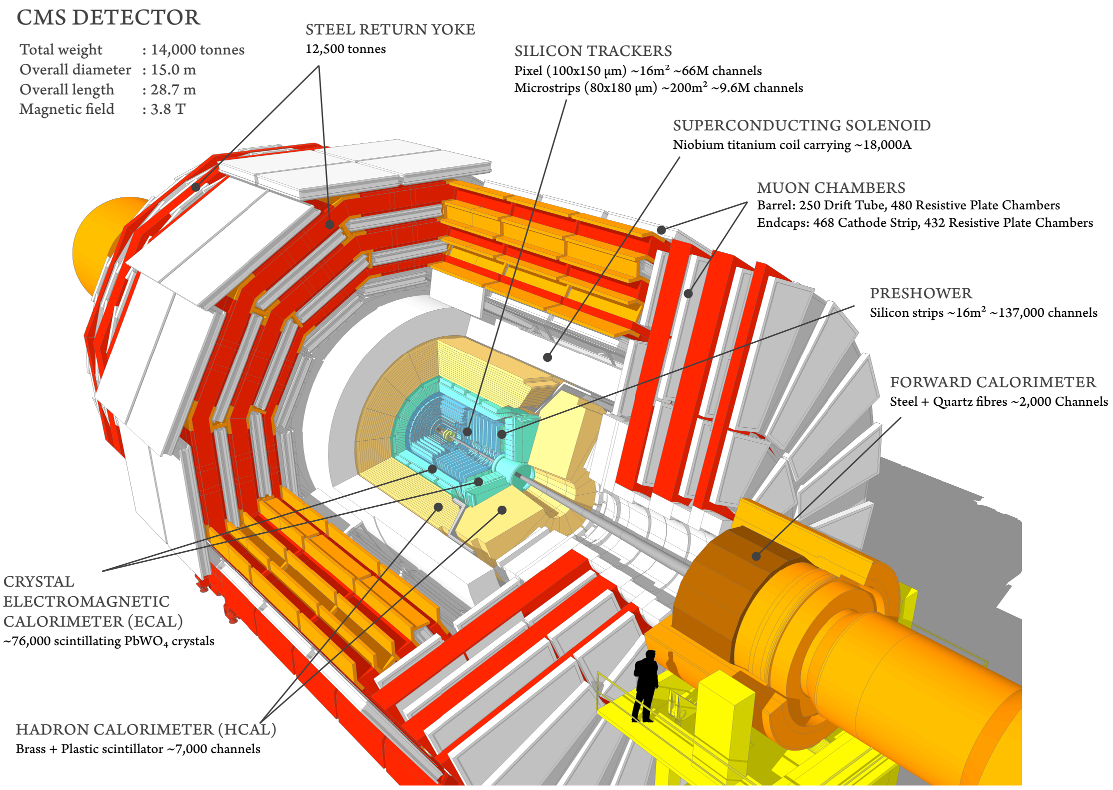


Figure 3.3: Sectional view of the CMS detector

the beamline in the anti-clockwise direction. The azimuthal angle ϕ is measured from the x-axis in the transverse $(x-y)$ plane and the polar angle θ is measured from the z-axis. The radial distance to the beamline is denoted by r. Pseudorapidity is defined as:

$$\eta = -\ln \tan \frac{\theta}{2}. \tag{3.3}$$

This implies that the particles moving in the transverse plane (perpendicular to the beamline) have a pseudorapidity of 0, whereas the beam direction has an infinite pseudorapidity. Considering the cylindrical shape of the detector, it has barrel and endcap regions, with the transition occurring at $\eta \sim 1.4$. The momentum and energy transverse to the beamline are denoted by p_T and E_T respectively; the imbalance of the energy measured in the transverse plane, called missing transverse energy, is denoted by $E_{\mathrm{T}}^{\mathrm{miss}}$.

The distance between any two objects (say, i and j) in the detector is often described in terms of the ΔR quantity, defined as follows:

$$\Delta R = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}. \tag{3.4}$$

The same quantity will be used to define cones in jet clustering algorithms, described in Section 3.5.3.

3.2.1 Inner Tracking System

The tracking system lies in the heart of the CMS detector and is the closest to the interaction point where the particle flux has the highest value. This imposes demanding requirements on the configuration of the system. At design luminosity of $\mathcal{L} = 10^{34}$ cm^{-2} s^{-1} with the bunch spacing of 25 ns, an average of 1000 particles from about 25 proton-proton interactions (pile-up vertices) is expected to traverse the tracker for each bunch crossing. However, up until the long shutdown a bunch spacing of 50 ns was used, which meant a higher number of protons in each bunch leading to approximately twice the number of pile-up vertices. Therefore, in order for the particle tracks to be identified reliably and separately for each bunch crossing, the tracker requires very fine granularity and fast response parameters. Another complication caused by the intense particle flux is the severe radiation damage, so the tracker has to be highly resilient in operating in the harsh environment for a reasonable lifetime.

To meet these requirements on granularity, response time and radiation resilience, the tracker design was chosen to be based on silicon detector technology. Although capable of meeting such conditions, this technology has a disadvantage of a high power density of on-detector electronics. This implies the necessity of an efficient cooling system. Moreover, a large amount of dense material interacting with the particles leads to higher multiple scattering, bremsstrahlung, photon conversions and nuclear interactions. Therefore, there are complications in the reconstruction of the tracks, meaning some loss of efficiency and precision. This will be discussed in detail later on in the object reconstruction section.

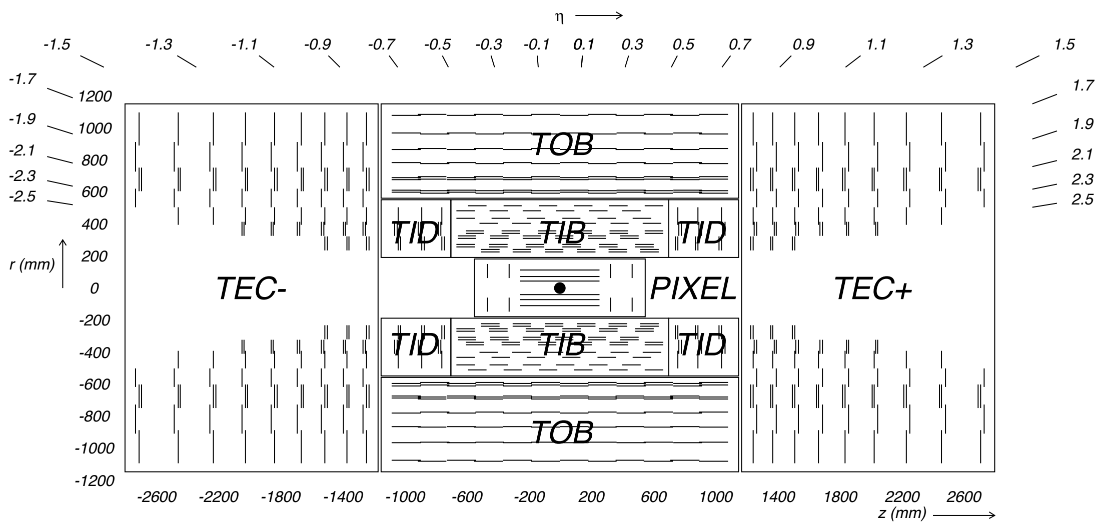


Figure 3.4: Cross section of the CMS tracker system [58]

Figure 3.4 shows the overall layout of the tracking system. It consists of the inner pixel detector, located in the vicinity of the interaction point, and silicon strip tracker detectors: inner barrel and disks (TIB and TID), outer barrel (TOB) and endcaps (TEC). The geometrical acceptance of the tracker system goes up to $|\eta|< 2.5$. The outer radius of the CMS tracker reaches approximately 110 cm, and its total length is about 540 cm.

The pixel detector consists of three layers of pixel sensors at radii of 4.4 cm, 7.3 cm and 10.2 cm from the beamline in the barrel region. In addition there are two endcap disks on each side at $|z| = 34.5$ cm and 46.5 cm. The pixel size equals 100×150 µm^2 in $r\phi \times z$ coordinates. The pixel detector has 66 million pixels and the total area of about 1 m^2.

The silicon strip tracker consists of several layers of silicon microstrip detectors. It covers the region between 20 cm to 110 cm in radius and extends up to ±280 cm in the z direction. The Tracker Inner Barrel (TIB) is made out of 4 layers and the Tracker Outer Barrel (TOB) has 6 layers in it. The tracker endcaps (TEC) comprise 9 disks, and there are also the tracker inner disks (TID) that consist of 3 disks filling the gap between TIB and TEC as shown in Figure 3.4. There are 9.3 million silicon strips covering the area of about 200 m^2. The silicon sensors' thickness varies between 320 and 500 µm and the strip pitch varies from 80 µm in the TIB to 180 µm in TOB and TEC.

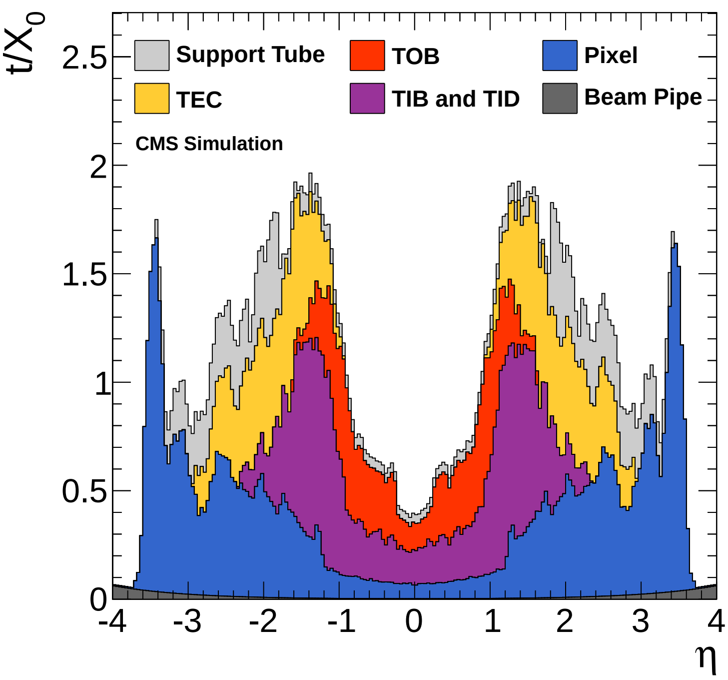


Figure 3.5: Material budget as a function of pseudorapidity η for the different subdetectors of the tracker in units of radiation length [59]

The silicon detectors of the tracker, the readout electronics and support structure form a considerable amount of material for the particles traversing from the interaction point. Figure 3.5 [59] shows the material budget of the CMS tracker in units of radiation lengths[1] (X_0). It grows from about 0.4 X_0 to 1.8 X_0 in the barrel region, and then decreases to about 1 X_0 in the endcaps. This causes a substantial conversion rate for photons and electrons in the tracker material; it also will be discussed in more detail in the electron reconstruction section.

3.2.2 Electromagnetic Calorimeter

The next detector subsystem which is surrounding the tracker is the electromagnetic calorimeter, or ECAL. It is of primary importance for the analyses described in this thesis, as it provides information for the electron and positron reconstruction. Combination of this information with that from the tracking system must ensure a precise measurement of electron position and momentum, and also sufficient background removal. It has to effectively distinguish the energy deposit shape of an electromagnetic particle from the one of a hadronic particle, which requires good segmentation and high resolution.

ECAL is a hermetic, high-granularity, high-resolution scintillating crystal calorimeter consisting of 61 200 lead tungstate ($PbWO_4$) crystals located in the central barrel region ($|\eta| < 1.479$), and 7324 crystals in each of the two endcaps ($1.479 < |\eta| < 3.0$). All crystals are followed by photodetectors reading and amplifying their scintillation: avalanche photodiodes (APD) are used in the barrel, and vacuum phototriodes (VPTs) are used in the endcaps. These different choices were caused by the configuration of the magnetic field and the expected level of radiation.

The layout of the ECAL sub-detector is shown in Figure 3.6. An additional preshower detector is used in the endcap region to lower the required detector depth. Its principal aim is to identify neutral pions in the endcaps, but it also helps to distinguish neutral pions and electrons from minimum ionising particles and improves the position determination of electrons and photons with high granularity.

The main geometrical characteristics of the ECAL crystals are shown in Table 3.2. The choice of lead tungstate was driven by the constraints of the CMS design. It is a very dense material (8.28 g/cm^3) with a short radiation length of $X_0 = 0.89$ cm, which allows the calorimeter to fit inside the compact magnet. Lead tungstate also has a small Molière radius[2] of 2.2 cm, which allows a calorimeter with fine granular-

[1] A material's radiation length is the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung.

[2] The Molière radius R_μ is a characteristic of a material giving the scale of the transverse dimension of the fully contained electromagnetic showers. It is defined as the radius of a cylinder containing an average of 90 % of the shower's energy deposition.

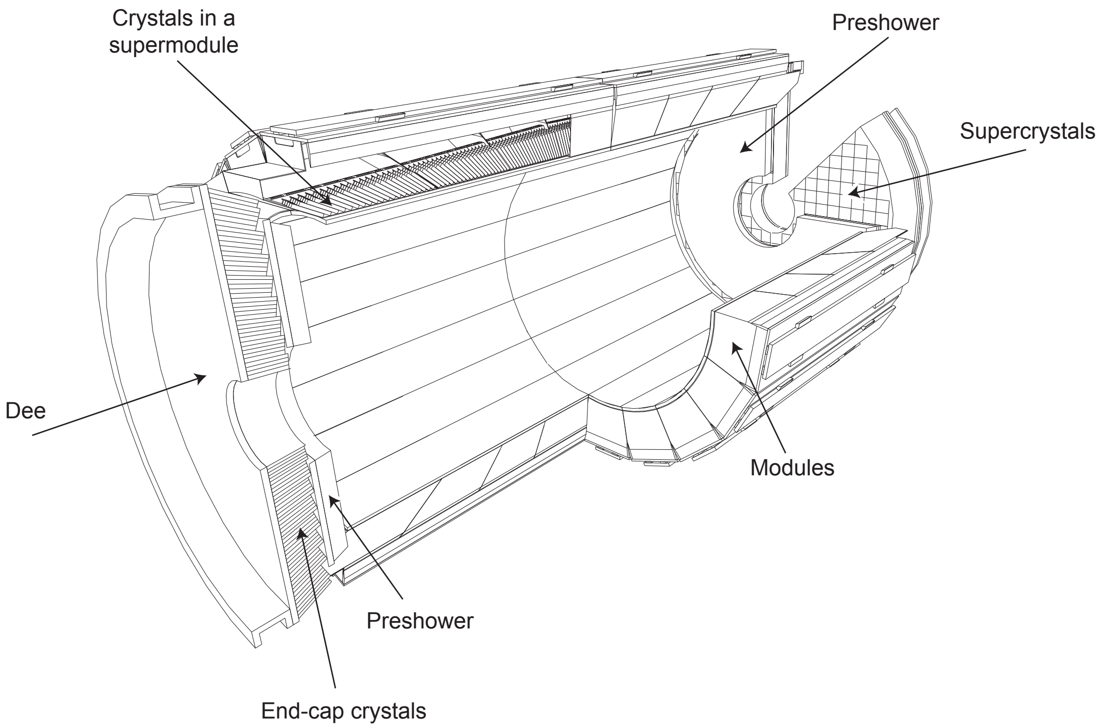


Figure 3.6: Layout of the CMS electromagnetic calorimeter [58]

Table 3.2: ECAL crystal characteristics [60]

Parameter	Barrel	Endcaps
Number of crystals	61 200	14 648
Crystal cross section in (η, ϕ)	0.0174×0.0174	varies
Crystal cross section at the front	22×22 mm^2	28.62×28.62 mm^2
Crystal cross section at the rear	26×26 mm^2	30×30 mm^2
Crystal length	230 mm ($25.8X_0$)	220 mm ($24.7X_0$)

ity. Finally, the crystals emit 80 % of their scintillation light in just 25 ns, however the light yield is relatively low. At 18 °C, about 4.5 photoelectrons per MeV are collected. The dependence of the light yield on temperature requires a cooling system capable of keeping the crystal temperature stable within ±0.05 °C to preserve energy resolution [60].

The energy-dependent resolution of the calorimeter can be parameterised as follows [58]:

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

where S is the stochastic term, N is the noise term, and C is the constant term. Figure 3.7 shows the energy resolution in bins of pseudorapidity, measured using Z $\rightarrow ee$ decays from MC simulation and 2012 data.

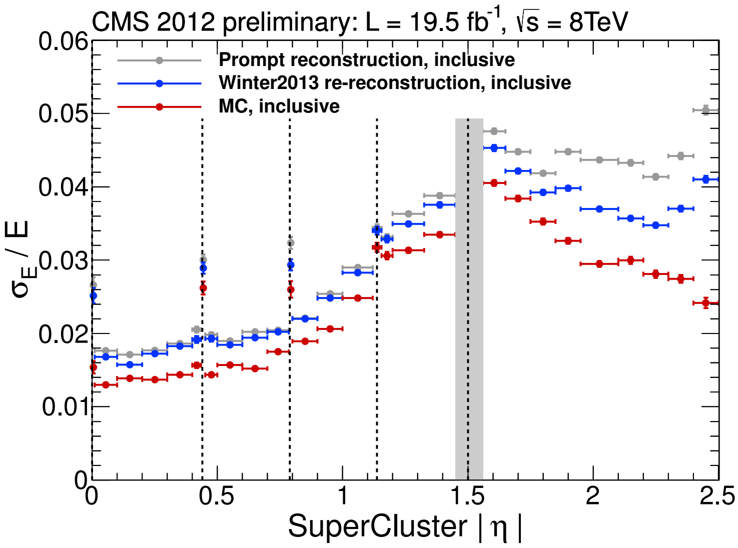


Figure 3.7: Relative electron (ECAL) energy resolution unfolded in bins of pseudorapidity η. Electrons from Z $\rightarrow ee$ decays are used. The resolution σ_{E}/E is plotted separately for data and simulated events. Clearly, it is affected by the amount of material in front of the ECAL, and is degraded in the vicinity of gaps in η between ECAL modules (shown by the vertical lines in the plot) [61].

3.2.3 Hadron Calorimeter

The hadron calorimeter (HCAL) is the next sub-detector located mostly inside the solenoid and completing the CMS calorimetry system. It is essential for the measurement of hadron jets and missing transverse energy.

As shown in Figure 3.8, HCAL consists of four subsystems: the hadron barrel calorimeter (HB), the hadron endcap calorimeter (HE), the hadron outer calorimeter (HO) and the hadron forward calorimeter (HF). The barrel and endcap parts (HB,

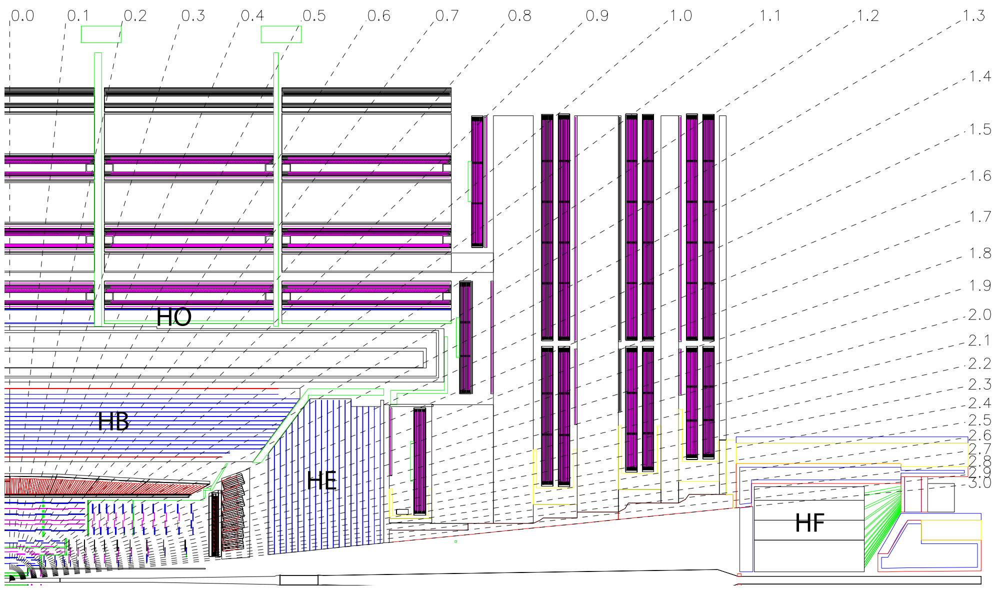


Figure 3.8: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [58].

HE) cover the pseudorapidity range up to $|\eta|< 3.0$, and the forward part (HF) extends it to a total coverage of $|\eta|< 5.0$. HCAL surrounds ECAL from its outer limit of 1.77 m from the beamline, to the inner limit of the magnet coil at 2.95 m from the beamline. However, due to space limitations the barrel calorimeters do not contain complete hadronic showers, therefore an outer calorimeter (HO) was designed to measure the energy leakage. It is placed in the muon system just outside of the solenoid in the barrel region.

HCAL is a sampling calorimeter consisting of alternating layers of brass absorbers and plastic scintillators, which form the active elements. The choice of the absorber material was caused by its short hadronic interaction length and its property of being non-magnetic, which is crucial in the strong magnetic field of the CMS magnet. The scintillation light is guided by embedded wavelength-shifting (WLS) fibres. The light from the WLS is then transmitted via a network of clear fibres, arranged in read-out towers, to hybrid photodiodes (HPDs) [58].

Both HB and HE scintillators have a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta|< 1.6$, and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta|\geq 1.6$. The tower segmentation of the forward calorimeter (HF) varies from $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ at $|\eta|= 3.0$ to $\Delta\eta \times \Delta\phi = 0.3 \times 0.35$ at $|\eta|= 5.0$. The HF is placed at about 11 m from the interaction point, and is essential to reconstruct very forward hadron jets. Together with the HO, it provides the hermeticity of the calorimetry system, making it possible to measure the transverse missing energy to a reasonable precision.

3.2.4 Superconducting Magnet

The superconducting solenoid is a central feature of the CMS apparatus, essentially giving it its name. The magnet has a length of 12.5 m, diameter of 6.3 m and mass of 220 t. Although it was initially designed to sustain a uniform magnetic field of 4 T within the 5.9 m diameter free bore, operation at 3.8 T was chosen in order to increase the lifetime. The magnetic field is returned by a massive iron yoke. The main parameters of the CMS magnet are shown in Table 3.3.

Table 3.3: Parameters of the CMS superconducting solenoid [60, 62]

Parameter	Value
Field	3.8 T
Inner bore	5.9 m
Length	12.5 m
Number of turns	2168
Current	18 160 kA
Stored energy	2.3 GJ

The large bending power of the solenoid is required to bend the tracks of high energy charged particles to an extent where good momentum resolution is achieved. The design requirement for the strength of the magnetic field was the ability to unambiguously determine the sign of the electric charge for muons with a momentum of ≈ 1 TeV [60].

The solenoid coil is constructed from four layers of superconducting high-purity niobium-titanium cable co-extruded with pure aluminium, which acts as a thermal stabiliser. The cold mass is cooled down to 4.5 K by liquid helium. If a fast discharge happens (e.g. caused by a magnet quench), about 3 days are necessary to re-cool the coil.

3.2.5 Muon System

The last sub-detector placed on the outermost part of CMS is the muon system. Since the muons are the most penetrating particles detectable by CMS, they have the cleanest signature and play an important role in many physics analyses. Due to their ability to travel through the many layers of the calorimeters, muons are relatively easy to identify and separate from the background.

The layout of the CMS muon system is shown in Figure 3.9. It consists of the drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The entire system surrounds the solenoid and covers the pseudorapidity region of

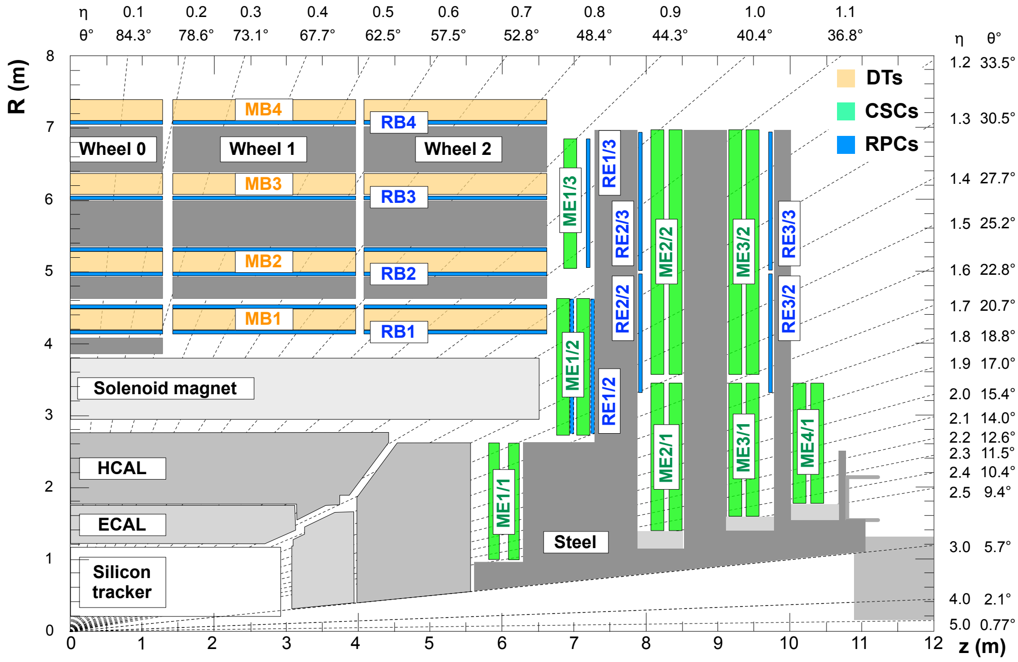


Figure 3.9: Layout of one quarter of the CMS muon system. Four drift tube (DT, in light orange) stations are labeled MB ("muon barrel") and the cathode strip chambers (CSC, in green) are labeled ME ("muon endcap"). Resistive plate chambers (RPC, in blue) are in both the barrel and the endcaps of CMS, where they are labeled RB and RE, respectively.

$|\eta|< 2.4$. Three different technologies are used in order to maximise the muon triggering, identification and reconstruction efficiency in both barrel and endcap regions of the detector.

The drift tubes are located in the barrel region ($|\eta|< 1.2$). Consisting of four stations, they form concentric cylinders around the beam line; there are 250 drift chambers with about 172 000 sensitive wires in total. When a muon passes through the volume, it knocks electrons off the atoms of the gas, which then follow the electric field and reach the positively-charged wires, providing information on the muon's position. The chambers are filled with the gas mixture of 85 % Ar and 15 % CO_2, where the muon drift time does not exceed 380 ns. Although this value is bigger than the typical bunch crossing time (25 or 50 ns), it is sufficient because of the small muon rate in this region.

In the endcaps, the cathode strip chambers cover the pseudorapidity region of $0.9 <|\eta|< 2.4$. Each of 468 CSCs is a trapezoidal multi-wire proportional chamber consisting of 6 gas gaps with a plane of radial cathode strips and a plane of anode wires which are roughly perpendicular. A charged muon traversing each plane of a chamber causes gas ionisation and a subsequent electron avalanche which produces a charge on the anode wire and an image charge on the cathode strips. The gas used in CSCs is a mixture of Ar, CO_2 and CF_4.

The resistive plate chambers system is complementary to both DT and CSC systems, and is located in both barrel and endcap regions ($|\eta| < 2.1$). RPCs also operate in an avalanche mode with a gas mixture of $C_2H_2F_4$, C_4H_{10} and SF_6, and due to an excellent time resolution of about 1 ns they provide fast information for triggering. The spacial resolution is, however, quite limited ($\approx$ 1 cm, compared to $\approx$ 100 µm for DTs and CSCs).

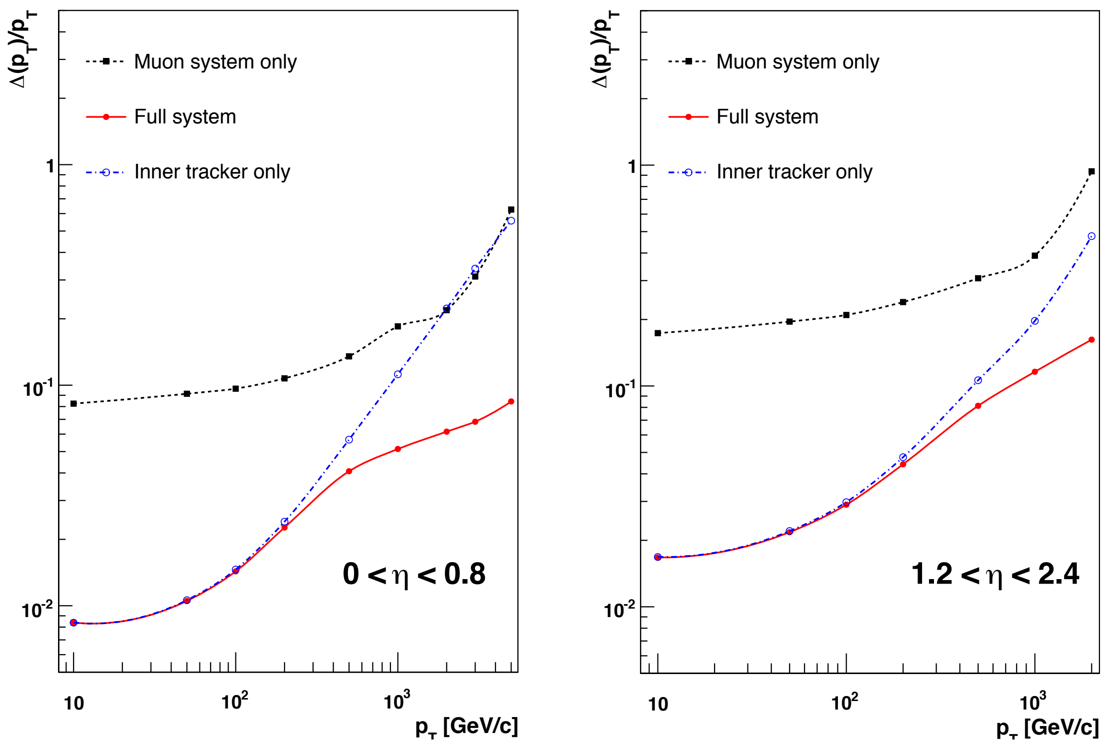


Figure 3.10: Muon transverse momentum resolution as a function of the transverse momentum (p_T) using the muon system only (black), the inner tracking only (blue), and both (red), in regions of $|\eta| < 0.8$ (left) and $1.2 < |\eta| < 2.4$ (right) [58].

The muon momentum is measured in both the tracker and the muon system. As it can be seen on Figure 3.10, both subsystems contribute to the momentum resolution at different p_T values. This happens due to the difference in the magnetic field and detector technology. For low-p_T muons, the best momentum resolution is obtained in the tracker, whereas in the high-p_T region the muon system provides a significant improvement. Therefore, by using information from both the silicon tracker and the muon chambers (i.e. reconstructing the "global muon"), the momentum resolution is improved in the whole p_T region up to a $\approx$ 1 TeV level.

3.2.6 Trigger and Data Acquisition

At design LHC luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}, approximately 25 collisions are expected to occur at each crossing of the proton bunches. The bunch spacing of 25 ns corresponds to a crossing rate of 40 MHz. Since every event produces ~ 1 MB of raw data, it corresponds to a total data production of 40 TB s^{-1}. Attempting to store all of these data is clearly beyond the available technology. Moreover, only a fraction of events contain hard scattering processes that are of interest, therefore an effective trigger system had to be implemented.

The CMS trigger is a two-level system, consisting of two independent parts: the Level-1 (L1) trigger and the High-Level Trigger (HLT). The L1 trigger is a hardware system implemented in programmable electronics residing partly on the detector, and partly in the underground control room located at approximately 90 m from the experimental cavern. The maximum latency between the collision and the L1 accept decision received by front-end electronics is 3.2 μs. During this amount of time, the complete event information is buffered in pipelined memories on the detector. The only information used for the L1 trigger decision is that from the muon system and the calorimetry. Since the reconstruction of tracks exceeds the time scale required for the L1 decision, the tracker information can't be used. The L1 trigger reduces the event rate from ~ 40 MHz to ~ 100 kHz, corresponding to a data flow of about 100 GB s^{-1}. These events are fed into the HLT system.

The High-Level Trigger is a software system implemented in a single CPU farm, sometimes referred to as the "Event Filter Farm". Having access to the full event information, customised algorithms of increasing complexity are used which results in a highly flexible trigger system. The event rate is reduced down to ~ 300 Hz, with the final data rate of approximately 300 MB s^{-1} being stored on a large disk cache at the experimental site (the Storage Manager) and later on transferred to CERN Tier 0 for further processing (see Section 3.3).

Since the start of the LHC running, the operating conditions have been changing drastically. During the start-up year of 2010, the instantaneous luminosity went up from about 10^{27} cm^{-2}s^{-1} to approximately 0.2×10^{32} cm^{-2}s^{-1}. In 2011 the luminosity ramped up to a factor of 20 above that of 2010, reaching approximately 4×10^{33} cm^{-2}s^{-1}. This required a lot of continuous effort to control the trigger rates at a reasonable level, whilst also keeping its efficiency acceptable. In 2012 the luminosity was more stable, peaking at $\approx 7.6 \times 10^{33}$ cm^{-2}s^{-1} which is just a factor of 2 above the 2011 values. However, it still came as a challenge because of the impact of pile-up. At a bunch spacing of 50 ns and increased centre of mass energy of 8 TeV, the average number of pile-up vertices nearly doubled comparing to that in 2011, which required a major CPU extension and implementation of sophisticated PU mitigation techniques at the HLT level. The author's contribution to the HLT development of the trigger paths important for top physics is described in Chapter 4.

3.3 Computing

The vast amounts of data delivered by the CMS detector impose high requirements on the offline computing system. During 2010–2012 operation, CMS collected ~ 10 PB of raw data per year. Including Monte Carlo simulation, reconstructed data and analysis skims, the total annual amount of data essentially doubles. To handle the distributed storage and processing of these data, not just for CMS but for the entire high energy physics community using the LHC, a worldwide LHC computing grid (WLCG) has been put in place.

WLCG is a global collaboration of more than 150 computing centres in about 40 countries. The grid has a tiered architecture, comprising 4 tiers with different resources and services. The first one, Tier 0, is based at CERN and is responsible for data-taking. It accepts raw data from the data acquisition system and repacks it into primary datasets according to the trigger information. The raw data are archived to tape, and is also prompt-reconstructed (within 48 hours) before being distributed to the Tier 1 (T1) centres around the world. There are 8 T1 sites based at large national laboratories in collaborating countries (e.g. Rutherford Appleton Laboratory in the UK and Fermi National Accelerator Laboratory in the US). Each of the T1 centres is used for large-scale centrally organised data-processing activities. The data are then distributed in a reduced format (see Section 3.3.1) to a more numerous set of Tier 2 centres, typically located at collaborating universities. Each of these centres is used for the grid-based analysis and Monte Carlo simulation for the whole experiment, as well as local services for groups maintaining them. The last stage of computing system, Tier 3, is meant solely for a local institution's user analysis.

3.3.1 Event Data Model

At the heart of the CMS Event Data Model lies the concept of an event, which is physically a result of a single collision in the LHC. From a software point of view, the event is a C++ object container storing raw data from a single readout of detector electronics (e.g. hits in various sub-detectors), as well as reconstructed data which is based on this information, such as tracks, clusters and physics objects. All these C++ objects are stored in ROOT format [63].

The EDM makes use of three main data formats, based on different levels of detail and precision:

- RAW format, containing full information from the detector as well as L1 and HLT trigger decisions, with an event size of ~ 1.5 MB.

- RECO (reconstructed data) format, which is obtained from raw data by application of pattern recognition and compression algorithms. This data includes reconstructed detector hits, clusters and physics objects (electrons, muons, etc.). The typical event size is ~ 250 kB.

- AOD (Analysis Oriented Data) format, produced by filtering the RECO data from the reconstructed detector objects, leaving just the high-level physics objects required for analysis. The event size is reduced down to ~ 50 kB.

The RECO and AOD data are analysis-ready data formats, produced centrally and used by many physics analysis groups. However, further simplification of the data is also a common practice. By transforming the C++ objects produced by CMS software into plain basic types or vectors of them, only including the analysis-specific content, the event size can be reduced down to ~ 3 kB level depending on the needs of a particular analysis. This data format is often referred to as private "ntuples", and it requires specific analysis software capable of restructuring the data into user-defined classes. By following this approach, the analysis can be run locally and generally much faster than processing the RECO or AOD data. However, it requires "ntuplising" these data every time when new centrally-recommended physics objects or corrections are produced.

3.3.2 Analysis Software

Both of the analyses described in this thesis use the CMS software framework (CMSSW [64]), as well as Bristol Analysis Tools (BAT [65]). The differential cross section analysis also uses an additional level of python scripts for post-processing [66].

CMSSW is the key CMS software framework built around the Event Data Model (see Section 3.3.1). The framework is essential for purposes of Monte Carlo simulation, detector calibration and alignment, as well as data reconstruction and analysis. CMSSW has a modular architecture, consisting of one configurable executable (cmsRun) and a large set of plug-in modules that contain all the code needed for event processing (reconstruction algorithms, calibration, etc.). Different versions of CMSSW were used for different analyses:

- CMSSW_4_2_8 for the top mass analysis on 2011 data;

- CMSSW_4_4_4 for the missing transverse energy analysis on 2011 data;

- CMSSW_5_3_9 for the top pair cross section analysis on 2012 data.

Corresponding versions were used to produce ntuples for processing by BAT, which was used to read the data, apply selections, calculate high-level variables and to create various histograms of distributions. BAT was originally started in 2010 by Łukasz Kreczko for the needs of the Bristol top group, later on also developed by the author and other researchers from Bristol and affiliated top groups. Like CMSSW, this framework has a modular structure, with its classes falling in four main categories:

- readers, for translating plain data types from ROOT files into C++ objects;
- RECO objects, i.e. output of the readers (physical objects like leptons, jets and its collections);
- selections, for application of event selections;
- analysers for creating histograms, applying selections, algorithms, and filling histograms.

All analysers are independent from each other, making the analysis chain stable and reliable. The final set of python scripts is used to prepare the histograms, perform fitting and unfolding procedures (in case of cross section analysis), and producing final tables and plots. Rootpy package [67] was used to access ROOT libraries in python interface, and matplotlib [68] was used to create plots.

3.4 Monte Carlo Simulation

To develop and test any analysis technique in modern particle physics, simulated events from Monte Carlo (MC) generators as well as a working simulation model of the detector are inevitably needed. All current theoretical knowledge of the Standard Model processes is incorporated in Monte Carlo simulation, and can be validated by direct comparison of the simulated events with real data. Any deviations may indicate the signs of new physics processes, which are also simulated according to various BSM physics models. On the other hand, these deviations can also be caused by deficiencies in the simulation process. Therefore, adequate and robust MC simulation is crucial for both precision measurements and searches for new physics.

Different MC simulation techniques exist and can be used in various MC generators. In this section, the basic concepts of MC simulation are introduced. All the event generators exploited in this thesis as well as different MC tunes used for evaluation of theoretical uncertainties are briefly described. Additionally, the Geant4-based [69] simulation of the CMS detector is mentioned.

3.4.1 Generation of hadron collision events

Simulation of hadron collision events is a rather complicated process. It is usually factorised into the following steps, schematically illustrated in Figure 3.11:

- incoming hadrons (protons);
- hard scattering of partons;
- parton showers;
- hadronisation of partons and hadron decays;
- underlying event.

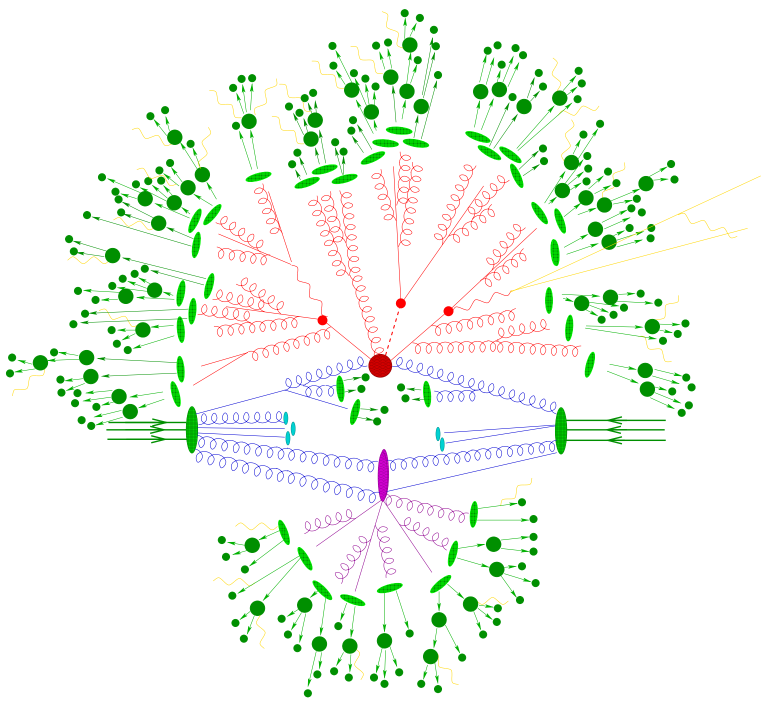


Figure 3.11: Schematic overview of a hadron collision process [70]

The simulation of hadron collisions makes use of the factorisation between the short-distance hard scattering process and long-distance soft interactions described by parton distribution functions, as shown in Section 2.2.5. In matrix-element MC generators such as MadGraph, the hard scattering process is described by calculating the tree-level matrix elements and including higher-order corrections, i.e. additional partons in the final state. The generator produces all possible Feynman diagrams for a renormalisable Langrangian-based model, and generates its matrix elements at the tree level, performing the integration over all phase-space. This calculation is then used to produce the cross section of various processes and sub-processes, as well as partons and parton decays.

Partons from matrix element calculations are then matched to parton showers. Accelerated coloured partons emit QCD radiation in the form of gluons, similarly to QED radiation emitted by accelerated electric charges. But since gluons also carry a colour charge, they emit further radiation, which results in parton showering. It is convenient to describe it in terms of the initial state radiation (ISR), which refers to the radiation of incoming partons, and the final state raditation (FSR), i.e. radiation of outgoing partons. Theoretically, parton showers can be regarded as higher-order corrections to the hard scattering process. However, precise calculation of these corrections is not feasible, therefore an approximation scheme is used, based on QCD evolution equations first described by Dokshitzer, Gribov, Lipatov, Altarelli and Parisi (DGLAP) [71, 72, 73].

After the parton showering is modelled, colour-neutral hadrons have to be formed in the hadronisation process according to the colour confinement, as described in Section 2.1.3. The most widely used hadronisation model (implemented in PYTHIA generator) is the Lund string model, where gluons are split into $q\overline{q}$ pairs and then turned into hadrons via the string fragmentation model [74]. Another hadronisation model (HERWIG) is based on cluster fragmentation, where colourless clusters are formed from quarks and gluons with low invariant masses, which then are turned into clusters [75].

Finally, apart from the hard scattering process, it is also important to model all the additional activity in proton-proton collisions, which is referred to as underlying event activity [76]. This activity includes multiple interactions of the remaining partons in hadron collisions, and beam remnants which do not participate in the hard scattering. Both the multiple parton interactions and the beam remnants carry a colour charge and produce additional soft hadrons. The colour charges from all the activity in the event, including the hard scattering process, ISR and FSR, beam remnants and multiple parton interactions can interact in a very non-trivial manner. This leads to additional uncertainty called colour reconnection [77], which also affects the whole hadronisation process and may result in biased precision measurements. To account for all known hadronisation and underlying event effects, phenomenological models with free parameters need to be tuned to data. Different tunes can then be used to estimate the systematic uncertainties arising from imperfect understanding of these effects.

One of the most significant systematic uncertainties can arise from the procedure of matching the matrix elements to parton showers. There is a certain threshold for partons transverse momenta from matrix element calculations to be matched with parton showers, often referred to as ME-PS threshold. To estimate the systematic uncertainty arising from it, additional MC samples are generated with the varied value of the threshold. In this thesis, this procedure was performed for signal $\mathrm{t\bar{t}}$+jets and background W + jets and Z + jets samples. Similarly, the systematic effect of the choice of factorisation scale Q^2 (described in Section 2.2.1) is also determined.

3.4.2 Monte Carlo event generators

In the analyses presented in this thesis, all physics processes were generated with a set of different MC generators, including MadGraph [78], PYTHIA [79, 80], MC@NLO [81], and POWHEG [82]. The performance of MC generators is usually optimised for one or more of the simulation steps described above, therefore two or more generators are often used in combination. For example, the nominal $\mathrm{t\bar{t}}$ signal samples used in this thesis are modelled with MadGraph and PYTHIA. Standalone generators can also be used for modelling theory systematic effects, which is also exploited in this thesis.

MadGraph [78] is a multi-purpose matrix-element Monte Carlo event generator. It automatically generates matrix elements for decays and $2 \to n$ scatterings. The matching of matrix elements to parton showers is performed according to the so-called MLM prescription [83] if a parton-jet pair satisfies a certain ΔR separation criterion. If none or more than one matched jets are found, events are rejected. As mentioned above, there is also a certain transverse momentum threshold requirement for the partons to be considered in the matching. Default values as well as variations used to estimate the systematic effect of the matching threshold in this work can be found in Table 5.4.

PYTHIA [79, 80] is one of the most widely used MC generators in high energy physics. It is a standard tool used for simulation of initial protons fragmentation, multi-particle production, quark and gluon hadronisation, beam remnants and the underlying event. In this work, PYTHIA is used to generate QCD multi-jet production, and to simulate the underlying event on top of other generators providing partons from hard processes.

MC@NLO [81] is a parton shower MC generator providing next-to-leading order (NLO) corrections for a simulated process. It represents a major improvement in precision and accuracy of physics simulations compared to the leading order (LO) generators like PYTHIA. An NLO implementation can handle additional partons in the final state coming from the hard scattering, which is not possible with LO calculations. Therefore it includes additional dynamic features, giving a major advantage for heavy flavour physics.

POWHEG (Positive Weight Hardest Emission Generator) [82] also works at NLO precision. Its main difference to the MC@NLO generator lies in the order of process generation: POWHEG initially generates the hardest radiation. This technique allows to avoid double-counting coming from subsequent radiation which may happen with the MC@NLO method when matching QCD NLO calculations with parton showers.

3.4.3 CMS detector simulation

Detector simulation is an essential final step on top of the full physics simulation of proton-proton collisions. All generated particles resulting from these collisions are propagated to the simulated layers of the detector so that their interaction with the detector material and detector response can also be modelled. The CMS simulation is based on the Geant4 (GEometry ANd Tracking [69]) package, which is implemented in the CMS software CMSSW [64]. This package fully recreates the geometry of the detector, including all its subsystems as well as a detailed map of the magnetic field.

In general, the simulation procedure includes modelling of the following steps:

- interaction region;
- particles traversing all the layers of the detector and the interaction processes;
- multiple interactions per beam crossing (pile-up) and event overlay effects;
- digital readout of the detector (digitisation).

The data stream generated with this process is very similar to the output of the real detector, although naturally in addition to reconstructed objects, generated events also include collections of generated objects. This allows users to access the Monte Carlo truth information, which is essential for any particle physics analysis.

3.5 Object Reconstruction

Most CMS analyses, including the ones described in this thesis, adopt a reconstruction technique called Particle Flow (PF) [84]. This algorithm is used to obtain a global event description at the level of individually reconstructed particles by means of combining information coming from all sub-detector systems. The ultimate goal is to determine type, energy and momentum of all the particles in the event with the highest possible precision and in the most optimal way. The types of these particles include electrons, muons, charged hadrons, neutral hadrons and photons. All these particles are then used to reconstruct jets (Section 3.5.3), missing transverse energy (Section 3.5.1) and tau leptons from their decay products.

3.5.1 Electron reconstruction

The reconstruction of the $\mathrm{t\bar{t}}$ pair with an electron in the final state (see Section 2.5) imposes strict requirements on the electron identification and its energy-momentum measurement, precision of which is of major importance for both top mass and $\mathrm{t\bar{t}}$ cross section measurements.

Although the CMS detector is equipped with highly accurate ECAL and tracker systems, electron identification and reconstruction is still a challenging task due to the large amount of tracker material (see Section 3.2.1). This results in significant Bremsstrahlung photon emission, which often causes an ECAL energy deposit to be widely spread in the azimuthal direction because of the high magnetic field. Therefore, dedicated algorithms were developed in order to collect all Bremsstrahlung energy deposits in the calorimeter (Bremsstrahlung recovery), and also to take into account the kinks in the electron trajectory caused by photon emission.

Electron reconstruction in CMS has the following distinct stages: seeding; track finding; pre-identification; Bremsstrahlung recovery; track-cluster linking; and final identification. Historically, the original seeding algorithm was designed and optimised for isolated high-p_{T} electrons. This approach starts from ECAL clusters, and therefore is called the "ECAL-driven" seeding. It is based on the property of ECAL energy deposits to have a narrow width in the η coordinate, and to be widely spread in ϕ (azimuthal direction) as mentioned above. The electron and all the associated Bremsstrahlung energy deposits form a single "super-cluster", and the ability to correctly identify it affects the overall performance of this method. Super-clusters are then matched to pairs or triplets of hits in the inner tracker layers, forming the track seeds upon which the electron tracks are built.

The performance of the ECAL-driven method is not very well suited for non-isolated and low-p_{T} electrons. This occurs mainly due to the fact that the super-cluster position and energy can be highly biased by the impact of overlapping particles, especially if the electron happens to be within a jet and therefore non-isolated. Also, high track multiplicity complicates the backward propagation from a super-cluster, because it can be consistent with a number of track seeds corresponding to other particles.

Within the particle flow method, efficient reconstruction of non-isolated and low-p_{T} electrons is particularly important since it affects the reconstruction of jets and missing transverse energy. Therefore, a different ("tracker-driven") seeding algorithm is used, which starts from reconstruction of tracks. The baseline of the CMS track reconstruction is the Kalman filter (KF) [85], which is a linear least-squares estimator based solely on Gaussian probability density functions. It is particularly suitable for muon reconstruction, since the interaction of muons with matter is dominated by multiple Coulomb scattering, which is well modelled by Gaussian fluctuations. However, this approach usually fails for electrons because Bremsstrahlung photon emission is highly non-Gaussian. To accommodate the resulting kinks in the electron trajectory, the Gaussian-Sum Filter (GSF) [86] is used, which is essentially a non-linear generalisation of the Kalman Filter. In this method, Bremsstrahlung energy loss is modelled by a Gaussian mixture, therefore the GSF track fit provides a better estimate for the inner and outer track momentum compared to the KF algorithm. The downside of this approach is its high CPU usage, which means it can be run only on a limited number of seeds.

The GSF tracks are reconstructed upon all ECAL-driven seeds. In the case of "tracker-driven" seeds, a pre-identification based on high-purity KF tracks has been adopted. This procedure starts with the tracks reconstructed with very tight criteria, thus substantially decreasing the fake rate. Then an iterative-tracking strategy is carried out by means of removing hits unambiguously assigned to tracks from the previous iteration, and also progressively relaxing track seeding criteria. This approach leads to both high efficiency and low fake rate, which is crucial for low-p_{T} and non-isolated electrons.

In the next step, track-cluster matching is performed. In the case where an electron has negligible Bremsstrahlung emission, the track is well reconstructed with the KF algorithm all the way to the ECAL internal surface, where the closest cluster is matched to the track. The corresponding cluster energy is compared with the track momentum, and if the ratio (E/p) is close to unity, the track is selected. On the contrary, if the electron experiences a significant Bremsstrahlung emission, other track characteristics are used. In this case a selection based on the number of hits in the tracker and the χ^2_{KF} of the KF fit is applied before running a GSF refit. Finally, the number of hits, the GSF refit χ^2_{GSF}, $\chi^2_{\mathrm{KF}}/\chi^2_{\mathrm{GSF}}$ ratio, the energy loss measured by the track and the quality of the ECAL cluster-track matching are fed into a multivariate analysis using a Boosted Decision Tree (BDT) estimator.

Both tracker-driven and ECAL-driven seeds are used to obtain the GSF track collection of electron candidates. In the particle flow algorithm, it is necessary to link both electron and Bremsstrahlung energy deposits to the GSF track. A super-cluster is linked to a track if the extrapolated position from the outermost tracker measurement is within the boundaries of one of the ECAL cells at the expected depth of the electron shower maximum. The preshower-ECAL and ECAL-HCAL links are made in a similar way.

Another important particle flow procedure, also driven by GSF tracks, is Bremsstrahlung recovery. In order to reconstruct an electron with correctly assigned energy and momentum, it is crucial to identify all energy deposits from Bremsstrahlung photons, thus forming a super-cluster. This procedure is carried out for each tracker layer by computing a straight-line extrapolation tangent to the track, up to the calorimeter. To determine a Bremsstrahlung photon, track-cluster linking is performed as described above. To limit the charged hadron contamination, clusters already assigned to KF tracks are not included in the calculation.

3.5.1.1 Electron identification

Electron reconstruction in CMS is based on a characteristic signature that electrons leave in the tracker and calorimetry systems. However, other objects like charged hadrons, jets or photon conversions can produce very similar signatures and therefore may be reconstructed as electrons. Therefore, in order to distinguish these “fake” electrons from “real” ones, a further selection has to be applied. This procedure is referred to as electron identification (or electron ID).

Initially, the electron identification is performed at the final stage of the electron reconstruction process. The working points of the cuts applied are selected to be loose enough in order to satisfy the requirements of most CMS analyses. Afterwards, more specific (tighter) ID cuts are applied for each individual analysis, defining the working point in the trade-off between selection efficiency and contamination from fakes. This will be discussed separately in the selection description for each analysis in corresponding chapters.

Different algorithms are used for electron identification at different stages of physics analysis. These algorithms use information from both the tracker and calorimeter systems combined in different ways, including the track quality variables, ECAL shower shape variables, track and super-cluster matching variables and Bremsstrahlung variables. The full list of variables used by electron ID algorithms described below is given in Table 3.4.

Table 3.4: Variables used in electron identification algorithms

Variable	**Description**
	Track quality variables
p_{T}	Transverse momentum of the GSF track
η	Pseudorapidity of the GSF track
GSF $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}}$	Transverse momentum resolution of the GSF track
$\#\mathrm{hits}_{\mathrm{KF}}$	Number of reconstructed KF track hits
χ^2_{GSF} and χ^2_{KF}	GSF and KF goodness-of-fits
	ECAL shower variables
$\sigma^2_{i\eta,i\eta}$	Cluster shape variable that gives a measure of the width of the cluster in η, using the distribution of energy in a 5×5 block of crystals around the seed crystal (the one with the highest energy) [87]: $\sigma^2_{i\eta,i\eta} = \sum_{5\times5\ \mathrm{crystals}} \left(\eta_i - \eta_{\mathrm{seed\ cluster}}\right)^2 E_i/E_{\mathrm{seed\ cluster}}$
$\sigma^2_{i\phi,i\phi}$	Cluster shape in ϕ
η_{SC} (ϕ_{SC})	Width of the super-cluster in η (ϕ)
$1 - E_{1\times5}/E_{5\times5}$	$E_{1\times5}$ is the energy in the central 1×5 strip of the 5×5 electron cluster, and $E_{5\times5}$ is its total energy
$E_{3\times3}/E_{\mathrm{SC,\ raw}}$	Ratio of the energy of a cluster of 3×3 to the uncorrected (raw) energy of the super-cluster
$E_{\mathrm{PS}}/E_{\mathrm{SC,\ raw}}$	Ratio of the energy in the preshower detector to the raw super-cluster energy (only in the endcap region)
	Longitudinal shower shape variables
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
$H/(H + E_{\mathrm{e}})$	Hadron fraction of the shower, where H is the energy of the hadron cluster linked to the GSF track

Table 3.4: Variables used in electron identification algorithms (continued)

	Track/super-cluster matching variables
$\Delta\eta_{\mathrm{in}}$ ($\Delta\phi_{\mathrm{in}}$)	Distance in η (ϕ) between the super-cluster position and the extrapolated track position
$\Delta\eta_{\mathrm{vtx}}$ ($\Delta\phi_{\mathrm{vtx}}$)	Distance in η (ϕ) between the super-cluster position and the position of the GSF track at vertex
$(E_{\mathrm{e}} + \sum E_{\gamma})/p_{\mathrm{in}}$	Ratio of the super-cluster energy to the inner track momentum
$E_{\mathrm{e}}/p_{\mathrm{out}}$	Ratio of the electron cluster energy to the outer track momentum
$1/E_{\mathrm{SC}} - 1/p_{\mathrm{T}}$	Difference between inverse super-cluster energy and inverse track momentum
	Bremsstrahlung variables
f_{brem}	Measured bremsstrahlung fraction, defined as: $f_{\mathrm{brem}} = (p_{\mathrm{in}} - p_{\mathrm{out}})/p_{\mathrm{in}}$, where p_{in} is the initial track momentum at the vertex and p_{out} is the track momentum at the last hit
$\sum E_{\gamma}/(p_{\mathrm{in}} - p_{\mathrm{out}})$	Ratio between the Bremsstrahlung photon energy as measured by ECAL and by the tracker
EarlyBrem	Flag of $(E_{\mathrm{e}} + \sum E_{\gamma}) > p_{\mathrm{in}}$ inequality, corresponding to an electron emitting an “early” Bremsstrahlung photon, i.e. before it has crossed at least three tracker layers
LateBrem	Flag of $E_{\mathrm{e}} > p_{\mathrm{out}}$ 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

There are several electron identification algorithms used by various CMS analyses, and four of them are used in the analyses described in this thesis: simple cut-based (SCB ID), cuts in categories (CiC ID), particle flow (PF ID) and multivariate analysis (MVA ID).

Simple cut-based ID [88] is used in the high-level trigger, and therefore has to be as simple, fast and robust as possible. This algorithm only uses a limited number of electron ID variables shown in Table 3.4, including H/E ratio, $\Delta\eta_{\mathrm{in}}$ and $\Delta\phi_{\mathrm{in}}$ matching variables, cluster shape variables and isolation variables (to be described in the next section). Although this identification method has the advantage of simplicity, it does not show the best signal efficiency and background rejection.

Cuts in categories ID [89] is one of the more complex methods, exploiting the categorisation of electrons. It is optimised to select electrons from different sources (W, Z and J/ψ decays) and reject fakes from jets or conversions. In order to achieve higher efficiency, electron candidates are split into categories with different signal to background ratio, allowing for better tuning of the working points of the cuts. The categories are formed by selection based on the mentioned ID variables, as well as isolation and conversion rejection variables; these are discussed in Section 3.5.1.2 and Section 3.5.1.3, respectively.

The cuts within CiC ID are applied in order to maximise the signal to background ratio. Depending on the needs of different analyses, nine levels of cut severity are implemented: *VeryLoose, Loose, Medium, Tight, SuperTight, HyperTight(1-4)*. Each step decreases the fake rate by about a factor of two for electrons with $E_{\mathrm{T}} >$ 20 GeV [89]. The CiC ID is used as a primary electron identification method in the top quark mass analysis (Chapter 5).

Particle flow ID [84] is the final step of the particle flow electron reconstruction. Following the particle flow concept, it uses information from all the CMS sub-detectors obtained in previous reconstruction steps to build new observables for electron identification. The variables shown in Table 3.4 are combined into a single discriminator by a multivariate analysis technique (BDT method), which has been trained on signal and background Monte Carlo samples. PF ID is a relatively loose identification method, since it has to satisfy the needs of all analyses using particle flow collections.

MVA ID is used in top cross sections analysis on 2012 data. It is another multivariate analysis technique, optimised to select isolated electrons from W and Z decays. The variables shown in Table 3.4 are also combined into a single discriminator, which is optimised by "training" in different selection categories.

3.5.1.2 Electron isolation

Isolation is an observable that allows prompt electrons (i.e. the ones from W and Z decays) to be distinguished from jets faking electrons and electrons within jets. It is essentially a measure of activity around the particle. In CMS, there are two different ways of quantifying isolation: detector-based and particle-based. The detector-based isolation is defined separately for each detector subsystem (tracker, ECAL and HCAL) [90]:

- Tracker isolation, calculated as the sum of the transverse momenta of all tracks within a cone of $\Delta R = 0.3$ around the electron, excluding the electron momentum itself;

- ECAL isolation, i.e. the sum of the transverse energy of all ECAL clusters within a cone of $\Delta R = 0.3$ around the super-cluster position. The footprint of the original electron is also removed;

- HCAL isolation, defined as the sum of the transverse energy of all HCAL towers within a cone of $\Delta R = 0.3$ centred at the super-cluster position.

Particle-based isolation exploits the particle flow information: it is calculated as the sum of the transverse energy of all PF particles in a cone of $\Delta R = 0.3$ around the electron. Both detector-based and particle-based isolation definitions are often normalised to the electron transverse momentum (or energy in the case of calorimeter isolation) in order to improve signal efficiency. The normalised sum of the tracker, ECAL and HCAL isolation variables is referred to as detector-based relative isolation (or reliso), similarly normalised particle-based isolation is called PF reliso. These quantities are crucial in various methods to estimate the QCD background contribution for many analyses, and will be used for both electrons and muons throughout this thesis.

3.5.1.3 Identification of photon conversions

Interacting with detector material, photons can convert into electron-positron pairs. Due to the large material budget in the tracker (Figure 3.5), especially in the end-cap region, there is a high chance of conversions happening. The resulting electrons can successfully fake prompt signal electrons, passing all the identification criteria and appearing isolated if the initial photon was isolated, too. Therefore conversions constitute a large proportion of the QCD background to top signal. The electron-positron pair may not be symmetrical in transverse momenta, therefore a veto on a second electron is not sufficient to reject such electrons and other conversion identification criteria are necessary.

There are three main methods of conversion rejection adopted by CMS [90]. The simplest method is based on counting the number of missing hits in the tracker. Conversions are most likely to happen at some distance from the interaction point, essentially anywhere between the point of photon production and the end of the tracker. Therefore electrons produced in such conversions are likely not to traverse through all the pixel layers. To separate these electrons from the prompt ones, a cut on the number of missing layers can be used. However, this approach fails if a conversion occurs in the beam pipe, or if the reconstructed electron track is paired with unrelated hits in the tracker, making it look like a prompt electron.

A slightly more sophisticated approach is the partner track method, based on a search for a second track close to the electron, but with opposite curvature (hence opposite charge). After optimised geometrical cuts, a pair of tracks is interpreted as an electron/positron pair from a photon conversion.

These two methods for conversion identification were used in the top mass analysis on 2011 data. For the top cross sections analysis on 2012 data, along with the number of hits method, a more advanced technique called vertex fit was used. It is essentially a full vertex fit of all pairs of tracks, with the selection being made on the fit probability. The method benefits from the full use of track uncertainties and covariances, combining all information into a single discriminator. However, increased complexity of this technique makes it more CPU-intensive than the partner track method.

3.5.2 Muon reconstruction

A good muon reconstruction and identification was one of the main CMS design requirements. Initially, muons are reconstructed independently in the tracker ("tracker tracks") and in the muon system ("standalone muon tracks") using the Kalman Filter technique [85]. Based on these objects, two different reconstruction methods are used [91]:

- Global muon reconstruction. Each standalone muon track reconstructed in the muon system is matched with the tracker track by propagating onto a common surface. A global fit is performed on the combined collection of hits from both tracks, and the resulting muon is referred to as a *global muon*.

- Tracker muon reconstruction. All tracker tracks above a certain threshold (p_{T} > 0.5 GeV, $p > 0.5$ GeV) are extrapolated to the muon system, taking into account possible energy losses, magnetic field and multiple Coulomb scattering.

If the track matches at least one muon segment in the muon system, i.e. a short track stub made of DT or CSC hits, it is considered a *tracker muon*.

Global muon reconstruction has a high efficiency for high-p_{T} muons, penetrating through more than one muon station. On the contrary, tracker muon reconstruction is more efficient for low-p_{T} muons ($p_{\mathrm{T}} < 5$ GeV), as it requires just a single muon segment. Since the $\mathrm{t\bar{t}}$ analyses described in this thesis have a semileptonic signature with exactly one energetic muon in the final state (in the case of the muon channel), only the global muon reconstruction method is used, as its momentum resolution at high transverse momenta benefits from both the tracker and the muon system (Figure 3.10).

Following the reconstruction, the quality of the muon objects is verified by applying identification criteria. This is done in order to suppress hadronic punch-through, muons from decays in flight and cosmic muons. The selection is applied on the following observables:

- normalised χ^2 (χ^2/number of degrees of freedom) of the global muon fit;
- number of muon chamber hits in the global muon fit;
- number of muon stations with muon segments;
- transverse impact parameter d_{xy} (closest approach of the track to the primary vertex);
- longitudinal distance d_z of the tracker track w.r.t. the primary vertex;
- number of hits in the pixel detector;
- number of hits in the tracker layers.

The muon identification also exploits the CMS particle flow event reconstruction, making use of information coming from all sub-detectors. Particle flow muons (PF muons) are identified by imposing selection on all muon candidates reconstructed with the global muon method. This selection was optimised for identification of muons in jets, minimising the fake rate from misidentified charged hadrons, which is crucial for correct reconstruction of jets and missing transverse energy (described in more detail in the following sections). Depending on the analysis specifics, isolation requirements can also be applied; isolation definitions closely follow the ones for electrons (Section 3.5.1.2).

3.5.3 Jet reconstruction

Hadronisation of quarks and gluons leads to the production of narrow cones of particles moving in approximately one direction, called jets. This happens due to colour confinement, as particles carrying a colour charge cannot exist in free form, they have to fragment into hadrons before they can be detected directly. Therefore, to measure the initial parton's momentum and energy, all these particles must be combined into jets.

The CMS particle flow algorithm implies reconstruction of individual particles (charged and neutral hadrons, electrons, muons, photons) before combining (or clustering) them into jets. A few different jet clustering techniques exist, but the one used predominantly in CMS and exclusively in this work is the anti-k_t algorithm [92], which defines the distance between constituent particles as:

$$d_{ij} = \min\left(\frac{1}{k_{t,i}^2}, \frac{1}{k_{t,j}^2}\right)\frac{\Delta_{i,j}^2}{R^2} \tag{3.6}$$

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 respectively transverse momenta, rapidities and azimuth angles of particles i/j, and R is the radius parameter.

The clustering proceeds by identifying the smallest distances between particles and recombining them until all jets are formed and no particles are left. An event typically has a few well-separated hard (high-p_T) particles and a large amount of soft (low-p_T) particles. The distance d_{1i} between a hard particle 1 and a soft particle i will be fully determined by the transverse momentum of the hard particle and the Δ_{1i} separation. On the other hand, distance between soft particles with similar separation will be substantially larger. Therefore, soft particles tend to cluster around the hard ones before they cluster amongst themselves. On the output, the anti-k_t algorithm forms conical jets with boundaries resilient to soft radiation.

3.5.3.1 Jet energy corrections

Due to the non-linear and non-uniform response of the calorimetry systems, jets reconstructed using detector inputs typically have energies different to those of the corresponding Monte Carlo particle jets (or generator jets), reconstructed by clustering the four-momenta of all stable particles generated in Monte Carlo simulation. Therefore, some mapping procedure is necessary. Corrections applied to reconstructed jets in order to translate the measured energy to the true particle (or parton) energy are referred to as jet energy corrections, or JEC.

CMS has adopted a factorised approach of applying jet energy corrections [93], meaning that each correction level takes care of a different effect. The set of corrections is applied sequentially, i.e. the output of each step is the input to the next one. Essentially, each level of correction is a scaling of a jet four-momentum, with a scale factor depending on various parameters of the jet, typically pseudorapidity and transverse momentum. Currently, the correction levels go as follows:

- offset correction;
- relative (η) correction;
- absolute (p_{T}) correction.

The goal of the offset correction is the subtraction of pile-up and electronic noise contributions from the jet energy. The energy from pile-up vertices can be deposited in calorimeters since the additional proton-proton collisions occur close enough in time to the hard scattering process. Electronic noise in calorimeter readouts also creates additional energy offset which needs to be corrected for. The scale factors for offset correction are derived using data-driven methods (zero-bias collisions).

The relative η correction is designed to flatten the jet response in pseudorapidity. These corrections are extracted with respect to the barrel region in bins of p_{T}, and therefore are uncorrelated with the following absolute p_{T} corrections. Both Monte Carlo and data-driven (dijet balance) methods are used to derive the relative correction scale factors. Once a jet is corrected for η dependence, it is corrected back to the particle level by applying absolute p_{T} correction. The goal of this correction is to flatten the jet response in transverse momentum. Derivation of these scale factors is done by using Monte Carlo truth information, or data-driven Z/γ+jet balance techniques.

Due to the fact that CMS simulation is not perfectly tuned to the data yet, additional residual corrections are applied in order to achieve better agreement between data and simulation. It is essentially a small residual η- and p_{T}-dependent calibration applied exclusively to data, which will remain in place until CMS develops a perfectly tuned simulation reproducing the data features out of the box.

3.5.3.2 Particle Flow jet identification

To ensure the quality of the jets, final identification criteria are applied to all jet objects. Particle flow jet identification (or PF jet ID) is used to reduce the noise and rate of electrons reconstructed as jets. The cuts are applied on the following observables :

- number of constituent particles;

- neutral hadron energy fraction (NHF);
- charged hadron energy fraction (CHF);
- neutral electromagnetic energy fraction (NEF);
- charged electromagnetic energy fraction (CEF);
- number of charged particles (NCH).

A rather loose set of identification cuts (referred to as "loose PF jet ID") was used in this work: a requirement of more than one constituent particle in the jet, NHF< 0.99, NEF< 0.99. In addition, for a pseudorapidity region of $\eta < 2.4$ the following requirements are imposed: CEF< 0.99, CHF> 0 and NCH> 0.

3.5.3.3 b-tagging

The identification of jets originating from b-quarks, or b-tagging, is one of the most important tools in top quark physics, capable of significantly decreasing the background contamination of signal processes. A variety of b-tagging algorithms has been developed by CMS [94], and the one that was used in this work is called the combined secondary vertex (CSV) algorithm. It is based on the fact that b-hadrons have a significant lifetime ($\sim 10^{-12}$ s) and can travel a distance of a few centimetres before decaying. Therefore jets originating from b-quarks are likely to have a secondary vertex located at a considerable distance from the primary vertex, which can be used as an efficient discriminator between light jets and b-jets. In order to maximise its efficiency, apart from the secondary vertex information the CSV algorithm also exploits the track-based lifetime information. The following variables are used in the algorithm:

- number of tracks in the jet;
- number of tracks at the secondary vertex;
- secondary vertex category;
- secondary vertex invariant mass;
- ratio of the total track energy at secondary vertex with respect to all tracks in the jet;
- pseudorapidities of tracks at secondary vertex with respect to the jet axis;

- impact parameter significance (i.e. ratio of IP to its uncertainty) of the first track that increases the invariant mass above the charm threshold of 1.5 GeV (tracks are ordered by IP significance; the mass of the system is recalculated after adding each track);

- impact parameter significances of each track in the jet.

All these observables are combined into a single discriminator, the "medium" working point of which is used in this work. It provides ~ 70 % b-tagging efficiency for a mis-tag rate of approximately 1 % [94]. Figure 3.12 shows the efficiency of the CSV algorithm as a function of the discriminator threshold, as measured using b-jet enriched sample from 7 TeV data and MC simulation. The scale factors represent the difference in b-tagging efficiency between data and MC samples (see Section 6.1.4).

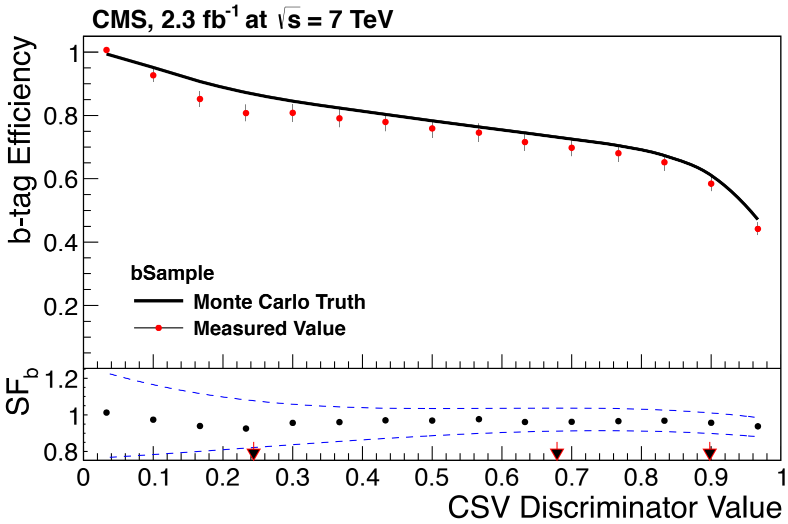


Figure 3.12: Measured b-tagging efficiency as a function of the discriminator threshold for the CSV algorithm. The absolute b-tagging efficiency measured from data and predicted from simulation is shown in the upper histogram. The scale factors SF_b are shown in the lower histogram, where the blue dashed lines represent the combined statistical and systematic uncertainty. The middle arrow indicates the "medium" working point used in the analyses [94].

3.5.4 Missing transverse energy

Due to momentum conservation, the sum of transverse momenta of all particles in the final state of proton-proton collisions is expected to be zero. However, some particles can escape the detector without being reconstructed, therefore creating the imbalance in transverse momentum which is referred to as missing transverse energy, defined as:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_i \vec{p^i}_{\mathrm{T}} \tag{3.7}$$

where $\vec{p^i}_{\mathrm{T}}$ are the transverse momentum vectors of all reconstructed particles. The modulus of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ vector is denoted by $E_{\mathrm{T}}^{\mathrm{miss}}$.

Accurate reconstruction of missing transverse energy is crucial for precise measurements of Standard Model processes with neutrinos in the final state. Top quark pair semileptonic decay is one of such processes, therefore analyses covered in this thesis implicitly (top mass) or explicitly ($\mathrm{t\bar{t}}$ cross section with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$-related variables) rely on efficient reconstruction of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. Misidentification and misreconstruction of any visible particles in the event contribute to $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ measurement, therefore it is a rather demanding task.

Just like in the case of leptons and jets, particle flow was used for $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction: in Equation 3.7 the transverse momentum vectors are of the particles reconstructed using the particle flow algorithm. This procedure gives so-called raw $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ on the output. The raw $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is systematically different from the true $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, which denotes the transverse momentum carried by invisible particles. This happens mainly due to the non-compensating nature of the calorimeters, effects of pile-up, noise, etc. Therefore, a set of corrections is applied:

- Type-0, which corrects $E_{\mathrm{T}}^{\mathrm{miss}}$ for pile-up;
- Type-I, a propagation of jet energy corrections (Section 3.5.3.1) to $E_{\mathrm{T}}^{\mathrm{miss}}$;
- xy-shift correction, reducing the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ ϕ modulation.

The causes of systematic $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ ϕ modulation include detector misalignment, beam spot displacement, inactive calorimeter cells and anisotropic detector response. The xy-shift correction mitigates these effects, making the measured $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution closer to the true $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution which is flat in ϕ because of the rotational symmetry of the collisions.

All these corrections were applied to missing transverse energy in the analyses described in this work. However, the xy-shift correction was not applied in the top mass analysis, since it was not available at the time. As this analysis is not particularly sensitive to $E_{\mathrm{T}}^{\mathrm{miss}}$, the ϕ modulation is not expected to affect the top mass measurement and its resolution.

3.6 Summary

In this chapter, the Large Hadron Collider (LHC) has been introduced to the reader. The CMS experiment including all its subsystems has been described in detail, the overview of the CMS computing model and analysis software have been shown. Reconstruction and identification methods of various analysis objects including particle flow algorithm have been discussed in detail.

4. High-Level Triggers for Top Quark Physics

The LHC is often referred to as a top quark factory, producing a $t\bar{t}$ pair nearly every second of its nominal operation. While the production rate of ≈ 1 Hz seems manageable in terms of recording the data, it is significantly complicated by background processes with similar signatures occurring at much higher rates.

The trigger is the starting point of any physics event selection process, and therefore is clearly important for any physics analysis. As mentioned in Section 3.2.6, the CMS L1 trigger rate is limited to ~ 100 kHz. In order to meet the data recording constraints of approximately 300 $\mathrm{MB\,s^{-1}}$, this rate is further reduced down to ~ 300 Hz, which is done by the HLT system. The total rate budget has to be allocated to maximise the acceptance across the full range of the CMS physics programme, which can be a matter of serious debate.

During the LHC operation in 2011 and 2012 under conditions of gradual increase of instantaneous luminosity and pile-up, but very limited rate budget, trigger developers constantly tackled the challenge of finding the best compromise between growing rates and maintaining reasonable signal acceptance. While the simplest approach is tightening the cuts on physical quantities like lepton or jet transverse momenta, it is not favourable since it lowers the number of stored signal events and decreases the phase space which is crucial for new physics searches as well as Standard Model precision measurements. Therefore, the development of more efficient algorithms is the most preferable solution, allowing a high level of acceptance to be kept for signal events, whilst effectively rejecting background events. This can often be achieved by reducing the level of approximation of the online (HLT) object reconstruction, making the algorithms closer to their sophisticated offline counterparts. However, it leads to a higher execution time, which is limited by computing resources available for HLT reconstruction. Hence, the CPU timing is another major constraint faced by the HLT developers.

This chapter covers the author's contribution to development and verification of high-level triggers for top quark physics with a semileptonic signature, where one of the W bosons decays into an electron and a neutrino. The description of top triggers, efficiency measurement, validation of jet energy corrections and pile-up subtraction applied at the HLT level are discussed in relevant sections of this chapter.

4.1 Level-1 triggers

The CMS L1 trigger [95] is built of custom electronics processing the data from the calorimeters and the muon system. It is the first filtering stage of any physics selection, therefore it has high requirements on efficiency and phase space acceptance. The input rate of events, i.e. the beam crossing frequency of $\sim 5 \times 10^8$ Hz must be reduced down to ~ 100 kHz. This represents a major challenge considering a gradual increase of instantaneous luminosity throughout the LHC operation. The rate budget of ~ 100 kHz is shared between several L1 triggers corresponding to different physics objects being present in the event. For top physics signatures with a single electron in the final state, the following triggers have been used:

- L1_SingleEG_18
- L1_SingleEG_20
- L1_SingleEG_22

These triggers are referred to as electron/photon (e/γ, or EG) triggers. They scan the ECAL for groups of trigger cells[1] (4×4) with a summed energy above the threshold that is suggested by the name of the trigger (18 GeV, 20 GeV and 22 GeV). The choice of threshold is determined by sustainability of the trigger rate at a given instantaneous luminosity.

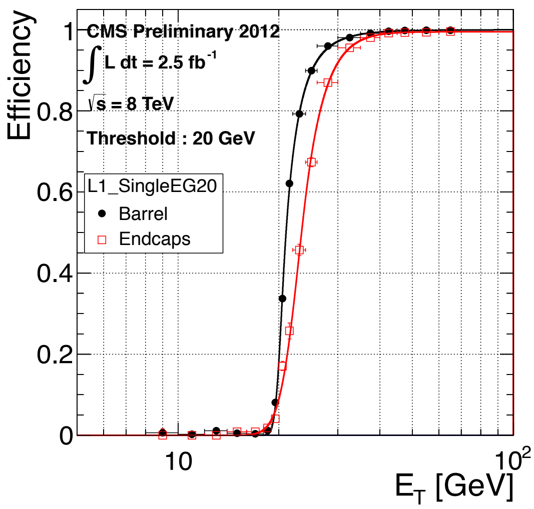


Figure 4.1: Efficiency of the 20 GeV e/γ trigger as a function of offline E_{T}, shown separately for barrel and endcap regions [96].

[1]Each trigger cell is formed by 5×5 crystals in the barrel region, corresponding to dimensions of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$; these numbers vary in the endcap region [95].

The efficiency of the L1_SingleEG_20 trigger as a function of the electron transverse momentum reconstructed offline is shown in Figure 4.1. Clearly, the turn-on curve of the trigger is sharper in the barrel region than in the endcap region. The plateau region of the curve (≈ 95 % efficiency) is reached at ≈ 30 GeV in the barrel and ≈ 35 GeV in the endcap region. This difference in trigger efficiency needs to be taken into account in the construction of the High-Level Trigger, as the L1 trigger decisions are used as input "seeds" to the HLT system, described in the following section.

4.2 High-Level Triggers

The High-Level Trigger [97] is the crucial part of the CMS event selection process. As mentioned in Section 3.2.6, it is based on software algorithms running on the Event Filter Farm, i.e. a large cluster of commercial CPUs. The HLT reconstruction, often referred to as online reconstruction, is implemented in the same software framework (CMSSW) which is used for offline reconstruction, and the algorithms can be very similar. However, the key difference between online and offline reconstruction is the running time. Since the HLT selection has to be performed in real time, it imposes a significant constraint on computing resources, enforcing compromises on the robustness and efficiency of online algorithms. With an exception of small samples for performance monitoring, data rejected by the HLT is lost irrevocably. Therefore, correct and efficient operation of the HLT is of major importance for the CMS physics programme.

The modular structure of CMSSW provides high flexibility in use of selection and reconstruction algorithms, allowing their continuous optimisation according to changes in physics needs and data-taking conditions. Various modules account for reconstruction of different physics objects and their matching with L1 objects, filtering, logging, monitoring, etc. All these modules are grouped into so-called trigger "paths", ultimately giving trigger decisions on whether to accept an event or not. A typical example of a trigger path used for top physics is an electron plus jets trigger, which requires the presence of isolated electron and at least three energetic jets in the event.

A trigger path can be prescaled or unprescaled. Prescaling is a method of reducing the trigger rate by only recording a fraction of events that pass the trigger selection. For example, if the prescale value of a particular trigger is 10, only one out of ten events that fire the trigger will be actually recorded. Usually prescaled triggers are used for performance monitoring and background estimation as they can impose a much looser selection criteria yet still have manageable rate. However, such triggers are unfavourable for signal selection, as they greatly reduce signal acceptance.

A set of trigger paths and their prescale factors is combined in the HLT configuration, referred to as the HLT menu or table [98]. Since the start of data-taking in 2010, the CMS trigger coordination adopted a "trigger train" model, implying the schedule of regular deadlines for updates in the trigger menu. These deadlines are usually imposed fortnightly, according to changes in instantaneous luminosity or other alterations in data-taking conditions. In order for a trigger path to be included in the trigger menu, it has to be implemented in the HLT configurations database, validated by measuring the trigger rate and signal efficiency, tested for CPU timing and finally approved by the trigger studies group.

As mentioned before, the top quark pair decay covered in this work has a semileptonic signature, containing an electron, at least four jets and a neutrino giving rise to $E_{\mathrm{T}}^{\mathrm{miss}}$. Due to the nature of the decay, all these objects are highly energetic and can be triggered on with rather high thresholds on transverse energy. During the start-up year of 2010 when the instantaneous luminosity went up from $\sim 10^{27}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ to $\sim 10^{31}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, top analyses with the signature of interest used the single electron trigger, requiring just one electron with certain isolation and transverse momentum criteria. However, by extrapolating the trigger rates on higher luminosities foreseen in the following years it became obvious that the single electron trigger rate would quickly become too high for the rate budget restrictions at the time. Therefore, alternative ways of adding other objects from the $\mathrm{t\bar{t}}$ signature were explored, naturally leading to electron plus jets trigger solution.

A typical electron plus jets trigger consists of a standard electron module, a jet module, a cleaning module to remove overlap between jet and electron collections, and a jet multiplicity filter. The electron module is constructed from the following sub-modules:

1. L1 object and ECAL super-cluster matching;
2. ECAL transverse energy filter;
3. ECAL electron ID and isolation filter;
4. HCAL electron ID and isolation filter;
5. Tracker electron ID and isolation filter.

To minimise the total running time, all the sub-modules are ordered by speed, starting from the fastest. The matching module (1) finds the ECAL super-cluster closest to the L1 object in η and ϕ dimensions. If the super-cluster is located within a certain η and ϕ range, the event is passed to the next module, otherwise it is rejected.

The following ECAL module (2) calculates the super-cluster transverse energy (E_{T}), and as long as it is above 25 GeV, the electron ID and isolation criteria are imposed using ECAL (3), HCAL (4) and the tracker (5) information (see Section 3.5.1).

The working points for identification and isolation criteria and their naming conventions are shown in Table 4.1. Working points studied for 2011 analysis (Chapter 5) were known by the following descriptive labels: VeryLoose (VL), Loose (L), Tight (T), and VeryTight (VT). In 2012, the working point known as WP80 was introduced for the use in single electron trigger, described later on.

Table 4.1: Naming conventions for electron trigger working points. The given values are for the barrel region of the detector and the values in brackets are for the endcap region. The used cut variables are explained in Section 3.5.

Working point	CaloID	CaloIso	TrkId	TrkIso
VeryLoose (VL)	$H/E < 0.15$ (0.10)	ECAL iso/$E_{\mathrm{T}} < 0.2$ (0.2)	$\Delta\eta < 0.01$ (0.01)	track iso/$p_{\mathrm{T}} < 0.2$ (0.2)
	$\sigma_{i\eta i\eta} < 0.024$ (0.040)	HCAL iso/$E_{\mathrm{T}} < 0.2$ (0.2)	$\Delta\phi < 0.15$ (0.10)	
Loose (L)	$H/E < 0.15$ (0.10)			
	$\sigma_{i\eta i\eta} < 0.014$ (0.035)			
Tight (T)	$H/E < 0.10$ (0.075)	ECAL iso/$E_{\mathrm{T}} < 0.125$ (0.075)	$\Delta\eta < 0.008$ (0.008)	track iso/$p_{\mathrm{T}} < 0.125$ (0.125)
	$\sigma_{i\eta i\eta} < 0.011$ (0.031)	HCAL iso/$E_{\mathrm{T}} < 0.125$ (0.125)	$\Delta\phi < 0.07$ (0.05)	
VeryTight (VT)	$H/E < 0.05$ (0.05)			
	$\sigma_{i\eta i\eta} < 0.011$ (0.031)			
WP80	$H/E < 0.10$ (0.05)	ECAL iso/$E_{\mathrm{T}} < 0.15$ (0.1)	$\Delta\eta < 0.007$ (0.007)	track iso/$p_{\mathrm{T}} < 0.05$ (0.05)
	$\sigma_{i\eta i\eta} < 0.01$ (0.03)	HCAL iso/$E_{\mathrm{T}} < 0.1$ (0.1)	$\Delta\phi < 0.06$ (0.03)	

In the main trigger path used for signal rather than background estimation in 2011 analysis, the Very Tight (VT) working point for calorimeter identification (CaloID) and the Tight (T) working point for calorimeter isolation (CaloIso) were used. This was motivated by consistency with similar criteria for the single electron trigger used in 2010.

For the events that pass all described criteria in the electron module, a simplified algorithm for offline iterative tracking is applied. This algorithm has a faster running time at the expense of worse resolution. In order to decrease the CPU usage, the Combinatorial Track Finder (CTF) [99] is used instead of the GSF one, and the tracking is confined to a small region around the electron. The events are required to pass the tight criteria of the tracker electron ID (TrkId) which is calculated at this stage. Finally, the tight working point of the tracker isolation (TrkIso) is applied, where the track p_{T} is summed over all tracks within $\Delta R = 0.3$ cone around the electron track, excluding the electron itself.

Provided that event passes all the electron sub-modules, the jet module is then executed. The jet reconstruction is performed with the anti-k_{t} algorithm (see Section 3.5.3) with a radius parameter of $R = 0.5$. Historically, at the early stages of LHC operation, CMS analyses mostly used calorimeter jets, i.e. jets reconstructed using calorimeter information exclusively. Therefore, initially the trigger jet module used the calorimeter jet collection, which provided fast reconstruction but worse resolution, hence lower trigger efficiency comparing to PF jets which were introduced later on during 2011 data-taking. In order to work online, Particle Flow (PF, Section 3.5.3.2) jet reconstruction was modified to be compatible with the simplified tracking algorithm. Despite the fact that running the jet module with PF jet reconstruction is highly CPU-intensive and therefore can not be used as an unprescaled standalone trigger, it works effectively together with the electron module since it reduces the input event rate considerably. Both calorimeter and PF jet versions of the module impose a 30 GeV cut on transverse momentum of the jets as well as a pseudorapidity cut of $|\eta| < 2.6$. Subsequently, the jet collection is passed to the cleaning module.

Within CMS, electron and jet collections are reconstructed independently, and since they often leave similar footprints in the detector, these collections can overlap, causing potential double-counting. The electron-jet cleaning module removes electron candidates found by the electron module from the collection of jet candidates found by the jet module. A jet is removed from the jet collection if there is an electron within a cone of $\Delta R = 0.3$ around the jet axis. However, occasionally a genuine jet can be incorrectly identified as an electron. In this case the fake electron can be removed from the jet collection, biasing the performance of the jet multiplicity filter. To tackle this issue, separate jet collections are created for all electrons in the event, where jets are only cleaned from additional (non-signal) electrons. All these jet collections are passed to the following jet multiplicity filter, which accepts an event if at least one collection contains a desired number of cleaned jets.

Electron plus jets triggers successfully functioned throughout 2011 and 2012 years of data taking. During 2011, they were used in the top mass analysis, described in this thesis. The 2012 cross section analysis used a single electron trigger with a lepton p_{T} threshold of 27 GeV, reasonably tight (WP80) lepton isolation requirements, but no specific jet requirements. In contrast to 2011 running, this trigger was decided to be unprescaled in the trigger menu for the whole period of data-taking in 2012, despite having a significant rate. It is favourable for many analyses as it is much more straightforward in terms of calculating efficiency and acceptance scale factors, also reducing the complexity of estimating the systematic errors associated with the triggering.

4.3 Trigger efficiency measurement

One of the most important characteristics of a trigger path is its efficiency. In a broad sense, selection efficiency can be defined as a conditional probability that a single event passes the selection, given all other conditions (detector configuration, preselection, etc). In the context of the trigger, efficiency highly depends on the offline selection it is measured with respect to. As a matter of fact, different studies can apply different offline selections, therefore the same trigger may have different efficiencies for each of them.

The trigger efficiency can be measured in both data and Monte Carlo simulation using various methods. For all of them, the study has to be performed on some initial preselected dataset. Ideally, this preselection needs to be unbiased with respect to the selection imposed by the trigger and offline selection. However, in reality it implies running on vast amounts of data, with very few events satisfying the trigger selection, leading to insufficient statistics or non-feasible computing resources needed for the study. Therefore, some other trigger is used for preselecting the initial dataset, somewhat correlated to the trigger under study in order to obtain reasonable statistics. To correctly measure the efficiency of a trigger path, this correlation has to be taken into account.

The efficiency of a trigger A can be written as the following conditional probability [100]:

$$\epsilon_A = P(A|\vec{x}, T, D) \tag{4.1}$$

where $\vec{x}$ are reference quantities used for triggering (e.g. lepton p_{T}), and T is an offline selection, and D is a combination of all other factors like detector effects. Assuming that these factors, along with the offline selection and quantities used for triggering are fixed, we can denote $\epsilon_A = P(A)$ for brevity.

If B is the trigger used for preselection, then

$$P(A, B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \tag{4.2}$$

Therefore, as it immediately follows from Bayes' theorem,

$$P(A) = \frac{P(A|B) \cdot P(B)}{P(B|A)} \tag{4.3}$$

Here the conditional probability $P(A|B)$ can be estimated by taking the ratio of the numbers of events that pass the triggers A and B, given that they all pass offline selection. Efficiency of the auxiliary trigger $\epsilon_B = P(B)$ can be estimated from data by using a dataset obtained with looser ("minimum bias") selection. Unfortunately, conditional probability $P(B|A)$ can not be measured using real data, as by the nature of the study the trigger A only considers events preselected by the trigger B. However, the trigger B is usually chosen in such a way that it is safe to assume that $P(B|A) = 1$ with negligible uncertainty. This is the case if the preselection trigger criteria are considerably looser than those of the trigger under study. To summarise, the trigger efficiency can be measured as:

$$\epsilon_A = P(A|B) \cdot P(B) = \frac{N_A}{N_B} \cdot \epsilon_B \tag{4.4}$$

where $N_{A(B)}$ is the number of events that pass the offline selection and fire the trigger A (B).

In case of electron plus jets triggers, the efficiency can be factorised in contributions of leptonic and hadronic parts of the trigger:

$$\epsilon_{ele+jets} = \epsilon_{ele} \cdot \epsilon_{jets} \tag{4.5}$$

This method can be used under assumption that the probability of finding an electron in the event is independent of the presence of the jets in the event, i.e. leptonic and hadronic "legs" of the trigger are uncorrelated. Although such correlations do exist, it has been shown that most of them are negligible and the factorisation method provides a meaningful estimate of the trigger efficiency [101].

The trigger efficiency can be parametrised by various physical quantities of the triggered objects, e.g. reconstructed jet pseudorapidity and transverse momentum. As the background spectrum and trigger rate fall down as a function of jet p_{T}, the trigger efficiency can be described by a turn-on curve approaching a plateau region, where the efficiency is maximum and constant within the statistical uncertainty. It is important for the trigger to have a sharp turn-on, with the nominal 95 % efficiency point being below or approximately at the cut value of offline reconstructed jet p_{T}. This can be achieved by adjusting the online cuts to be lower than offline ones: if these thresholds are close, then the trigger would bias the analysis distributions. In this case, non-flat p_{T}-dependent scale factors have to be applied to the total number of signal events, which also complicates estimation of the systematic uncertainty attributed to the trigger. As for the efficiency dependency on the jet angular distribution, it is expected to be flat – however, as it can be seen from the following section, this is not always the case.

4.4 Validation of jet energy corrections and charged hadron subtraction

In the end of 2011, to match the jet reconstruction method used by most CMS analyses, the calorimeter-based jet module in the electron plus jets HLT path was replaced with a PF-based one. This was done in an attempt to improve the p_{T} resolution of online jet reconstruction, reduce the rate and obtain sharper turn-on curves of the trigger efficiency. In the first iteration of PF-based trigger paths, no jet energy corrections were applied to the jets online, which resulted in the observation of a non-flat trigger efficiency with respect to jet pseudorapidity. After a few months of work, the responsible physics object group (known as “JetMET”) provided a series of jet energy corrections to be applied at the HLT level, which were tested by the author.

Pile-up turned out to be another major issue particularly during the 2012 data-taking period. Increased centre of mass energy and instantaneous luminosity at a bunch spacing of 50 ns effectively doubled the average number of pile-up vertices compared to 2011 conditions. This led to a significant increase in the rates of PF-jets triggers, requiring implementation of pile-up mitigation techniques at the HLT level. As a result, charged hadron subtraction (CHS) for PF jets, also known as “PFnoPU”, was introduced online. As the name suggests, it essentially removes the charged hadrons associated with secondary vertices from PF candidates, which helped to reduce the trigger rate dramatically. As top triggers operate with PF jets, they were greatly affected by pile-up, therefore this correction also had to be validated for the electron plus jets triggers.

In this work, the impact of both jet energy corrections and pile-up subtraction on the electron plus jets trigger efficiency was investigated. A rather complicated name of the main signal trigger under study (HLT_Ele25_CaloIdVT_CaloIsoT_-TrkIdT_TrkIsoT_TriCentralPFJet30) suggests working points of tracker and calorimeter ID and isolation criteria for the signal electron (see Table 4.1), as well as transverse momenta requirements for both electron (25 GeV) and each of the three particle flow jets (30 GeV). Since only the hadronic part of the trigger was of interest for this particular study, the efficiency was measured using a “Single Electron” primary dataset of raw 2012 data. This primary dataset was constructed by accepting events passing at least one of the several single electron triggers with various p_{T} and isolation requirements. One of the loosest triggers, referred to as HLT_Ele27_-WP80, operates with working point “WP80”, also shown in Table 4.1, and electron p_{T} threshold of 27 GeV.

The offline selection used to obtain a clean $\mathrm{t\bar{t}}$ sample is outlined in Section 5.2, with the exception of the jet multiplicity requirement: here at least three jets were required. A tighter requirement of at least four jets was also of interest, as described below. The trigger efficiency is then calculated as:

$$\epsilon = \frac{N_{\text{fired}}}{N_{\text{selected}}} \tag{4.6}$$

where N_{selected} is the number of events from the primary dataset passing the described offline selection, and N_{fired} is the number of events which in addition fired the electron plus three jet trigger.

All the efficiencies were measured in bins of jet p_{T} and η. For these variables, different fit functions were used. The efficiency in p_{T} was fitted with the default function describing the turn-on curve:

$$\epsilon(p_{\mathrm{T}}) = A \times e^{\left(B \times e^{C \times p_{\mathrm{T}}}\right)} \tag{4.7}$$

while the efficiency in η was fitted with a quadratic form:

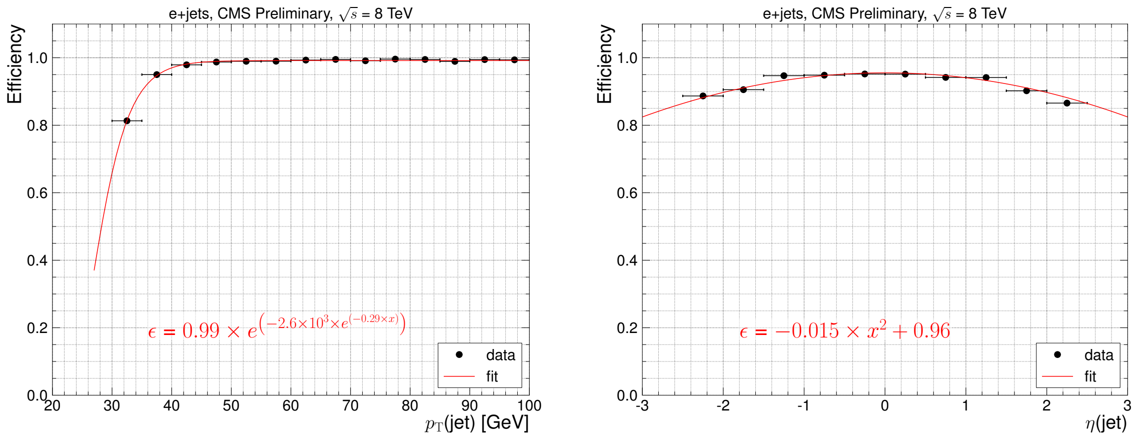
$$\epsilon(\eta) = A \times \eta^2 + B \times \eta + C \tag{4.8}$$

where A, B and C are the fit parameters.

Although the η dependency was expected to be flat, the observed structure in the η distribution of the PF jet triggers suggested the use of a parabolic fit function. In fact, this drop of the efficiency in the endcap region was the main motivation for the study, as it was believed to be due to the fact that the p_{T}- and η-dependent jet energy corrections were not applied online, causing inconsistency between HLT and offline reconstructed jets. The efficiency of the main electron plus jets trigger with no jet energy corrections applied is shown on Figure 4.2. Clearly, a parabolic shape is seen on the efficiency plot with respect to pseudorapidity of a third reconstructed jet (all the jets were sorted by the value of transverse momentum).

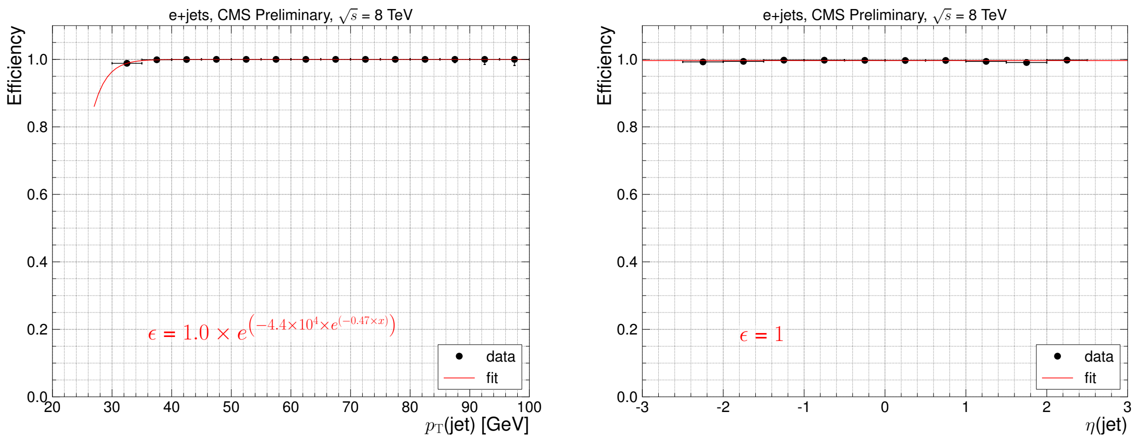
It has to be emphasised that only efficiencies with respect to offline selection with at least three jets were affected. Requiring a fourth jet in the event increases the trigger efficiency to nearly 100 %, as can be seen on Figure 4.3. This implies that observed inefficiency in the endcap region was not critical for analyses requiring at least four jets in the event (including the ones in this thesis), however, it was still important to investigate this effect in order to better understand any possible systematic errors due to the triggering.

For the purpose of the study, a modified HLT menu was created, containing new versions of the main signal trigger with jet energy corrections and charged hadron subtraction applied. This menu was run on a $\mathrm{t\bar{t}}$ skim of a RAW dataset mentioned



(a) (b)

Figure 4.2: Trigger efficiency for HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_-TrkIsoT_TriCentralPFJet30 (uncorrected) as a function of the third jet p_{T} (a) and η (b) for events with at least three jets.

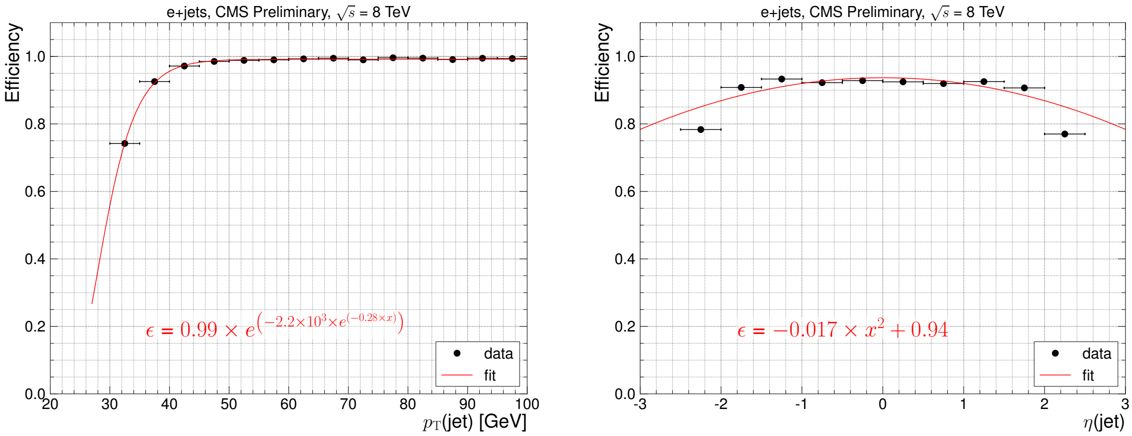


(a) (b)

Figure 4.3: Trigger efficiency for HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_-TrkIsoT_TriCentralPFJet30 (uncorrected) as a function of the third jet p_{T} (a) and η (b) for events with at least four jets.

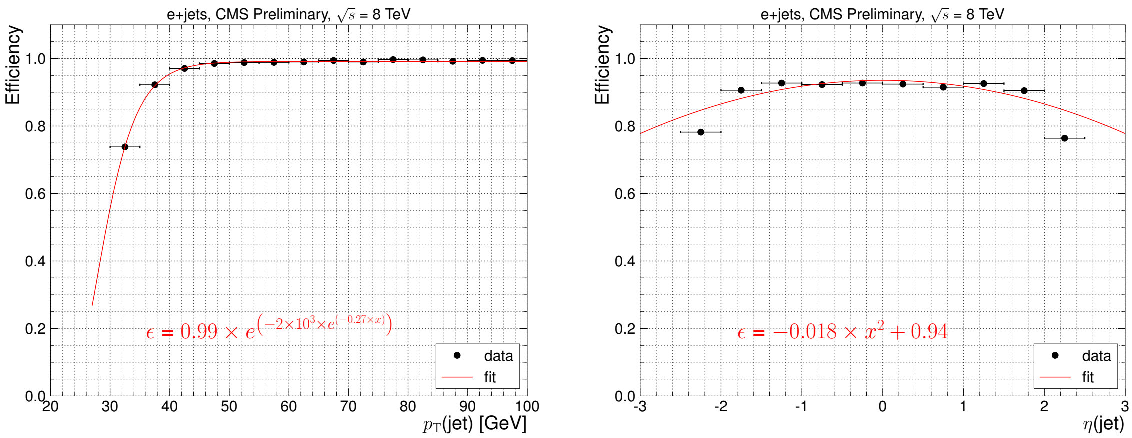
above, and the resulting trigger flags as well as jet collections were used to produce the final plots.

Figure 4.4 presents the results for the aforementioned trigger path but with jet energy corrections applied, and Figure 4.5 shows it for the same path with both charged hadron subtraction and jet corrections applied. It is obvious that all these corrections haven't solved the inefficiency problem in the endcaps, which caused some confusion in the group that provided them and triggered additional scrutiny from the trigger and jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ experts.



(a) (b)

Figure 4.4: Trigger efficiency for HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_-TrkIsoT_TriCentralPFJet30 with jet energy corrections applied online, as a function of the third jet p_{T} (a) and η (b) for events with at least three jets.



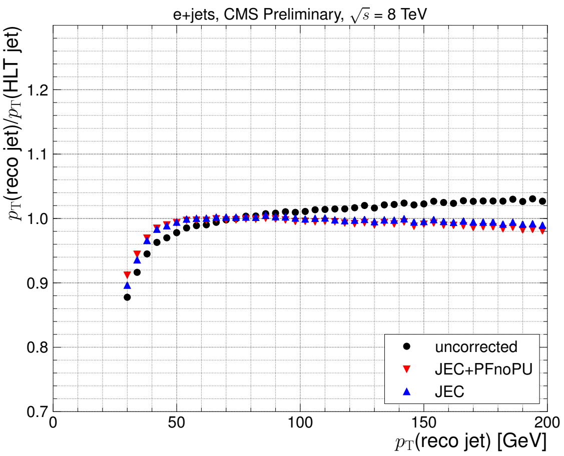
(a) (b)

Figure 4.5: Trigger efficiency for HLT_Ele25_CaloIdVT_CaloIsoT_TrkIdT_-TrkIsoT_TriCentralPFJet30 with jet energy corrections and charged hadron subtraction applied online, as a function of the third jet p_{T} (a) and η (b) for events with at least three jets.

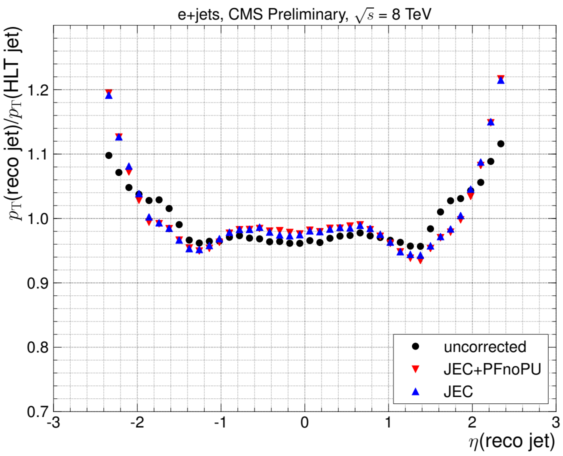
In order to gain further understanding, the jet response distributions were investigated. Jet response is a ratio of the transverse momenta of offline-reconstructed and HLT jets, ideally produced from the same objects in the detector. To find this ratio, all the HLT jets were matched with offline jets found within a cone of $\Delta R = 0.3$ for each online jet. Matching efficiency proved to be close to unity, and subsequent distributions were produced only for the successfully matched jets. Figure 4.6 shows the response plots for trigger paths with different corrections: uncorrected, only jet energy corrected (JEC) and with both jet energy corrections and charged hadron subtraction (JEC+CHS) applied. It can be seen that corrections work reasonably well in the barrel region, but substantially underestimate the energy of online jets in the endcap region. The response plot shown in slices of offline jet transverse momenta (c) suggests that corrections work better for jets with higher energy, which can also be seen from the distribution with respect to jet p_{T}, however, the overall response in the endcap region is still far from being perfect.

Shorty after these results were presented to the trigger coordination, they were confirmed independently by other experts within the JetMET group. Because of the time constraints associated with the luminosity increase schedule, a following workaround solution was suggested. In order to re-gain the trigger efficiency, the threshold for the third jet was dropped from 30 GeV to 20 GeV, which allowed application of jet energy corrections and charged hadron subtraction with a minimum loss in the efficiency in the endcap region.

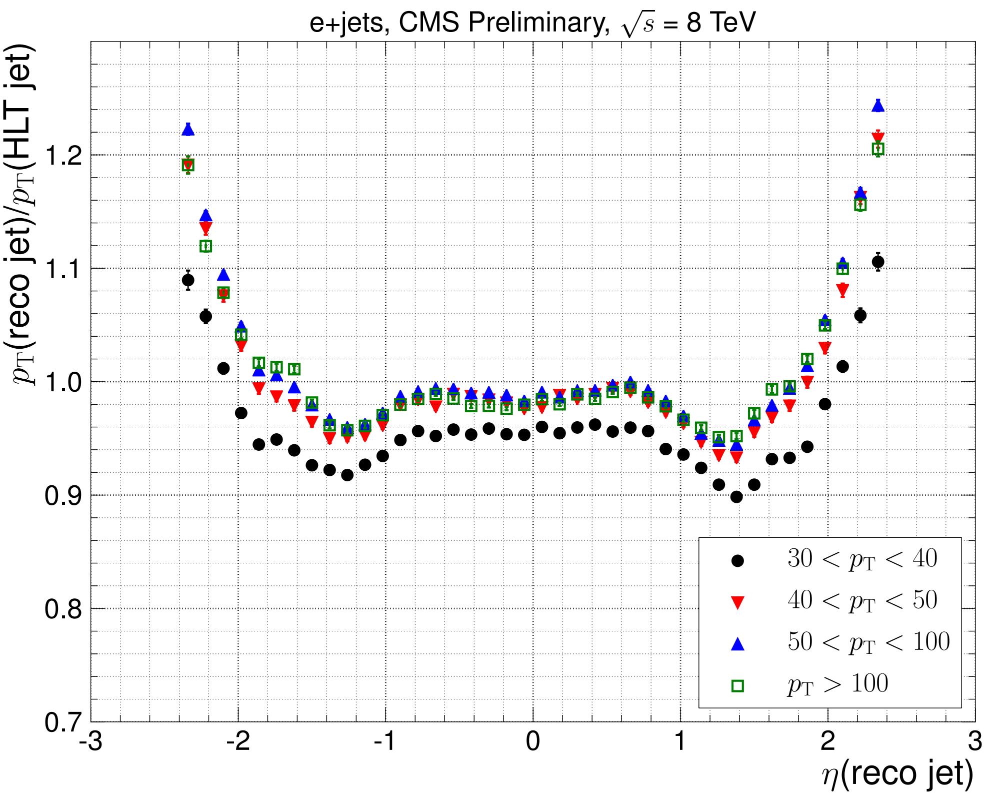
At the time of writing this thesis, the source of inadequacy of the jet response in the endcap region remains not fully understood. One of the possible explanations is based on the observation that the number of fake tracks reconstructed online is substantially larger than that in the offline reconstruction, which can potentially bias the PF jet reconstruction on the HLT level. However, a full solution to this problem (and therefore derivation of adequate jet energy corrections for online application) is still being sought for by a relevant group of experts.



(a)



(b)



(c)

Figure 4.6: Ratio of matched offline and online reconstructed jet transverse momenta as a function of offline jet p_T (a) and η (b) for different jet corrections and in slices of offline jet p_T (c).

4.5 Summary

This chapter covered the service work performed by the author for the CMS collaboration, namely development, validation and maintenance of high-level triggers for top quark physics, specifically for the electron channel of semileptonic $t\bar{t}$ decay signature. The importance and typical challenges of trigger development have been discussed. The modular structure of high-level triggers has been described in detail for electron plus jets trigger paths. A core of author's contribution: the investigation of the effect of jet energy corrections applied on the HLT level to top trigger paths has been presented, showing how it affects trigger efficiency and jet response distributions. An impact of charged hadron subtraction applied online has also been studied and demonstrated.

5. Top Quark Mass Measurement

The top quark mass measurement using 2011 LHC data recorded by the CMS detector at a centre of mass energy of $\sqrt{s} = 7$ TeV is presented in this chapter. The analysis was a cross-check to the official CMS top quark mass measurement at 7 TeV in the lepton plus jets channel published in 2012 [102], which currently remains the most precise single measurement of the top quark mass. Only the electron plus jets channel is described in this thesis, as the muon side of the analysis was performed by a different group at CERN, although in close collaboration.

The mass extraction technique used in this analysis, referred to as the ideogram method, is essentially the same as the one used in the CMS measurement of the mass difference between top and antitop quarks [103] and the top quark mass measurement in 2010 [104]. In this chapter, this technique is described in detail. Additionally, data and Monte Carlo samples, event selection process, the kinematic fit used for $\mathrm{t\bar{t}}$ reconstruction, as well as the calibration of the ideogram method, systematic uncertainties and the final result are presented in relevant sections of the chapter.

5.1 Data and Simulation

5.1.1 Data

The data used in this work is the full 2011 dataset recorded by the CMS detector with a total integrated luminosity of 5.0 ± 0.1 fb^{-1}. As only the electron channel is covered by this particular analysis, all data were pre-selected by single electron plus jets high-level triggers, described in detail in Chapter 4.

5.1.2 Monte Carlo samples

The nominal signal $\mathrm{t\bar{t}}$ MC sample used in this analysis was generated using the MadGraph generator with a top quark mass of $m_{\mathrm{t}} = 172.5$ GeV, interfaced with parton showering implemented in PYTHIA. To calibrate the mass extraction technique, as explained in Section 5.5, 8 additional $\mathrm{t\bar{t}}$ samples were used with different top quark masses ranging between 161.5 GeV and 184.5 GeV. The W + jets and Z + jets background processes (see Section 2.2.3) were also generated with MadGraph + PYTHIA, whereas the single top background was generated with POWHEG generator, also interfaced with PYTHIA for parton showering. A full list of signal $t\bar{t}$ and background MC samples with values of cross section, numbers of generated events and corresponding integrated luminosities is shown in Table 5.1. All samples were produced using the CTEQ6L Parton Distribution Functions (PDF) [105].

The QCD background samples are available in two different sets, one of which is pre-filtered to contain jets with high electromagnetic content (e/γ enriched), and the second one is enriched with electrons coming from decays of heavy flavour quarks ($b/c \rightarrow qe\nu$). This pre-selection is performed at generator level, to ensure reasonable statistics after the $t\bar{t}$ selection imposed in top quark analyses with electrons in the final state. Both sets of QCD samples are generated with PYTHIA generator in different exclusive bins of $\hat{p}_{\mathrm{T}}$ (i.e. transverse momentum of the outgoing partons defined in the rest frame of the hard process). The γ+jets MC samples were produced with MadGraph in different bins of H_{T} at generator level, which denotes the sum of transverse momenta of all high-p_{T} objects. Table 5.2 gives a list of QCD and γ+jets samples with values of production cross sections, pre-selection efficiencies, numbers of generated events and corresponding integrated luminosities.

Finally, Table 5.3 lists the additional samples generated to allow estimation of the systematic uncertainties due to factorisation scale and matching threshold choices, as discussed in Section 3.4.1. A summary of central values and variations used to estimate these systematic uncertainties is shown in Table 5.4.

5.2 Event Selection

The event selection applied in this analysis essentially follows the so-called reference selection, i.e. the one recommended by the CMS Top Quark Physics Analysis Group for SM top quark measurements. It is designed to select the semileptonic signature of a $t\bar{t}$ decay with an isolated lepton, four jets (including two jets from b-quarks) and a neutrino in the form of missing transverse energy. All objects are reconstructed using the particle flow method as described in Section 3.5.

As this analysis focuses on the electron channel, the event selection proceeds as follows:

1. preselection;
2. trigger;
3. electron candidate selection;
4. dilepton veto;
5. conversion veto;
6. muon veto;
7. jet selection;
8. b-tagging.

Table 5.1: Signal and background Monte Carlo samples with cross sections at $\sqrt{s} =$ 7 TeV, numbers of generated events and corresponding integrated luminosities.

Process	Generator	σ (pb)	# events	$\int \mathcal{L}dt$ (fb^{-1})
$\mathrm{t\bar{t}} + \mathrm{jets}$	MadGraph + PYTHIA			
$m_\mathrm{t} = 161.5$ GeV		157.5	1620072	10.3
$m_\mathrm{t} = 163.5$ GeV		157.5	1633197	10.4
$m_\mathrm{t} = 166.5$ GeV		157.5	1669034	10.6
$m_\mathrm{t} = 169.5$ GeV		157.5	1606570	10.2
$m_\mathrm{t} = 172.5$ GeV		157.5	7490162	47.6
$m_\mathrm{t} = 175.5$ GeV		157.5	1538301	9.8
$m_\mathrm{t} = 178.5$ GeV		157.5	1648519	10.5
$m_\mathrm{t} = 181.5$ GeV		157.5	1665350	10.6
$m_\mathrm{t} = 184.5$ GeV		157.5	1671859	10.6
$\mathrm{W} + \mathrm{jets}$ ($\mathrm{W} \rightarrow l\nu$)	MadGraph + PYTHIA	31314	81345381	2.6
$\mathrm{Z}/\gamma^* \rightarrow l^+l^-$ + jets, $m(ll) > 50$ GeV	MadGraph + PYTHIA	3048	36222153	11.9
Single top	POWHEG + PYTHIA			
top t-channel		42.6	3814228	89.5
anti-top t-channel		22.0	1944822	88.4
top s-channel		2.72	259971	92.6
anti-top s-channel		1.49	137980	92.6
top tW-channel		5.3	814390	153.7
anti-top tW-channel		5.3	809984	152.8

Table 5.2: QCD multi-jet background and γ + jets MC samples with cross sections at $\sqrt{s} = 7$ TeV, numbers of generated events and corresponding integrated luminosities.

Process	Generator	σ (pb)	filter efficiency	# events	$\int \mathcal{L}dt$ (fb^{-1})
QCD (e/γ enriched)	PYTHIA				
20 GeV $< \hat{p}_\mathrm{T} <$ 30 GeV		2.355×10^8	7.3×10^{-3}	34720808	2.0×10^{-2}
30 GeV $< \hat{p}_\mathrm{T} <$ 80 GeV		5.93×10^7	0.059	70375915	2.0×10^{-2}
80 GeV $< \hat{p}_\mathrm{T} <$ 170 GeV		9.06×10^5	0.148	8150669	6.1×10^{-2}
QCD ($\mathrm{b/c} \rightarrow e\nu$)	PYTHIA				
20 GeV $< \hat{p}_\mathrm{T} <$ 30 GeV		2.355×10^8	4.6×10^{-4}	2002588	1.8×10^{-2}
30 GeV $< \hat{p}_\mathrm{T} <$ 80 GeV		5.93×10^7	2.34×10^{-3}	2030030	1.5×10^{-2}
80 GeV $< \hat{p}_\mathrm{T} <$ 170 GeV		9.06×10^5	0.0104	1082690	0.1
γ + jets	MadGraph				
40 GeV $< H_\mathrm{T} <$ 100 GeV	+PYTHIA	23 620	1	12730863	0.5
100 GeV $< H_\mathrm{T} <$ 200 GeV		3476	1	1536287	0.4
$H_\mathrm{T} >$ 200 GeV		485	1	9377168	19.3

Table 5.3: Systematic MC samples with cross sections at $\sqrt{s} = 7$ TeV, numbers of generated events and corresponding integrated luminosities. Factorisation scale Q and matching threshold systematic uncertainties (see Section 3.4.1) are estimated with variations of $\mathrm{t\bar{t}}$ + jets, W + jets and Z + jets samples.

Process	Generator	σ (pb)	# events	$\int \mathcal{L}dt$ (fb^{-1})
$\mathrm{t\bar{t}}$ + jets	MadGraph + PYTHIA			
0.5 $\times$ matching threshold		157.5	1607808	10.2
2 $\times$ matching threshold		157.5	4029823	25.5
$0.5 \times Q$		157.5	4004587	25.4
$2 \times Q$		157.5	3696269	23.4
W + jets (W $\to l\nu$)	MadGraph + PYTHIA			
0.5 $\times$ matching threshold		29690	9956679	0.3
2 $\times$ matching threshold		30290	10461655	0.3
$0.5 \times Q$		33300	10092532	0.3
$2 \times Q$		32000	9784907	0.3
Z + jets (Z $\to ll$)	MadGraph + PYTHIA			
0.5 $\times$ matching threshold		2888	1614808	0.6
2 $\times$ matching threshold		2915	1641121	0.6
$0.5 \times Q$		3312	1658855	0.5
$2 \times Q$		2954	1592742	0.5

Table 5.4: Central values and variations of factorisation/renormalisation scale (Q^2) and matrix element to parton shower (ME-PS) matching in MadGraph MC samples with PYTHIA parton showering.

parameter	factorisation scale	matching threshold
	$\mathrm{t\bar{t}}$ events	
central value	$Q^2 = m_\mathrm{t}^2 + \sum_{\mathrm{jets}} p_\mathrm{T}^2$	20 GeV
variations	$(0.5 \cdot Q)^2, (2 \cdot Q)^2$	10 GeV and 40 GeV
	W + jets and Z + jets events	
central value	$Q^2 = m_\mathrm{W/Z}^2 + \sum_{\mathrm{jets}} p_\mathrm{T}^2$	10 GeV
variations	$(0.5 \cdot Q)^2, (2 \cdot Q)^2$	5 GeV and 20 GeV

Preselection

Preselection is performed in order to reduce the number of events stored for local analysis. All events are required to have at least one electron with E_{T} above 30 GeV within pseudorapidity region of $|\eta| < 2.5$, or at least one muon with p_{T} above 20 GeV and $|\eta| < 2.1$. Additional event cleaning is performed at this stage to decrease the number of events with substantial detector noise. A small fraction of events with anomalous HCAL noise is removed by an HCAL noise filter. Furthermore, contribution from beam halo (mentioned in Section 6.2.3) is mitigated by the appropriate filter requiring that at least 25 % of tracks be of high purity if there are more than 10 tracks in the event. Finally, the preselection stage includes the requirement of at least one good primary vertex in the event which has to be located within 24 cm in the z-direction from the centre of CMS and within a radial distance of 2 cm. The primary vertex is also required to have at least four degrees of freedom which is determined from the vertex fit and is strongly correlated with the number of tracks compatible with the primary interaction region [106].

Trigger

The trigger requirement is imposed as described in Chapter 4. Essentially, all the events are required to fire the electron-plus-three-jets trigger in the version current at the time of data-taking. A full list of triggers can be found in Appendix A.

Electron candidate selection

Exactly one electron candidate satisfying all the following criteria is required to be present in the event. As the electron plus jets trigger has a lepton p_{T} threshold of 25 GeV, the electron candidate transverse momentum is required to be above 30 GeV, in order to be in the trigger efficiency plateau region. Furthermore, the electron has to be within the tracker region of $|\eta| < 2.5$ excluding the ECAL barrel-endcap transition regions of $1.4442 < |\eta| < 1.566$. The CiC electron ID (see Section 3.5.1) with the tightest working point ("HyperTight") is used for electron identification purposes. This working point provides a misidentification rate of less than 1 %, and overall identification efficiency of 75 % [89].

In order to only select electrons originating at the interaction vertex, the x-y distance (d_{xy}) between the electron track and the interaction point (2D impact parameter) is required to be less than 0.02 cm. The contribution of electrons inside jets from QCD background events, as well as fake electrons, is reduced by imposing the PF-based relative isolation cut on the variable reliso defined in Section 3.5.1.2. A ΔR-cone with size of 0.3 is used in calculation of the reliso variable, and events with reliso < 0.1 are accepted.

Dilepton veto

A veto requirement on a second electron candidate (or dilepton veto) is used to reject events with any additional electrons. Looser criteria are used to identify the second electron in an event. Namely, the event is rejected if there is a second electron satisfying a lower E_{T} threshold of 20 GeV and a looser cut on reliso (< 0.2).

Additionally, in order to reduce Z+jets background, the following veto is applied. If the event contains a second electron with $E_{\mathrm{T}} > 30$ GeV and reliso < 1.0 that forms such an invariant mass with the signal electron candidate that it is close to the Z mass peak, it is also rejected. The invariant mass window around the Z peak is chosen to be 76 GeV $< m_{ee} <$ 106 GeV.

Conversion veto

As mentioned in Section 3.5.1.3, two methods are used in this analysis to remove electrons coming from photon conversions: missing pixel layers method and partner track matching method. If missing pixel layer hits or a matching partner track are present in the event, it is rejected.

Muon veto

Other $\mathrm{t\bar{t}}$ channels, including muon and dilepton decay modes, can contaminate events passing the electron plus jets selection. To reduce this contamination, events containing an isolated global muon (see Section 3.5.2) are rejected. The muon is required to have a p_{T} above 10 GeV, $|\eta| < 2.5$ and reliso < 0.2 with a ΔR cone of 0.4.

Jet selection and b-tagging

Final steps in the selection help to further reduce the background by imposing constraints on the number of jets. This is particularly effective to mitigate the $\mathrm{W + jets}$ contamination, as the number of $\mathrm{W + jets}$ events decreases exponentially with increasing number of jets. In this analysis, at least four jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ are required to be present in the event. These jets have to pass the loose PF jet ID (see Section 3.5.3). Finally, the CSV b-tagging algorithm with medium working point (Section 3.5.3.3) is used to identify the two b-quarks from the $\mathrm{t\bar{t}}$ decay.

In contrast to the official CMS top quark measurement [102] that was cross-checked by this analysis, there is no specific requirement on the number of b-tagged jets in the event in order to increase available statistics after full event selection. Along with the difference in the joint likelihood fit which will be described in Section 5.4, these are the main two differences between the main measurement and the cross-check.

Table 5.5 shows the effect of the event selection on the number of events in each contributing process, the total expected number of events and the number of events observed in data. A few kinematic distributions, showing the comparison between data and Monte Carlo simulation after applying the described selection criteria are shown in Figure 5.1. Overall, a good agreement is observed at this stage. However, the final selection criterion is applied later on in the analysis after the kinematic fit, which is discussed in the next section.

Table 5.5: Number of expected and observed events from MC and data after selection, before the fitting process. The $\mathrm{t\bar{t}}+\mathrm{jets}$ MC sample shown was generated with $m_{\mathrm{t}} = 172.5$ GeV.

	$\mathrm{t\bar{t}}+$ **jets**	$\mathrm{W}+$ **jets**	$\mathrm{Z}+$ **jets**	**Single top**	**QCD**	**Sum MC**	**Data**
Events	21425	2504	558	1248	2484	28220	27340
Fraction	75.9%	8.9%	2.0%	4.4%	8.8%	100%	–

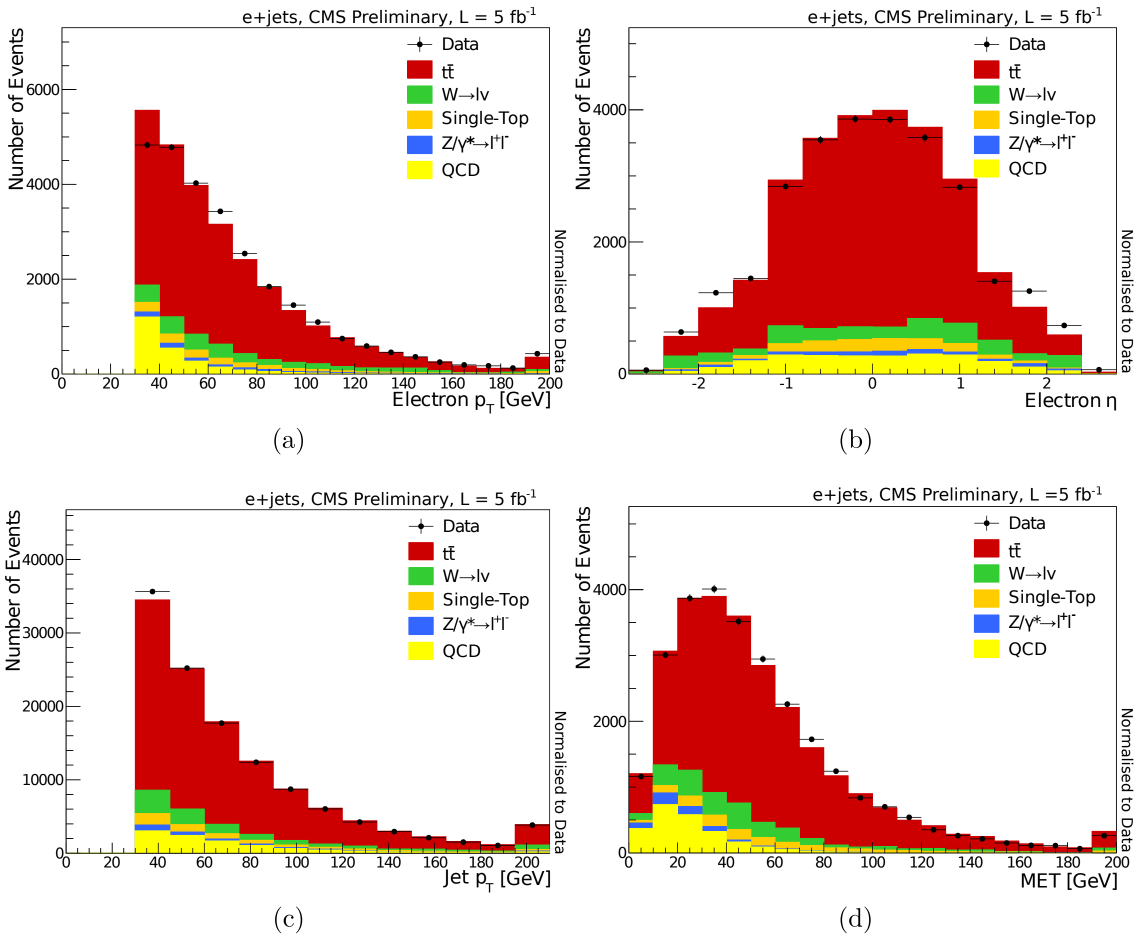


Figure 5.1: Kinematic variable distributions after all selection cuts: (a) electron p_{T}, (b) electron η, (c) jet p_{T} for all jets passing the selection, (d) $E_{\mathrm{T}}^{\mathrm{miss}}$.

5.3 Kinematic Fit

In order to precisely measure the top quark mass, objects in the final state need to be associated with the originating partons of the top quark pair decay. In this analysis, a kinematic fit is employed to fully reconstruct the event kinematics under the $\mathrm{t\bar{t}}$ hypothesis, thus improving the resolution of measured quantities by exploiting the knowledge of decay process.

The HitFit fitting package [107] used in this work originates from the DØ collaboration. Based on the SQUAW algorithm [108] developed in Lawrence Berkeley National Laboratory, the original HitFit code was written by Scott Stuart Snyder in Fortran for Run I of the Tevatron. It was successfully used in the first direct measurement of the top quark mass by DØ for lepton plus jets $\mathrm{t\bar{t}}$ events [109]. The package was then migrated to C++ for Run II of the Tevatron, and exploited in a series of $\mathrm{t\bar{t}}$ analyses including the updated top quark mass measurement using the ideogram method [110], first measurement of the forward-backward charge asymmetry in $\mathrm{t\bar{t}}$ production [111] and some others.

One of the key features of the HitFit package is its independence of specific experiments and detector geometries. The only external dependence is on CLHEP [112], a C++ library that provides utility classes for linear algebra, geometry, vector arithmetic, etc., which was specifically developed for high energy physics simulation and analysis software and is notably used by the Geant4 package. In 2011 HitFit was ported to CMSSW and later on successfully used in a number of CMS top quark analyses, including the top mass measurement in the lepton plus jets channel [102].

HitFit is designed to fit lepton plus jets events under the $\mathrm{t\bar{t}}$ hypothesis. Specifically, it strives to constrain the event to a hypothesis of production of two heavy particles, each decaying into a W boson and a b quark. According to the semileptonic signature, one of the W boson decays into an electron-neutrino pair, while the other decays into a light quark-antiquark pair.

The input to the fit includes the four-momenta of the lepton, four leading jets and missing transverse energy, as well as their respective resolutions that were calculated using Monte Carlo samples. Since only four leading jets are used, there are $4! = 24$ ways to associate these four reconstructed jets with the partons from the $\mathrm{t\bar{t}}$ decay. The number of permutations is reduced to $4!/2 = 12$ since the light jets from the hadronic W decay are interchangeable. However, for each hypothesis the z component of neutrino four-momentum needs to be determined. This is done by requiring the two heavy particles to have equal mass ($m_{\mathrm{t}} = m_{\mathrm{\bar{t}}}$), which yields a quadratic equation with either two real solutions, or two complex solutions that share the real part. In the case of the two real solutions, the fit is performed for each of them. If there are two complex solutions, the fit is performed once for the common real part of them, but the fit result is included in the event content twice. Therefore, each hypothesis has two possible values of the neutrino four-momentum, which doubles the number of permutations back to $4! = 24$.

The kinematic fit is performed by minimising the χ^2 function defined as

$$\chi^2 = (\mathbf{x} - \mathbf{x}^m)^T \mathbf{G} (\mathbf{x} - \mathbf{x}^m) \tag{5.1}$$

where $\mathbf{x}^m$ is the vector of measured observables, $\mathbf{x}$ is the vector of fitted variables and $\mathbf{G}$ is the inverse of the error matrix given by the resolutions of the observables.

The following kinematic constraints are imposed:

1. Reconstructed top quarks from both hadronic and leptonic legs are required to have equal masses:

$$m(\mathrm{t} \to \ell\nu\mathrm{b}) = m(\bar{\mathrm{t}} \to \mathrm{q}\bar{\mathrm{q}}\bar{\mathrm{b}}), \tag{5.2}$$

2. The W boson masses reconstructed in both hadronic and leptonic legs are fixed:

$$m(\ell\nu) = m(\mathrm{q}\bar{\mathrm{q}}) = m_{\mathrm{W}} = 80.4 \text{ GeV}. \tag{5.3}$$

The full description of the fitting algorithm can be found in the appendix of the HitFit author's PhD thesis [113]. Essentially, since the constraints are non-linear, an iterative technique is used. Starting with the measured observables, the constraint equations are linearised by expansion in a power series around the starting point. The minimisation is then solved with these linearised constraints, providing the new starting value for the next minimisation step which is repeated until the χ^2 converges or a maximum number of iterations is reached. Events are rejected if there is no permutation resulting in a converging fit.

Typically, an event has a few permutations resulting in a converged fit. The simplest approach of picking a permutation with the best χ^2 value is disfavoured, since Monte Carlo studies have shown that the correct permutation does not necessarily have the best χ^2 [113]. This value can change drastically with relatively small variations of the input parameters, as different permutations may obtain the smallest value of χ^2. On the other hand, some solutions yield very large χ^2 value, which typically happens for badly reconstructed or background events. Therefore, a loose cut of $\chi^2 = 20$ (rejecting approximately 15 % of combinations) is applied for all permutations on the output of the converged fit. This is the final selection criterion on top of the event selection described in Section 5.2.

To cross-check the performance of the kinematic fitter, the probability of finding the correct permutation was estimated by matching fitted objects to the true $t\bar{t}$ decay objects obtained using MC truth information in the $t\bar{t}$+jets sample. Figure 5.2 shows the residuals of the fitted top quark mass, i.e. the difference between fitted and true masses divided by the true generated top quark mass. This distribution is plotted for all fitted combinations with the best χ^2, and just those correctly matched to the true $t\bar{t}$ decay objects. It was found that the efficiency of the kinematic fitter to assign the correct permutation with the best χ^2 value is approximately 12 %. Some kinematic variables for permutations giving the best (smallest) χ^2 values are shown in Figure 5.3.

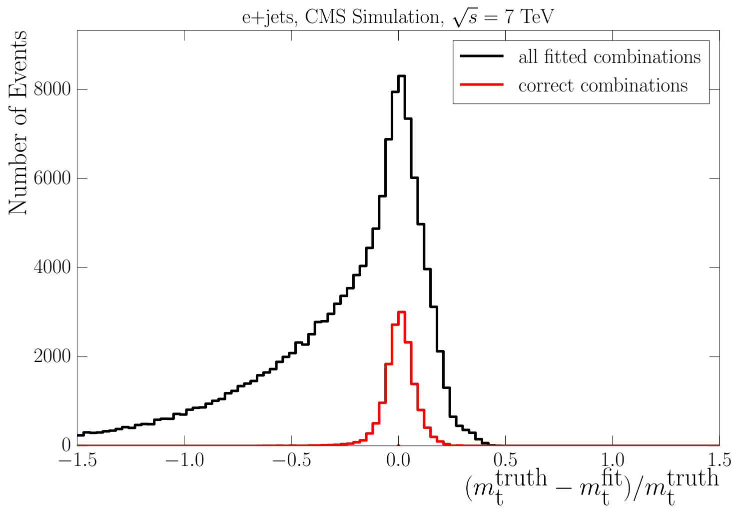


Figure 5.2: The difference between fitted and true top quark masses normalised by the true top quark mass for all best fitted combinations (black) and correctly found ones (red) as per Monte Carlo truth information from the $t\bar{t}$+jets sample generated with $m_t = 172.5$ GeV.

The compatibility of a permutation with the $t\bar{t}$ hypothesis is quantified by the following fit probability:

$$P_{\text{fit}} = \exp(-\frac{1}{2}\chi^2). \tag{5.4}$$

This probability is used in the next step of the analysis, referred to as the ideogram method.

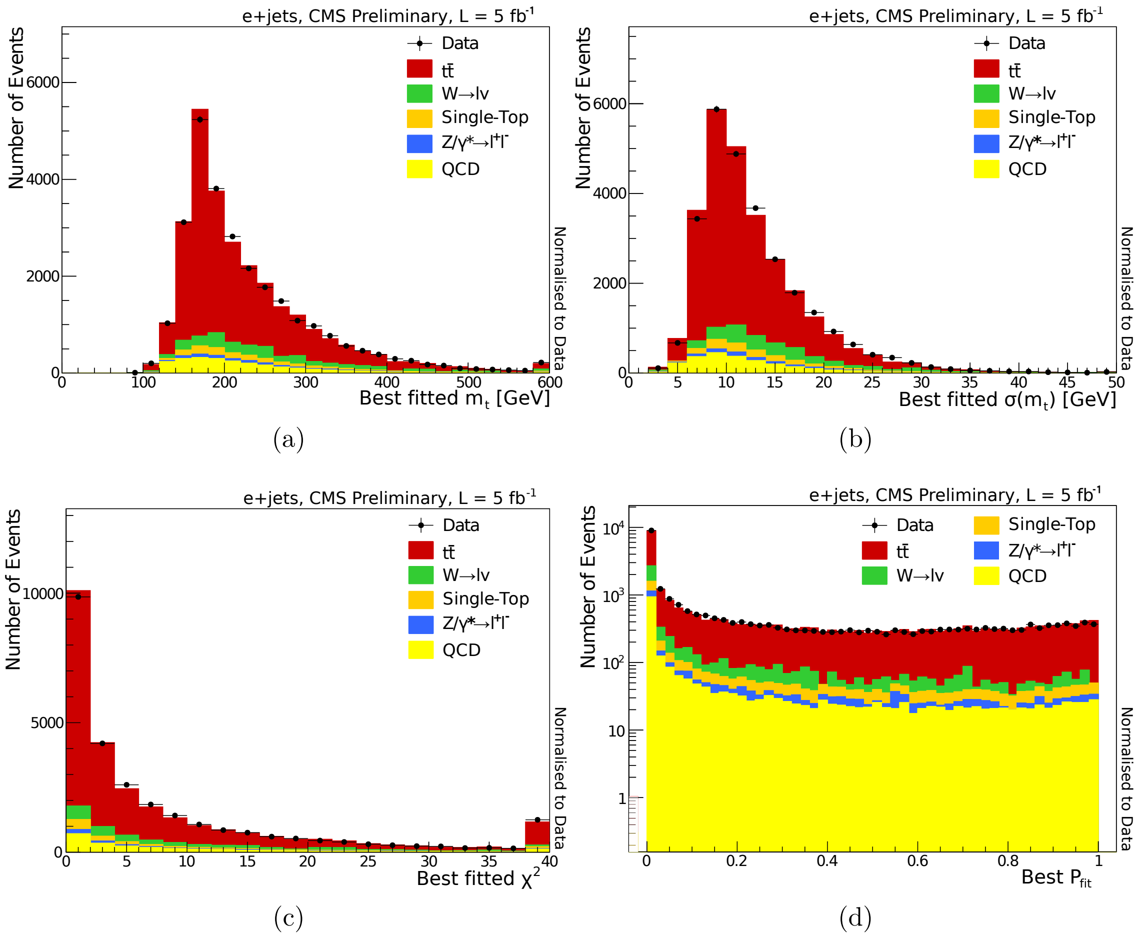


Figure 5.3: Distributions of (a) fitted top quark mass, (b) its standard deviation, (c) χ^2, and (d) fit probability for the best fitted hypotheses.

5.4 The Ideogram Method

The ideogram method is a powerful mass extraction technique. Its main idea lies in finding the likelihood distribution of the particle mass given the particular data sample, which is performed by evaluating the likelihood of measured observables given a generated particle mass for each event in the sample. The signal likelihood is evaluated from the theoretically expected Breit-Wigner distribution convoluted with the Gaussian experimental resolution on event-by-event basis. The whole procedure is calibrated using Monte Carlo simulation of a series of particle masses in the expected region.

One of the first applications of the ideogram method tracks back to the W boson mass measurement by the DELPHI Collaboration at CERN LEP collider [114, 115]. It was later used in a series of top quark mass measurements, including the DØ measurement in the lepton plus jets channel [110] and CDF measurement in the all-hadronic channel [116]. The main CMS top quark mass measurement in lepton plus jets channel at 7 TeV [102] which is cross-checked by this analysis is also based on the ideogram method with, however, *in situ* JES (jet energy scale) measurement in a joint likelihood fit. As JES is a dominant systematic in this analysis, this approach proved to be more precise, but let us not neglect the importance of a cross-check.

5.4.1 Event likelihood

In the basis of the ideogram method lies the event likelihood, also referred to as the ideogram. It is calculated for each event in the sample as a function of the hypothesised top quark mass m_{t} in the following way:

$$\mathcal{L}_{\text{event}}\left(\mathbf{x}|m_{\mathrm{t}}, f_{\mathrm{t\bar{t}}}\right) = f_{\mathrm{t\bar{t}}} P_{\mathrm{t\bar{t}}}\left(\mathbf{x}|m_{\mathrm{t}}\right) + \left(1 - f_{\mathrm{t\bar{t}}}\right) P_{\text{bkg}}\left(\mathbf{x}\right). \tag{5.5}$$

Here $\mathbf{x}$ is the set of event observables with the top mass information from the kinematic fit, $f_{\mathrm{t\bar{t}}}$ is the fraction of $\mathrm{t\bar{t}}$ events in the data sample, and $P_{\mathrm{t\bar{t}}}$ and P_{bkg} are probability densities for signal and background events, respectively. Essentially, by construction of the likelihood, the first term corresponds to the probability for the event to be signal, whilst the second term is its probability to be background. The kinematic information $\mathbf{x}$ includes the fitted mass m_i, the estimated standard deviation for each fitted mass $\sigma(m_i)$ and the goodness-of-fit χ_i^2 for all permutations, i.e. all possible jet-parton assignments and neutrino solutions surviving the selection criteria on the output of the fit.

Jet-parton assignment weights

The signal and background probabilities forming the event likelihood are calculated as a weighted sum over jet-parton assignments and neutrino solutions coming from the kinematic fit. Each permutation is weighted by the probability that it is correct, which is essentially the fit probability P_{fit} introduced in Equation 5.4. However, this particular implementation of the ideogram method also exploits the b-tagging information, thus the fit probability is multiplied by an additional term p_{btags}, which is the probability that the particular jet-parton assignment is compatible with the observed b-tags:

$$p_{\mathrm{btags}} = \prod_{j=1}^{4} p_j. \tag{5.6}$$

Here the index j runs over all four jets in the kinematic fit, and the probability p_j can be either ε_l, $(1-\varepsilon_l)$, ε_b, or $(1-\varepsilon_b)$, depending on the flavour of the jet as derived by the fit (light jet or b-jet), and whether the jet is b-tagged or not. As a recap from Section 3.5.3.3, for the medium working point of CSV b-tagging algorithm used in this analysis, b-tagging efficiency is approximately ~ 70 %, whilst the mis-tag rate (or $(1-\varepsilon_l)$) is approximately 1 %.

Therefore, for each such permutation the relative weight is w_i is calculated as:

$$w_i = p_{\mathrm{fit}} \cdot p_{\mathrm{btags}} = \exp(-\frac{1}{2}\chi^2) \cdot \prod_{j=1}^{4} p_j. \tag{5.7}$$

These weights are used in the calculation of signal and background probability densities, described hereafter.

Signal probability

The $\mathrm{t\bar{t}}$ signal probability in the event likelihood shown in Equation 5.5 is calculated as a weighted sum over all permutations:

$$\begin{aligned} P_{\mathrm{t\bar{t}}}(\mathbf{x}|m_{\mathrm{t}}) = \sum_{i}^{24} w_i \Bigg(f_{\mathrm{cp}} \cdot \int_{m_{\min}}^{m_{\max}} \mathrm{G}(m'|m_i,\sigma_i)\ \mathrm{RBW}(m'|m_{\mathrm{t}},\Gamma_{\mathrm{t}})\mathrm{d}m' + \\ +(1-f_{\mathrm{cp}})\mathrm{WP}(m_i|m_{\mathrm{t}}) \Bigg). \end{aligned} \tag{5.8}$$

Here f_{cp} is the fraction of events in which the maximum weight is assigned to the correct permutation, estimated from simulated $\mathrm{t\bar{t}}$ events. For each permutation, the first term expresses the probability that it has a correct jet-parton assignment (correct permutation). It is calculated as a convolution of a relativistic Breit-Wigner distribution $\mathrm{RBW}(m'|m_\mathrm{t}, \Gamma_\mathrm{t})$ and a Gaussian $\mathrm{G}(m'|m_i, \sigma_i)$.

The relativistic Breit-Wigner distribution describes the theoretical top quark mass distribution for a given hypothesised top quark mass value m_t with a top quark width Γ_t set to 2 GeV, according to the latest experimental results [117]. As suggested by Standard Model electroweak corrections and $s-$dependence of the phase space [118], the following form of RBW is used:

$$\mathrm{RBW}(m'|m_\mathrm{t}, \Gamma_\mathrm{t}) \propto \frac{m'^2}{\left(m'^2 - {m_\mathrm{t}}^2\right)^2 + m'^4 \frac{\Gamma_\mathrm{t}}{{m_\mathrm{t}}^2}}. \tag{5.9}$$

The Gaussian in the convolution represents the detector resolution. It is centred at the fitted mass value m_i with a width of the fitted mass uncertainty σ_i as estimated by the fit:

$$\mathrm{G}(m'|m_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(\frac{(m' - m_i)^2}{2\sigma_i^2}\right). \tag{5.10}$$

The fitted mass spectrum for $\mathrm{t\bar{t}}$ events with correct permutations is shown in Figure 5.4. The Monte Carlo sample with top mass of 166.5 GeV was used, for illustration.

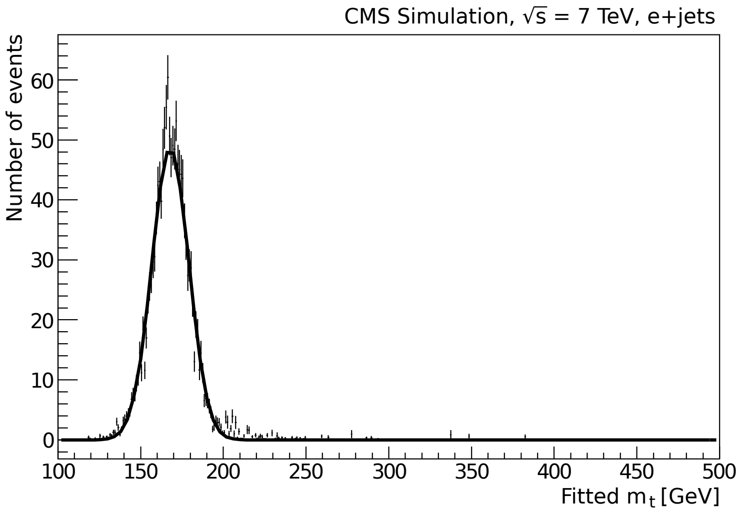


Figure 5.4: Fitted top quark mass distribution for $\mathrm{t\bar{t}}$ events with correct permutations. The points with error bars represent the numbers of events for values of fitted m_t from the output of the kinematic fit, obtained with MC truth sample generated with $m_\mathrm{t} = 166.5$ GeV.

The second term shows the probability of a wrong permutation in the signal event. This probability density, denoted $\mathrm{WP}(m_i|m_\mathrm{t})$, is calculated by fitting the Crystal Ball analytical function to the fitted top quark mass distribution for permutations with wrong jet-parton assignment, estimated purely from Monte Carlo

$t\bar{t}$ sample whilst taking the permutation weights (Equation 5.7) into account. The wrong permutations fall in two categories: the first one corresponds to events where the four leading jets come from $t\bar{t}$ decay, but are not assigned to the partons correctly; the second one are events with one or more of the four leading jets coming from soft gluon radiation (ISR or FSR). It was found that the fitted top quark mass distributions have similar shapes for both categories, therefore they were combined for the purpose of fitting. The Crystal Ball function [119, 120] has the form of:

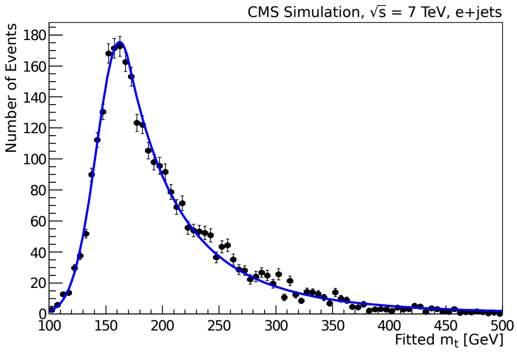
$$\mathrm{CB}(x;\mu,\sigma,\alpha,n) = N \times \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) & \text{if } \frac{x-\mu}{\sigma} > -\alpha, \\ \frac{1}{\sqrt{2\pi}\sigma}(\frac{n}{|\alpha|})^n \exp(-\frac{|\alpha|^2}{2})(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\mu}{\sigma})^{-n} & \text{if } \frac{x-\mu}{\sigma} \leq -\alpha. \end{cases} \tag{5.11}$$

The exponential order of $n = 5$ was used, and parameters μ, σ, and α were parametrised as functions of the top quark mass. The fitted distributions of wrong permutations spectra for three different top masses (166.5 GeV, 172.5 GeV and 178.5 GeV) are shown in Figure 5.5.

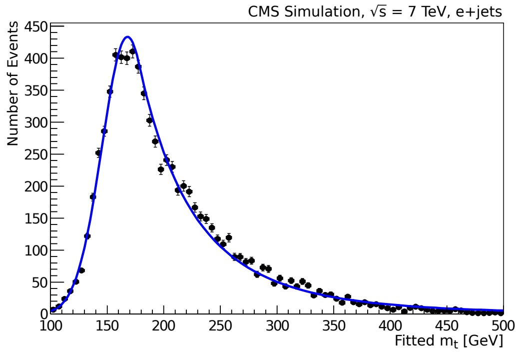
Background probability

The shape of the background probability density in Equation 5.5 is assumed not to depend on the top quark mass, and is estimated from the background Monte Carlo samples. The procedure is done in a similar way to the wrong permutations probability, and the jet-parton assignment weights (Equation 5.7) are also taken into account. Since the dominant background is $\mathrm{W}+\mathrm{jets}$, it was used exclusively for extracting the fitted top quark mass distribution by running the kinematic fit. The contribution from single top, $\mathrm{Z}+\mathrm{jets}$ and QCD was expected to be small, and it was confirmed that their probability density distributions have similar shapes to that of the $\mathrm{W}+\mathrm{jets}$ sample. Moreover, the Monte Carlo calibration described later on in Section 5.5 takes all mentioned background sources into account, therefore any possible bias due to this approximation is corrected for.

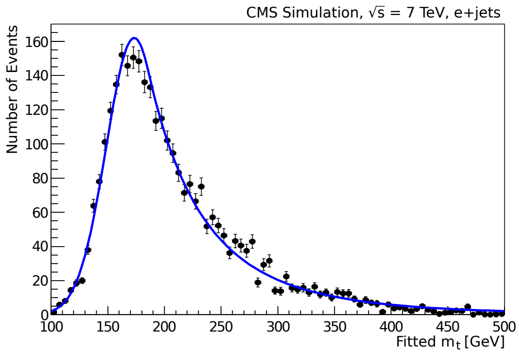
The fitted top quark mass spectrum for background events as estimated from the $\mathrm{W}+\mathrm{jets}$ MC sample is shown in Figure 5.6. This distribution is described by the Landau function, fitted by the ROOT package [121].



(a)



(b)



(c)

Figure 5.5: The fitted top quark mass distribution for $t\bar{t}$ events with wrong permutations. The points with error bars represent the numbers of events for values of fitted m_{t} from the output of the kinematic fit, obtained with MC truth sample generated with (a) $m_{\mathrm{t}} = 166.5$ GeV, (b) $m_{\mathrm{t}} = 172.5$ GeV, (c) $m_{\mathrm{t}} = 178.5$ GeV.

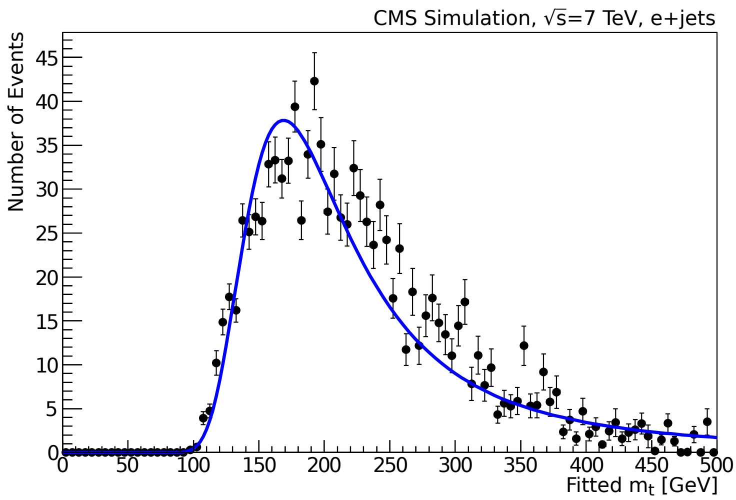


Figure 5.6: Fitted top quark mass distribution for background events. The points with error bars represent the numbers of events for values of fitted m_{t} from the output of the kinematic fit, obtained with a $\mathrm{W} + \mathrm{jets}$ Monte Carlo sample.

5.4.2 Combined likelihood fit

Finally, for the extraction of the top quark mass the total likelihood for the entire data sample is calculated as the product of all the event likelihoods:

$$\mathcal{L}_{\text{sample}} = \prod_{\text{all events}} \mathcal{L}_{\text{event}}\left(\mathbf{x}|m_{\text{t}}\right). \tag{5.12}$$

To extract the most probable top quark mass, this likelihood needs to be maximised. In practice, it is simplified by performing the minimisation of the summed negative log-likelihoods:

$$-2\log\mathcal{L}_{\text{sample}} = -2\sum_{\text{all events}} \log\mathcal{L}_{\text{event}}\left(\mathbf{x}|m_{\text{t}}\right). \tag{5.13}$$

The result with the full dataset is shown in Figure 5.7. The top quark mass and its statistical uncertainty are calculated by using the parabolic interpolation with three lowest points in the log-likelihood curve. However, this result can not be regarded as the final measurement, as it does not take into account the possible mass bias of the ideogram method. The mass calibration technique is described in the next section.

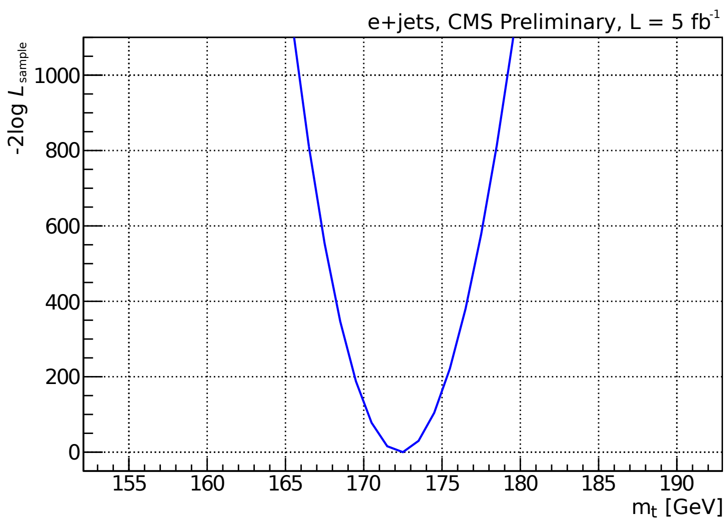


Figure 5.7: Log-likelihood curve as a function of top quark mass for electron plus jets dataset before calibration (the true minimum is subtracted in Equation 5.13).

5.5 Mass Calibration

By construction of the ideogram method, it has approximations which can lead to possible bias in the top quark mass measurement. To account for this bias, a calibration is performed by running the measurement on Monte Carlo simulation with known true top quark mass. Pseudo-experiments are used to measure the bias in the fitted mass, as well as to calibrate the statistical uncertainty of the measurement.

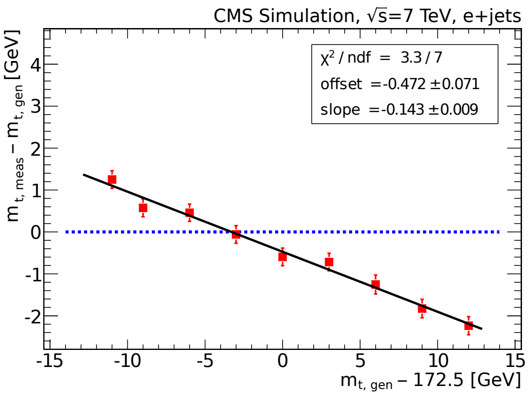
For each of the nine generated $\mathrm{t\bar{t}}$ samples with mass points between 161.5 GeV and 184.5 GeV, 5000 pseudo-experiments were performed with events picked from Monte Carlo samples representing different sources of events, so that the numbers of events follow a Poisson distribution centred around the corresponding numbers observed in real data. The observed value of 23 727 electron plus jets events with an average signal fraction of 76 % was used in each pseudo-experiment. The background was estimated by using events from the W + jets MC sample, whereas the impact of other background sources was accounted for separately as a systematic uncertainty (Section 5.6). For every pseudo-experiment all the event likelihoods are calculated, the total log-likelihood is formed and the top quark mass (m_i) with its statistical uncertainty (σ_i) are extracted in the same way as in the real measurement described above. These results are used to derive the bias and pull distributions, defined as follows:

$$\mathrm{bias} = m_{\mathrm{t}}^{\mathrm{meas}} - m_{\mathrm{t}}^{\mathrm{gen}} \quad \text{and} \quad \mathrm{pull}_i = \frac{m_i - m_{\mathrm{t}}^{\mathrm{meas}}}{\sigma_i}, \tag{5.14}$$

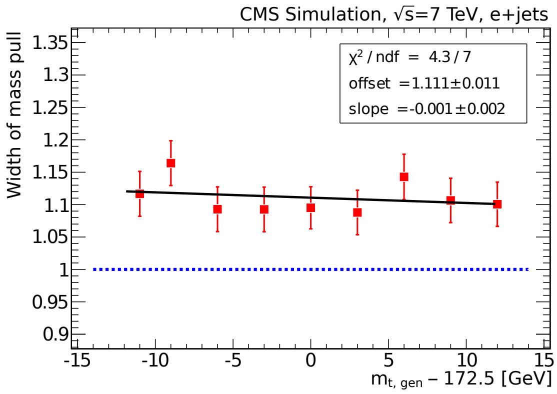
where $m_{\mathrm{t}}^{\mathrm{meas}}$ is the mean top quark mass over all pseudo-experiments, for a given generated mass point. The width of the pull is defined as the standard deviation of a Gaussian fitted to the pull distribution.

The mass bias and the width of the pull as a function of the generated top quark mass are shown in Figure 5.8 (a, b). Clearly, using the ideogram method out of the box reveals a significant bias which needs to be corrected for. It can also be seen that the width of the pull distribution is approximately 11 % larger than unity, suggesting that the ideogram method underestimates the statistical uncertainty.

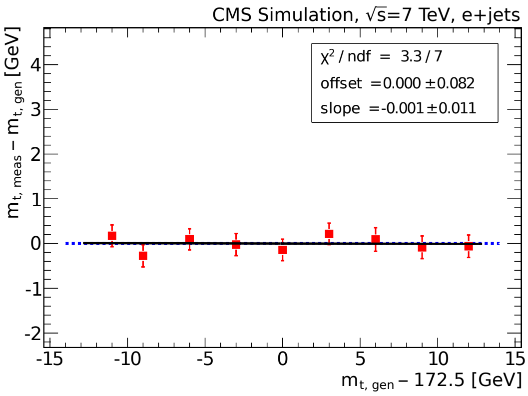
Due to the apparent bias, the fitted linear calibration curves in both distributions are used to correct the final measurement of the top quark mass. As a closure test, the mass bias and pull width as a function of true top quark mass are calculated after applying the full calibration. The distributions are shown in Figure 5.8 (c, d). Evidently, the closure tests confirm that the top quark mass and its statistical uncertainty are unbiased after calibration.



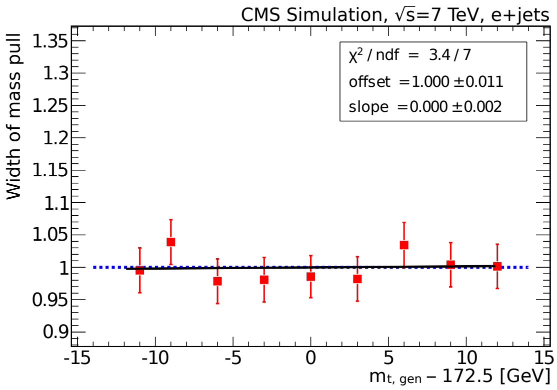
(a)



(b)



(c)



(d)

Figure 5.8: Average mass bias and width of the pull as a function of the true top quark mass before (a, b) and after (c, d) calibration.

5.6 Systematic Uncertainties

Like any measurement, the top quark mass measurement is prone to various systematic errors. In an attempt to evaluate them, several sources of systematic effects are considered in this analysis. These include theoretical uncertainties on Monte Carlo modelling, assumptions made within the ideogram method, as well as imperfect understanding of detector effects and reconstruction algorithms.

Systematic uncertainties are calculated by running the analysis with tuned MC samples where systematic effects are varied by ± 1 standard deviation, and comparing the result with the nominal one. The obtained difference in fitted Δm_{t} is regarded as an uncertainty due to a given systematic effect. The total systematic error is calculated as a sum of all components in quadrature, under the assumption that all the effects are uncorrelated. A summary of all systematic uncertainties is given in Table 5.6, whereas all corresponding sources are described below.

Table 5.6: Overview of the systematic uncertainties and their effects on the top quark mass measurement. Dominating uncertainties are emphasised in bold.

Source of the systematic uncertainty	Δm_{t} (GeV)
Fit calibration statistics	0.12
b-tagging	0.10
Jet Energy Scale (overall data/MC)	**1.21**
Factorisation scale	**1.38**
ME-PS matching threshold	**0.93**
Colour Reconnection	0.03
Underlying event	0.08
Non-$\mathrm{t\bar{t}}$ background	0.61
Pile-up	0.22
PDF	0.15
Total	2.17

Fit calibration statistics. To estimate the systematic effect of the calibration with pseudo-experiments described in Section 5.5, the statistical uncertainty of this calibration is propagated to the final measurement by regarding it as a systematic error. This uncertainty was evaluated with the central $\mathrm{t\bar{t}}$ MC sample with the top quark mass of 172.5 GeV.

b-tagging. The nominal ("medium") working point of the CSVM algorithm (Section 3.5.3.3) is varied in order to reflect the uncertainty of the b-tagging efficiency of ~ 4 % [94]. The average effect of this change on the mass measurement (approximately 0.1 GeV) is quoted as a systematic uncertainty.

Jet energy scale. Since the top quark mass measurement is based on knowledge about at least four energetic jets in semileptonic $\mathrm{t\bar{t}}$ decay, it is very sensitive to any uncertainties in jet energies. Hence, the jet energy scale is a dominant systematic in this analysis, particularly depending on our understanding of the CMS detector. To quantify this systematic effect, the jet energies are varied within $\pm 1\sigma$ uncertainty of the overall jet energy correction, obtained from a dedicated jet energy calibration and resolution study [122]. This p_{T}- and η-dependent uncertainty incorporates various contributions, including an offset due to noise and pile-up, absolute and relative discrepancies between data and MC, flavour-dependent corrections, etc. After applying asymmetric $\pm 1\sigma$ uncertainties, an average systematic effect of 1.21 GeV on the fitted top mass was observed.

Factorisation scale. As shown in Table 5.4, the factorisation scale is varied by factors of 0.5 and 2 with respect to the nominal value used in the central generated samples. The procedure is performed for $\mathrm{t\bar{t}}+$jets, W$+$jets and Z$+$jets Monte Carlo samples independently as the factorisation scales are considered uncorrelated. To combine the systematic uncertainty, the average absolute size of the up and down effects are added in quadrature. Due to the lack of statistics available in variations of W $+$ jets and Z $+$ jets samples, this systematic uncertainty is one of the largest in the top quark mass measurement.

ME-PS matching threshold. The systematic effect of the choice of ME-PS threshold, mentioned in Section 3.4.1, is estimated by varying the default value of 20 GeV by factors of 0.5 and 2. Similarly to factorisation scale systematic uncertainty, the matching thresholds are varied independently for $\mathrm{t\bar{t}}$ $+$ jets, W $+$ jets and Z $+$ jets Monte Carlo samples and then added in quadrature. The same issue with the lack of statistics appears here, although to a lesser extent.

Colour reconnection. Different modelling of colour reconnection (Section 3.4.1) result in a systematic uncertainty that is estimated by changing the central MC tune by the one with disabled colour reconnections. The observed shift in top quark mass is quoted as a systematic uncertainty.

Underlying event. The underlying event (Section 3.4.1) modelling uncertainty was estimated by using different PYTHIA tunes with increased and decreased underlying event activity. The central tune, referred to as Perugia 2011, was compared to Perugia 2011 mpiHi and Perugia 2011 Tevatron tunes [123] with more and less underlying event activity, respectively. The largest mass shift (≈ 0.08 GeV) represents the systematic uncertainty due to underlying event modelling.

Background modelling. As mentioned in the description of mass calibration (Section 5.5), the background in pseudo-experiments was estimated by using only W + jets MC sample. The impact of other background sources was studied separately and accounted for as a systematic uncertainty. Implementing the Z + jets, single top and QCD backgrounds in the calibration and separately varying their contributions by 30 % yielded the shifts in the measurement, the quadrature sum of which (≈ 0.61 GeV) was quoted as a systematic uncertainty due to the background composition.

Pile-up. Since the number of pile-up vertices in data can only be known after data-taking process, it does not usually match the according number in MC simulation. To account for this discrepancy, the simulated events are re-weighted to match the observed pile-up distribution in data. The systematic uncertainty due to pile-up re-weighting was estimated by scaling the pile-up weights up and down, leading to shifts in top quark mass measurement. The largest of these shifts (approximately 0.22 GeV) was quoted as a systematic uncertainty.

Parton distribution functions. The CTEQ parton distribution functions [105] used in the simulation of Monte Carlo samples can be described by 22 orthogonal parameters. To estimate the systematic uncertainty due to the choice of PDFs, the up and down variations of these parameters were employed to calculate the impact on the measurement. Events in nominal $\mathrm{t\bar{t}}$ + jets sample were re-weighted according to the deviation of each PDF parameter from the central value. The average absolute values of the up and down shifts for each variation are then summed in quadrature, yielding the value of 0.15 GeV which we quote as a PDF systematic uncertainty.

5.7 Results

After extracting the top quark mass using combined likelihood fit, performing mass calibration and estimating systematic uncertainties as shown in previous sections, the following result was obtained in the electron plus jets channel:

$$m_{\mathrm{t}} = 172.87 \pm 0.27\ (\text{stat.}) \pm 2.17\ (\text{syst.})\ \text{GeV}. \tag{5.15}$$

Being a cross-check result, this measurement is consistent with the official CMS top quark mass measurement at 7 TeV in the lepton plus jets channel [102]:

$$m_{\mathrm{t}} = 173.49 \pm 0.43\ (\text{stat.+JES}) \pm 0.98\ (\text{syst.})\ \text{GeV}. \tag{5.16}$$

This result expectedly has a significantly lower uncertainty due to *in situ* measurement of the jet energy scale (JES) in a joint likelihood fit. It was obtained using the combined fit in the electrons plus jets and muon plus jets channels, whereas separate fits in each channel yield following results:

$$\begin{aligned} \text{e+jets: } m_{\mathrm{t}} &= 173.72 \pm 0.66\ (\text{stat.+JES}) \pm 1.00\ (\text{syst.})\ \text{GeV}, \\ \mu\text{+jets: } m_{\mathrm{t}} &= 173.22 \pm 0.56\ (\text{stat.+JES}) \pm 1.06\ (\text{syst.})\ \text{GeV}. \end{aligned} \tag{5.17}$$

All results are consistent with the latest Tevatron and LHC combination result [11]:

$$m_{\mathrm{t}} = 173.34 \pm 0.27\ (\text{stat.}) \pm 0.71\ (\text{syst.})\ \text{GeV}.$$

5.8 Summary

In this chapter, a top quark mass measurement in the electron plus jets channel was presented using the full 2011 dataset recorded by the CMS detector with a total integrated luminosity of 5.0 fb^{-1}. The mass extraction technique called ideogram method was described in detail, as well other technicalities of the analysis. The result is consistent with the official CMS top quark mass measurement in the lepton plus jets channel at 7 TeV, as well as latest combination results from the Tevatron and the LHC experiments.

6. Top Quark Pair Differential Cross Section Measurement

Measurement of the differential cross section of the top quark pairs with respect to different variables is an important precision measurement, and provides a sensitive probe for new physics. For instance, deviations in the tail of the missing transverse energy distribution could signal new resonances with invisible particles produced in association with top quarks.

The analysis described in this chapter represents the top quark differential cross section measurement with respect to event-level (or global) distributions, including the missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$), the scalar sum of jet transverse momenta (H_{T}), the scalar sum of the transverse momenta of all objects in the event (S_{T}), and both the transverse mass ($M_{\mathrm{T}}^{\mathrm{W}}$) and transverse momentum ($p_{\mathrm{T}}^{\mathrm{W}}$) of the leptonically decaying W boson produced in the top quark decay [124, 125]. The analysis is focused on semileptonic $\mathrm{t\bar{t}}$ decays, where a lepton is either an electron or a muon. The cross section measurement is performed to validate different Monte Carlo generators and theoretical predictions of the effects due to variations of various modelling parameters.

6.1 Data and Simulation

6.1.1 Data

This analysis uses the full 2012 dataset collected by the CMS detector at a centre of mass energy of 8 TeV, with a total integrated luminosity of 19.7 fb^{-1}. Only certified events were used in the data, i.e. from such preriods of data-taking when all of the detector subsystems were functioning with no errors. Depending on the channel, the data were preselected with the single electron or single muon trigger. The preselection procedure as well as full event selection will be described in Section 6.2.

6.1.2 Monte Carlo samples

The Monte Carlo generators used in this work were presented in Section 3.4.2. The list of signal and background MC samples is shown in Table 6.1, and is broadly similar to that from the top quark mass analysis. In addition, the signal $\mathrm{t\bar{t}}$ + jets sample is also available with POWHEG and MC@NLO generators in order to be able to differentiate between them. Moreover, to extend the statistics of the W/Z + jets samples, they were generated in four exclusive jet multiplicity bins: W/Z boson plus one/two/three and at least four jets.

Table 6.2 presents the list of QCD multi-jet background and γ + jets samples used in the estimation of the QCD background in the electron plus jets channel. The analogous set of muon-enriched QCD samples used in the muon plus jets channel is shown in Table 6.3. Although the QCD background is estimated using data-driven techniques in both channels (see Section 6.3), the MC simulation is still used for normalisation purposes. Table 6.4 lists the samples used for estimation of the systematic uncertainties due to factorisation scale and matching threshold choices, as discussed in Section 3.4.1. Similarly to the top quark mass analysis, a summary of central values and variations used to estimate these systematic uncertainties is shown in Table 5.4.

Table 6.1: Signal and background Monte Carlo samples with cross sections at $\sqrt{s} =$ 8 TeV, numbers of generated events and corresponding integrated luminosities.

Process	Generator	σ (pb)	# events	$\int \mathcal{L}dt$ (fb^{-1})
$\mathrm{t\bar{t}}$ + jets, m_{t} = 172.5 GeV	MadGraph + PYTHIA	245.8	6854416	27.9
$\mathrm{t\bar{t}}$ + jets, m_{t} = 172.5 GeV	POWHEG + PYTHIA	245.8	21675970	88.2
$\mathrm{t\bar{t}}$ + jets, m_{t} = 172.5 GeV	MC@NLO + HERWIG	245.8	32706581	133.1
W + jets (W $\to l\nu$)	MadGraph + PYTHIA			
W + 1 jet		5400.0	23141598	4.3
W + 2 jet		1750.0	34044921	19.5
W + 3 jet		519.0	15539503	29.9
W + 4 jet		214.0	13349346	62.4
Z/$\gamma^* \to l^+l^-$ + jets, $m(ll) > 50$ GeV	MadGraph + PYTHIA			
Z + 1 jet		561.0	24045248	42.9
Z + 2 jet		181.0	21852156	120.7
Z + 3 jet		51.1	11015445	215.6
Z + 4 jet		23.04	6402827	277.9
Single top	POWHEG + PYTHIA			
top t-channel		55.5	3758227	67.7
anti-top t-channel		30.0	1935072	64.5
top s-channel		3.89	259961	66.8
anti-top s-channel		1.76	139974	79.5
top tW-channel		11.18	497658	44.5
anti-top tW-channel		11.18	493460	44.1

Table 6.2: QCD multi-jet background and γ + jets MC samples used in the electron plus jets channel with cross sections at $\sqrt{s} = 8$ TeV, numbers of generated events and corresponding integrated luminosities.

Process	Generator	σ (pb)	filter efficiency	# events	$\int \mathcal{L}dt$ (fb^{-1})
QCD (e/γ enriched)	PYTHIA				
20 GeV $< \hat{p}_{\mathrm{T}} <$ 30 GeV		2.886×10^{8}	1.01×10^{-2}	34339883	1.2×10^{-2}
30 GeV $< \hat{p}_{\mathrm{T}} <$ 80 GeV		7.433×10^{7}	6.21×10^{-2}	32537408	7.0×10^{-3}
80 GeV $< \hat{p}_{\mathrm{T}} <$ 170 GeV		1.191×10^{6}	0.154	34542763	0.19
170 GeV $< \hat{p}_{\mathrm{T}} <$ 250 GeV		30 990.0	0.148	22862259	5.0
250 GeV $< \hat{p}_{\mathrm{T}} <$ 350 GeV		4250.0	0.131	32505856	58.4
$\hat{p}_{\mathrm{T}} >$ 350 GeV		810.0	0.11	33981105	381.4
QCD (b/c $\to e\nu$)	PYTHIA				
20 GeV $< \hat{p}_{\mathrm{T}} <$ 30 GeV		2.886×10^{8}	5.8×10^{-4}	1740229	1.0×10^{-2}
30 GeV $< \hat{p}_{\mathrm{T}} <$ 80 GeV		7.433×10^{7}	2.25×10^{-3}	2048152	1.2×10^{-2}
80 GeV $< \hat{p}_{\mathrm{T}} <$ 170 GeV		1.191×10^{6}	1.09×10^{-2}	1945525	0.15
170 GeV $< \hat{p}_{\mathrm{T}} <$ 250 GeV		30 990.0	2.04×10^{-2}	1948112	3.1
250 GeV $< \hat{p}_{\mathrm{T}} <$ 350 GeV		4250.0	2.43×10^{-2}	2026521	19.6
$\hat{p}_{\mathrm{T}} >$ 350 GeV		810.0	2.95×10^{-2}	1948532	81.5
γ + jets	MadGraph				
200 GeV $< H_{\mathrm{T}} <$ 400 GeV	+PYTHIA	960.5	1	10479625	10.9
$H_{\mathrm{T}} >$ 400 GeV		107.5	1	1611963	15.0

Table 6.3: QCD multi-jet background MC samples used in the muon plus jets channel with cross sections at $\sqrt{s} = 8$ TeV, numbers of generated events and corresponding integrated luminosities.

Process	Generator	σ (pb)	filter efficiency	# events	$\int \mathcal{L}dt$ (fb^{-1})
QCD (μ enriched)	PYTHIA				
15 GeV $< \hat{p}_{\mathrm{T}} <$ 20 GeV		7.022×10^{8}	3.9×10^{-3}	1722681	6.3×10^{-4}
20 GeV $< \hat{p}_{\mathrm{T}} <$ 30 GeV		2.87×10^{8}	6.5×10^{-3}	8486904	4.5×10^{-3}
30 GeV $< \hat{p}_{\mathrm{T}} <$ 50 GeV		6.609×10^{7}	1.22×10^{-2}	9560265	1.2×10^{-2}
50 GeV $< \hat{p}_{\mathrm{T}} <$ 80 GeV		8.082×10^{6}	2.18×10^{-2}	10365230	5.9×10^{-2}
80 GeV $< \hat{p}_{\mathrm{T}} <$ 120 GeV		1.024×10^{6}	3.95×10^{-2}	9238642	0.23
120 GeV $< \hat{p}_{\mathrm{T}} <$ 170 GeV		1.578×10^{5}	4.73×10^{-2}	8501935	1.1
170 GeV $< \hat{p}_{\mathrm{T}} <$ 300 GeV		34 020.0	6.76×10^{-2}	7669947	3.3
300 GeV $< \hat{p}_{\mathrm{T}} <$ 470 GeV		1757.0	8.64×10^{-2}	7832261	51.6
470 GeV $< \hat{p}_{\mathrm{T}} <$ 600 GeV		115.2	0.102	3783069	322.0
600 GeV $< \hat{p}_{\mathrm{T}} <$ 800 GeV		27.01	0.0996	4119000	1531.1
800 GeV $< \hat{p}_{\mathrm{T}} <$ 1000 GeV		3.57	0.1033	4107853	11 139.0
$\hat{p}_{\mathrm{T}} >$ 1000 GeV		0.774	0.1097	3873970	45 625.6

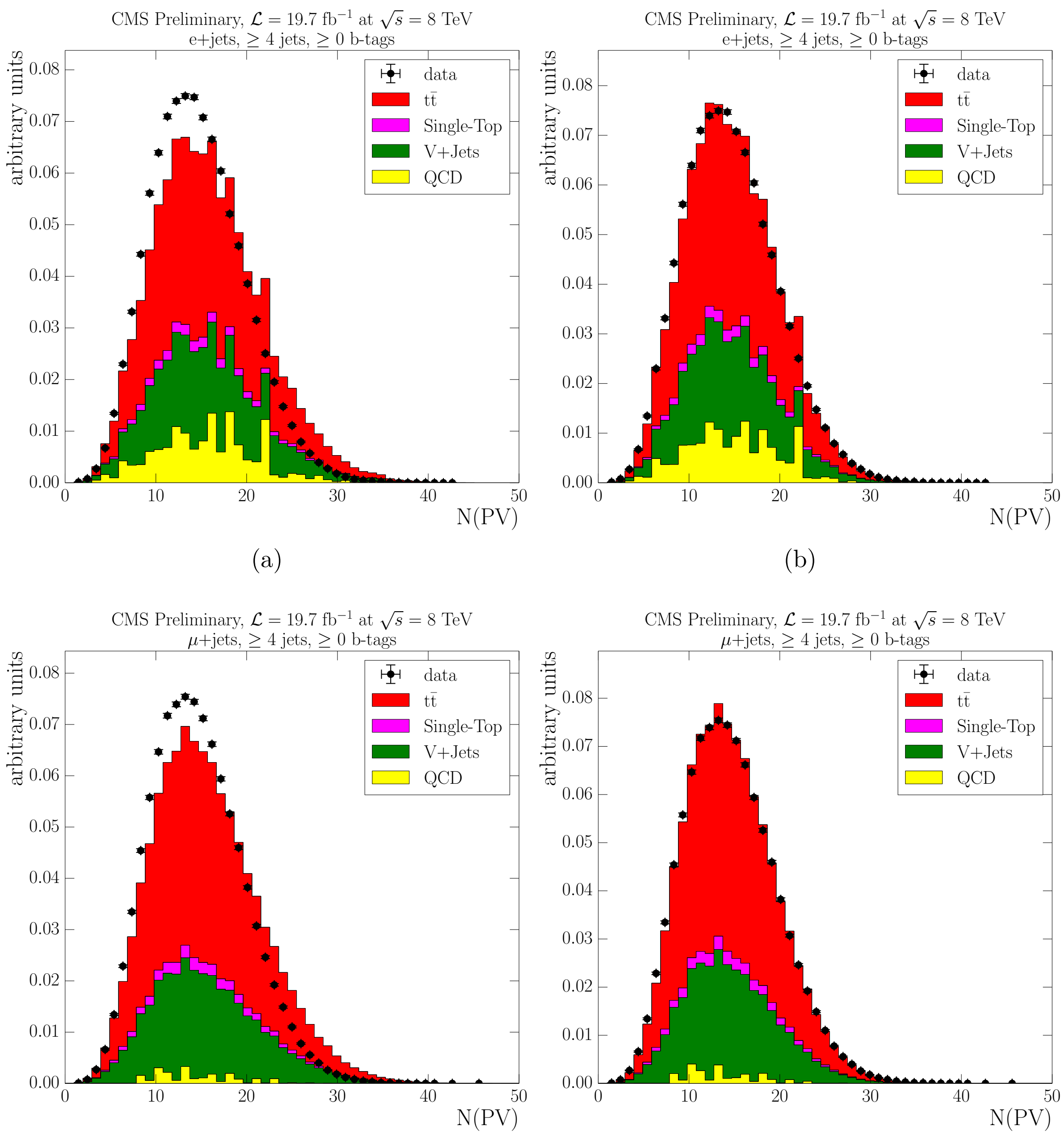
Table 6.4: Systematic MC samples with cross sections at $\sqrt{s} = 8$ TeV, numbers of generated events and corresponding integrated luminosities. Factorisation scale Q and matching threshold systematic uncertainties are estimated with variations of $\mathrm{t\bar{t}} + \mathrm{jets}$, $\mathrm{W} + \mathrm{jets}$ and $\mathrm{Z} + \mathrm{jets}$ samples.

Process	Generator	σ (pb)	# events	$\int \mathcal{L}dt$ (fb^{-1})
$\mathrm{t\bar{t}} + \mathrm{jets}$	MadGraph + PYTHIA			
0.5 × matching threshold		245.8	5476728	10.2
2 × matching threshold		245.8	5306710	25.5
$0.5 \times Q$		245.8	5387181	25.4
$2 \times Q$		245.8	5009488	23.4
$\mathrm{W} + \mathrm{jets}$ ($\mathrm{W} \to l\nu$)	MadGraph + PYTHIA			
0.5 × matching threshold		29690	21364637	0.7
2 × matching threshold		30290	20976082	0.7
$0.5 \times Q$		33300	20719363	0.6
$2 \times Q$		32000	20784770	0.6
$\mathrm{Z} + \mathrm{jets}$ ($\mathrm{Z} \to ll$)	MadGraph + PYTHIA			
0.5 × matching threshold		2888	2112387	0.6
2 × matching threshold		2915	1985529	0.7
$0.5 \times Q$		3312	1934901	0.6
$2 \times Q$		2954	2170270	0.7

6.1.3 Pile-up reweighting

Since the number of simulated pile-up interactions does not necessarily represent the same distribution in data, a reweighting procedure for MC events is required. To produce the weights, measured instantaneous luminosity is used to obtain a distribution of true pile-up vertices in data, i.e. before any vertex reconstruction inefficiencies. This distribution is then divided by the simulated one, and for each number of pile-up interactions, the weight is calculated. Figure 6.1 shows the number of interactions before and after reweighting for both electron and muon channels. Evidently, the reweighting procedure helps to achieve a good agreement between data and simulation.

To estimate the number of pile-up vertices in data, the measured instantaneous luminosity in each bunch crossing is multiplied by an average total inelastic proton-proton cross section. Therefore, two sources of uncertainty arise from these factors: the luminosity uncertainty, measured to be 4.4 % by a dedicated study [126], and the uncertainty on the total inelastic cross section. To obtain the total cross section value, the CMS measurement of 68.0 ± 4.5 mb based on 7 TeV data [127] was extrapolated to the 8 TeV value of 69.3 mb. The recommended uncertainty on this



(a) (b) (c) (d)

Figure 6.1: Number of reconstructed vertices per event before (left) and after pile-up reweighting (right) in the electron channel (top) and in the muon channel (bottom). Both data and sum of the MC samples are normalised to unit area.

value, set to be 5 %, covers the modelling and physics aspects of pile-up simulation that have not been fully studied to date.

The effect of the ± 5 % variations in total inelastic cross section on the estimate of the true number of vertices in data is shown in Figure 6.2, whereas its effect on the simulated distributions is shown in Figure 6.3 for both electron and muon channels.

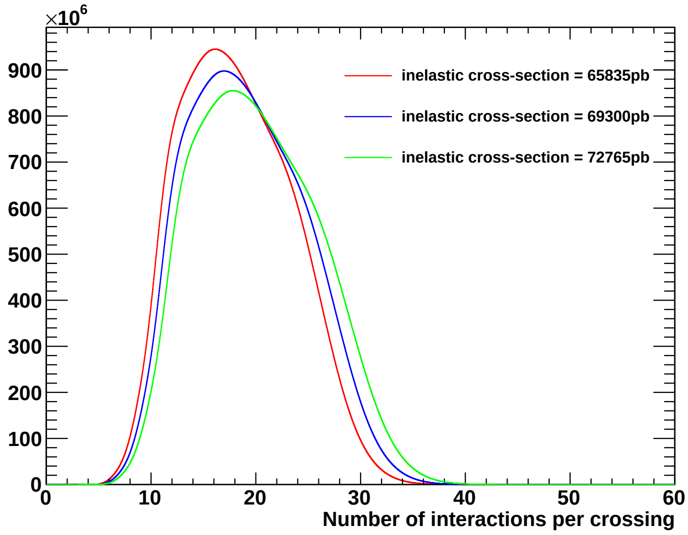


Figure 6.2: Number of expected vertices per event for central and $\pm 1\sigma$ variations of the total inelastic cross section for the 2012 data.

6.1.4 b-tagging corrections

In order to mitigate the background (QCD in particular), b-tagging is applied in the event selection of this analysis. Similarly to the top quark mass analysis, the CSV algorithm with the medium working point providing ~ 70 % efficiency and ~ 1 % mis-tag rate (Section 3.5.3.3) was used. However, in contrast with the 2011 analysis where the b-tagging corrections were implemented in the construction of the event likelihood, in this work scale factors are applied to Monte Carlo events. This is done in order to correct for the mismatch between the algorithm efficiency seen in data and MC [128]. p_{T} and η-dependent scale factors were derived by the CMS b-tag Physics Object Group [129] and are used to reweight the simulated events. The overall jet content of the event is taken into account, such as the number of jets originating from light and heavy quarks.

Figure 6.4 shows the results of the reweighting procedure for electron and muon channels. It can be seen that the numbers of simulated events in b-tag multiplicity bins of 0 and 1 are scaled up, whereas for other bins the number of MC events is scaled down. The systematic uncertainty due to b-tagging corrections is calculated by applying $\pm 1\sigma$ variations of the scale factors in the reweighting procedure.

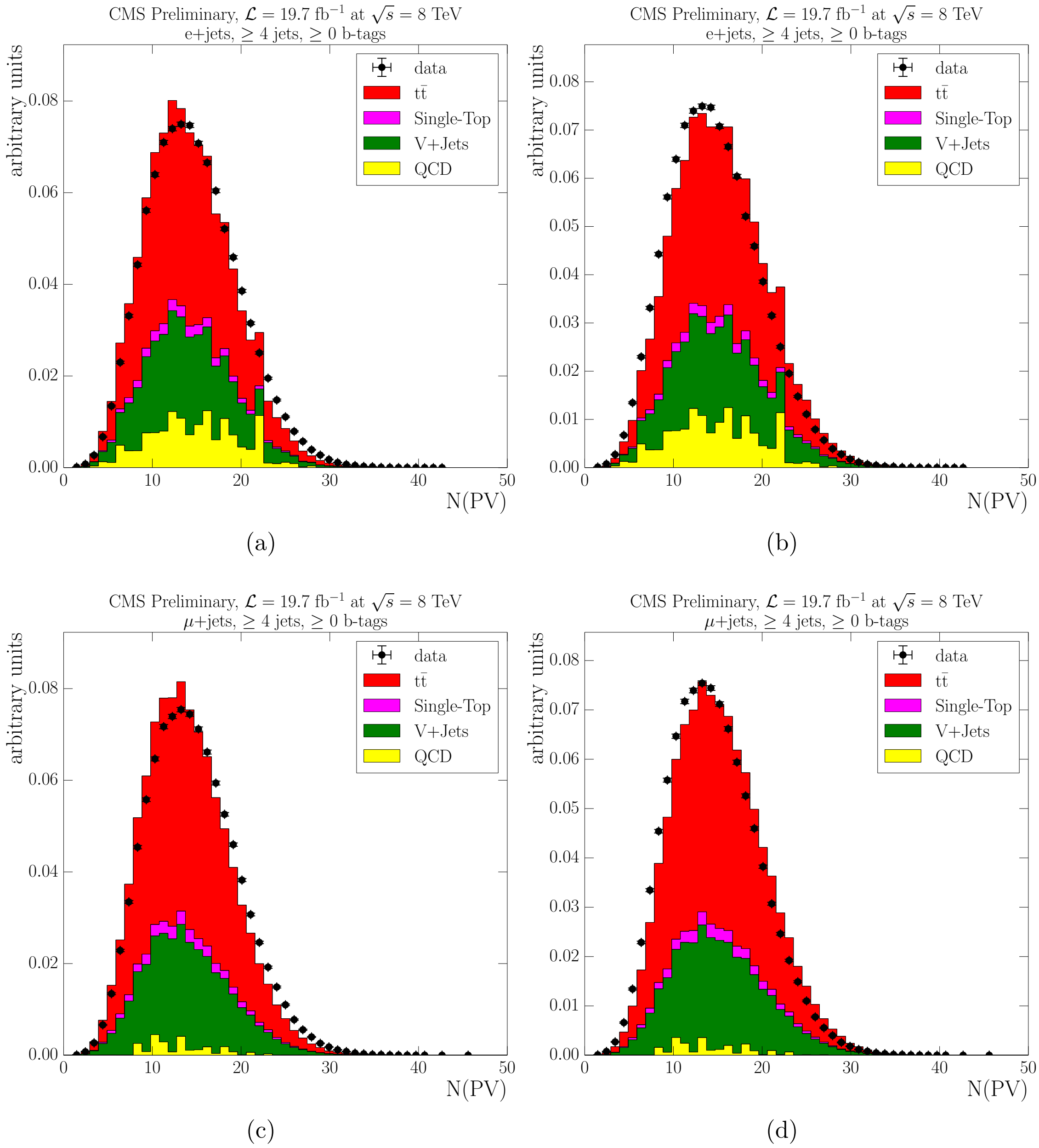


Figure 6.3: Number of reconstructed vertices per event for -1σ (left) and $+1\sigma$ variation (right) of the pile-up reweighting procedure for the electron channel (top) and the muon channel (bottom).

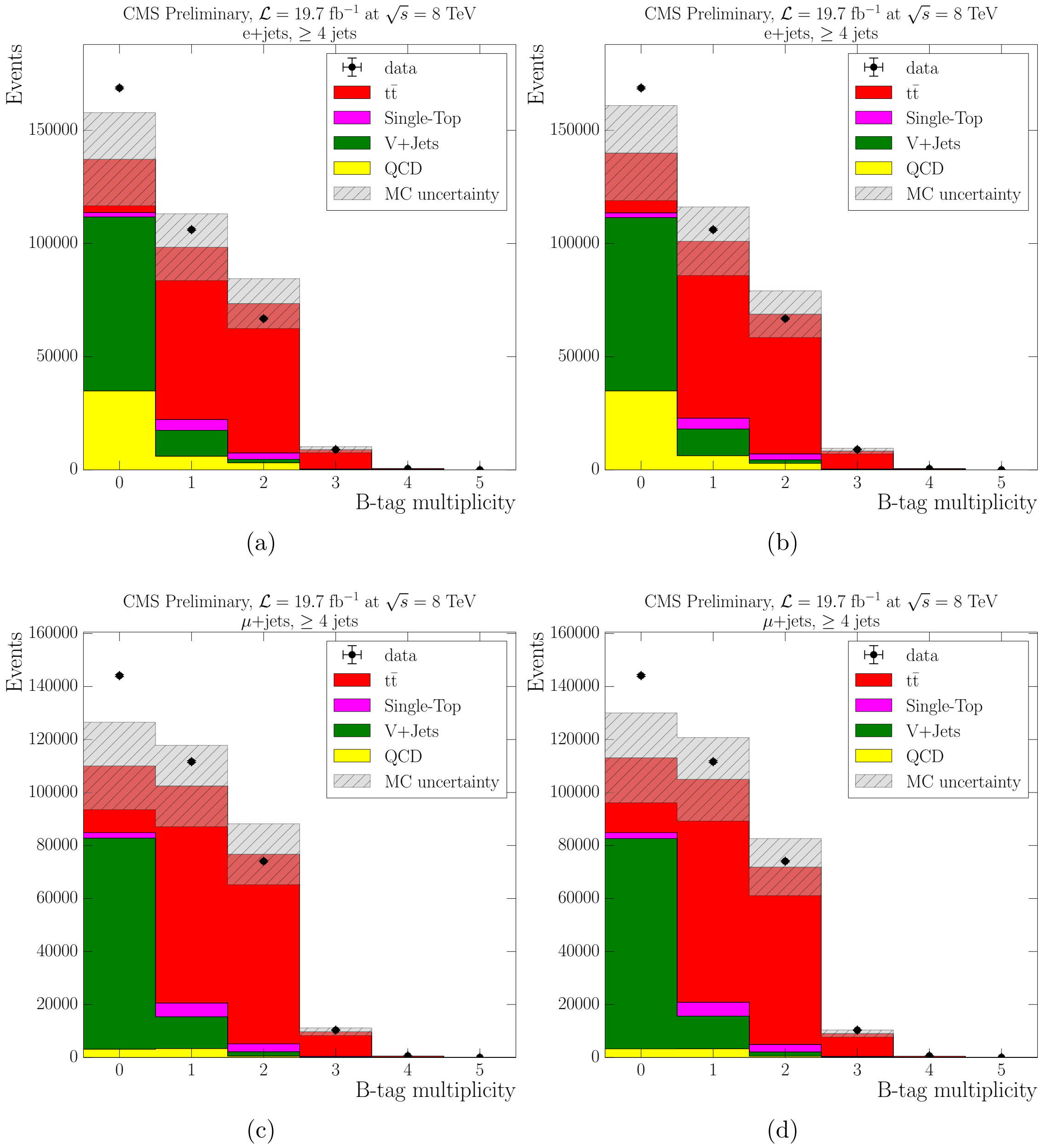


Figure 6.4: b-tag multiplicity before (left) and after applying the b-tagging scale factors (right) in the electron channel (top) and the muon channel (bottom).

6.2 Event Selection

As in the top quark mass analysis, the event selection in this work is based on the standard CMS selection optimised for Standard Model $\mathrm{t\bar{t}}$ production. It also exploits the semileptonic signature of a $\mathrm{t\bar{t}}$ decay with exactly one isolated lepton and at least four jets, two of which are b-tagged. One of the main goals of the analysis is the measurement of the missing transverse energy distribution represented by a neutrino from the leptonic decay of the W boson. Consequently, there is no specific requirement on the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the selection.

Comparing to the 2011 event selection described in Section 5.2, enhanced pile-up subtraction and lepton identification techniques are used. Improved understanding of the detector in the form of updated alignment and calibration constants resulted in better resolution of jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. The preselection (or skimming) step used in order to reduce the number of events for local analysis processing is identical to that of the 2011 analysis. However, additional filters are used to reject the events with artificial $E_{\mathrm{T}}^{\mathrm{miss}}$ caused by noise or known problems with one of the detector subsystems. These filters are discussed in Section 6.2.3.

Two semileptonic $\mathrm{t\bar{t}}$ decay modes are explored in this analysis: the electron plus jets and the muon plus jets channels. The event selection for each of these decay modes is described hereafter.

6.2.1 Electron plus jets channel

Similarly to the electron plus jets channel in the top quark mass analysis, the selection consists of the following steps:

1. preselection;
2. trigger;
3. electron candidate selection;
4. muon veto;
5. dilepton veto;
6. conversion veto;
7. jet selection;
8. b-tagging.

Trigger

In contrast to the 2011 analysis using cross triggers, a single electron trigger is used in this analysis. Referred to as HLT_Ele27_WP80, this trigger has a lepton p_{T} threshold of 27 GeV and "WP80" lepton isolation requirement (see Table 4.1), but no specific jet requirements. In 2012, the single electron trigger was unprescaled for the whole running period, despite a comparatively large rate. The decision to keep the trigger unprescaled was made due to simplicity of the trigger selection, requiring flat efficiency corrections.

Electron candidate selection

Similarly to 2011 selection, exactly one electron candidate passing is required in the event, passing the following criteria:

- $E_{\mathrm{T}} > 30$ GeV;
- $|\eta| < 2.5$, excluding the ECAL barrel-endcap transition regions of $1.4442 < |\eta| < 1.566$;
- MVA electron ID (Section 3.5.1) with a discriminator cut of > 0.5;
- transverse impact parameter with respect to the primary vertex $d_{xy} < 0.02$ cm;
- PF-based relative isolation (Section 3.5.1.2) reliso < 0.1 with a ΔR cone of 0.3.

Muon veto

To reduce contamination from other $\mathrm{t\bar{t}}$ decay modes (i.e. muon and dilepton channels), events containing a global or a tracker muon (see Section 3.5.2) are rejected. The requirements on the muon are identical to those of 2011 selection: $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.5$ and reliso < 0.2 with a ΔR cone of 0.4.

Dilepton veto

To reject events with additional electrons, a dilepton veto is applied. Events are rejected if they contain a second electron candidate with the same ID and η requirements, but looser E_{T} and reliso criteria (> 20 GeV and < 0.15, respectively).

Conversion veto

As described in Section 3.5.1.3, a vertex fit technique combined with the number of missing hits in the tracker is used for photon conversion rejection.

Jet selection and b-tagging

To help further reduce the background, at least four jets with $p_T > 30$ GeV and $|\eta| < 2.5$ are required in the event. These jets are required to pass the loose PF jet ID (see Section 3.5.3). Before the multiplicity and identification requirements, the collection of jets is cleaned against the signal electron, i.e. any jet within a ΔR cone of 0.3 around the signal lepton is excluded from the analysis. This is done due to the fact that an electron may be reconstructed as a jet by leaving a characteristic signature in calorimeters. Any additional jets with $p_T > 20$ GeV and the same cleaning and ID requirements are also used for calculation of H_T and S_T variables (Section 6.5.1). Finally, the combined secondary vertex (CSV) b-tagging algorithm with the medium working point (Section 3.5.3.3) is used to identify the jets originating from b-quarks in the $t\bar{t}$ decay. At least two b-tagged jets are required in this analysis.

6.2.2 Muon plus jets channel

The muon channel event selection follows a strategy similar to that of the electron channel:

1. preselection;
2. trigger;
3. muon candidate selection;
4. second muon veto;
5. electron veto;
6. jet selection;
7. b-tagging.

Trigger

An isolated single muon trigger, referred to as HLT_IsoMu24_eta2p1, is used in this channel. Partially suggested by the name of the trigger path, the following criteria on the muon are imposed at the HLT level:

- $p_T > 24$ GeV;
- $|\eta| < 2.1$;
- Combined tracker and calorimeter-based relative isolation (Section 3.5.1.2) reliso < 0.15 with a ΔR cone of 0.3.

Muon candidate selection

For selection of a muon candidate, exactly one muon satisfying the following requirements is required:

- $p_{\mathrm{T}} > 26$ GeV;
- $|\eta| < 2.1$;
- PF-based relative isolation (Section 3.5.1.2) reliso < 0.12 with a ΔR cone of 0.4;
- PF-muon ID criteria (Section 3.5.2):
  - identified as a particle flow muon and as a global muon;
  - a normalised χ^2 of the global muon fit < 10;
  - number of muon chamber hits > 0;
  - number of muon stations with muon segments > 1;
  - transverse impact parameter w.r.t. the primary vertex $d_{xy} < 0.02$ cm;
  - longitudinal distance of the tracker track w.r.t. the primary vertex $d_z < 0.5$ cm;
  - number of hits in the pixel detector > 0;
  - number of hits in the tracker layers > 5.

Second muon veto

Events containing an additional loose muon are rejected, with loose ID requirements as follows:

- $p_{\mathrm{T}} > 10$ GeV;
- $|\eta| < 2.5$;
- PF-based relative isolation (Section 3.5.1.2) reliso < 0.2 with a ΔR cone of 0.4;
- identified as a particle flow muon, and either global or tracker muon.

These criteria are identical to the muon veto requirements in the electron channel.

Electron veto

To further reduce background contamination, events with electrons are rejected. The requirements on additional loose electron are the same to the dilepton veto in the electron channel.

Jet selection and b-tagging

Finally, the jet requirements are applied. These are also identical to those of the electron channel, mentioned above.

6.2.3 Additional filters for missing transverse energy

There is no specific requirement on missing transverse energy in the selection, however, a set of corrections is applied to $E_{\mathrm{T}}^{\mathrm{miss}}$. As described in Section 3.5.4, jet energy corrections are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$, and it is also corrected for pile-up and various detector effects causing ϕ modulation.

Additionally, a set of optional filters, prescribed for analyses with high sensitivity to $E_{\mathrm{T}}^{\mathrm{miss}}$, was applied in the selection for both electron and muon channels. Most of these filters were put in place after discovering events with anomalously high missing transverse energy. A short description of each filter is given below.

CSC beam halo filter. Protons in the LHC beams can interact with residual particles of beam collimators, which produces secondary particle showers referred to as beam halo [130]. This machine-induced background can significantly contaminate collision events, particularly affecting the $E_{\mathrm{T}}^{\mathrm{miss}}$ observable due to specific trajectories of halo muons. A dedicated algorithm based on exploiting the halo signatures in the Cathode Strip Chambers (CSCs) helps to mitigate this background.

HCAL laser filter. The HCAL laser system is used for calibration and monitoring the detector response [60]. During the 2012 data-taking, laser pulses were accidentally fired and consequently polluted a small fraction of the recorded physics dataset. Using the fact that unlike physics events, laser pulses produce hits in calibration channels of the HCAL, the dataset is filtered from this contamination.

ECAL dead cell filter. Both ECAL endcap and barrel regions are known to have faulty crystals producing too much noise in detector readouts. There are also crystals corresponding to front-end cards with defective data link. All these channels are masked in reconstruction and constitute to about 1 % of the total amount. However, they can still cause a significant amount of energy deposits to be lost, therefore contributing to fake $E_{\mathrm{T}}^{\mathrm{miss}}$. One of the approaches to tackle this issue uses the so-called trigger primitive information, i.e. digital quantities produced by the Level 1 trigger electronics [95]. Another approach uses the energy deposits surrounding the masked channels (boundary energy). These methods allow to filter events with the estimated energy leakage above certain threshold (~ 10 GeV).

Tracking failure filter. Some events have been observed to have a lack of tracks corresponding to standard or large calorimeter deposits. There are two understood sources of this misreconstruction: in a first type of such events, the tracking algorithm (explained in Section 3.5.1) halts for some of its iterations due to an exceeding number of calorimeter clusters. In a second type the hard collision happens in a satellite colliding bunch at a distance of approximately 75 cm from the nominal interaction point along the z axis. A cut on the summed p_{T} of the tracks originating from the primary vertices passing the preselection criteria, divided by the summed p_{T} of all jets in the event, was found to clearly distinguish between pathological and unaffected events. A cut value of 10 % is imposed.

Bad ECAL endcap super-cluster filter. In 2012, two 5×5 super-clusters in the ECAL endcap regions were observed to occasionally produce anomalous pulses, leading to artificially high $E_{\mathrm{T}}^{\mathrm{miss}}$. The source of this issue can be related to the High-Voltage system, but is subject to further investigation. A filter removing the problematic events is used, configured to cut on the total energy of reconstructed hits not passing the nominal quality criteria in the two super-clusters, with the cut value of 1 TeV.

ECAL laser correction filter. Lead tungstate crystals of the ECAL naturally lose their transparency with radiation exposure. To maintain high precision, this effect is corrected for by frequently injecting laser pulses and reading out the response in order to calibrate each crystal [60]. During 2012 data-taking, a handful of crystals had unphysically large values of laser corrections, resulting in anomalously high $E_{\mathrm{T}}^{\mathrm{miss}}$ in recorded events. The filter removes such events from the dataset.

Strip tracker noise filter. A few events were affected by a large coherent noise in the silicon strip tracker, which resulted in a much larger number of the strip clusters compared to according pixel cluster multiplicity in the tracking algorithm. A dedicated filter rejects events with this anomaly.

Although the number of events removed by these filters is extremely small ($\ll 1$ %), most of them have high fake $E_{\mathrm{T}}^{\mathrm{miss}}$, which could lead to an artificial excess in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. Due to the analysis sensitivity to $E_{\mathrm{T}}^{\mathrm{miss}}$, application of these filters was of high importance.

6.3 Data-driven QCD Estimation

QCD multi-jet events constitute an important background to semileptonic $t\bar{t}$ decays, and one of the most difficult to model (Section 2.2.3). The requirement of two b-tagged jets in the event selection is mainly motivated by the necessity to mitigate the effect of this particular background. In this analysis, the final QCD yield is obtained from the fitting process as explained in Section 6.5. However, the initial estimate is performed using data-driven techniques. Only the shape of the distribution is of importance here as the normalisation is determined in the fitting process. The QCD estimation methods for both electron and muon channel are described in detail hereafter.

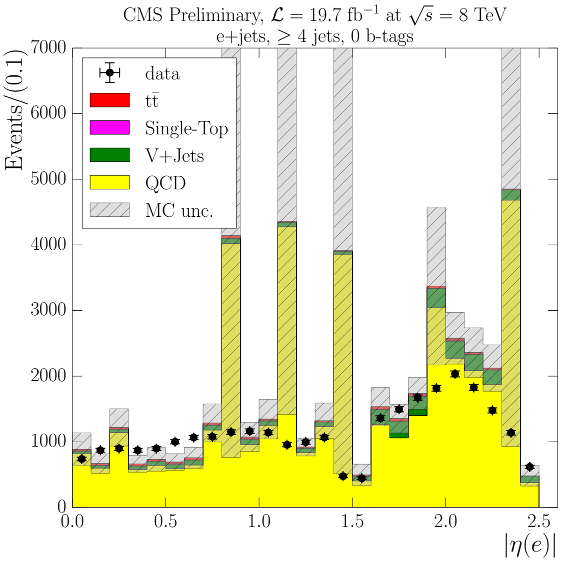
6.3.1 QCD multi-jet events with electrons

Two control regions have been investigated in the electron channel:

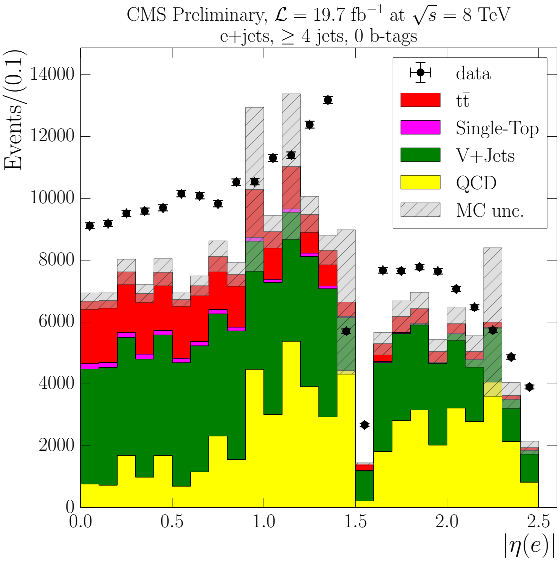
- conversion control region: conversion veto is inverted in the full event selection;
- non-isolated electron control region: the electron isolation requirement is changed from reliso < 0.1 to reliso > 0.2.

Additionally, only events with no b-tagged jets are taken into account to increase the statistics and QCD purity in both control regions. In Figure 6.5 the distributions of the absolute pseudorapidities of the electron are presented for both control regions. The distributions observed in data are compared with MC expectations. It can be seen that the distributions are not well described by the Monte Carlo, e.g. occasional high peaks in the QCD MC represent the lack of statistics in simulation. Moreover, the two control regions have considerably different shapes in the electron $|\eta|$ distributions. For the conversion control region, the shape is mostly flat in the barrel region ($|\eta| < 1.5$) and peaks in the endcap region ($|\eta| > 1.5$), which is expected as the conversions are more likely to happen in the region with higher material budget. In contrast, the non-isolated control region peaks in the pseudorapidity region of $1.0 < |\eta| < 1.5$. Such behaviour is correlated with the jet $|\eta|$ distribution, since this control region favours non-isolated electrons coming from heavy quark decays.

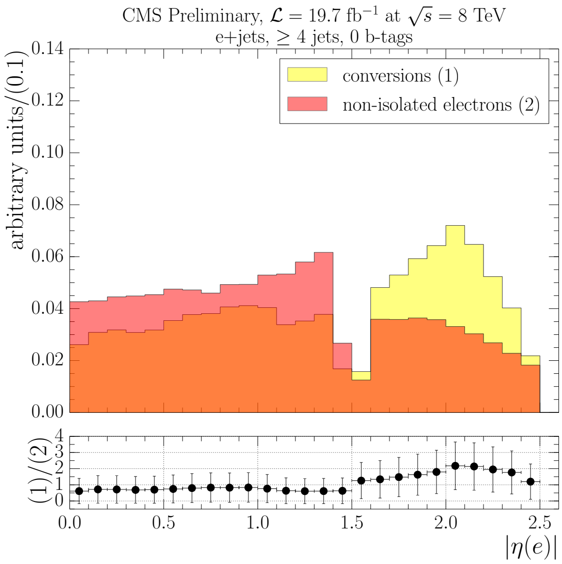
Also, in Figure 6.5 we see that while the conversion region is predicted to be a rather pure QCD sample, the non-isolated control region has significant contamination from $t\bar{t}$ signal and W/Z + jets events. For this reason, the conversion region was chosen to produce the QCD template for the electron channel. However, since the relative contributions of the QCD events from both control regions in the final



(a)



(b)



(c)

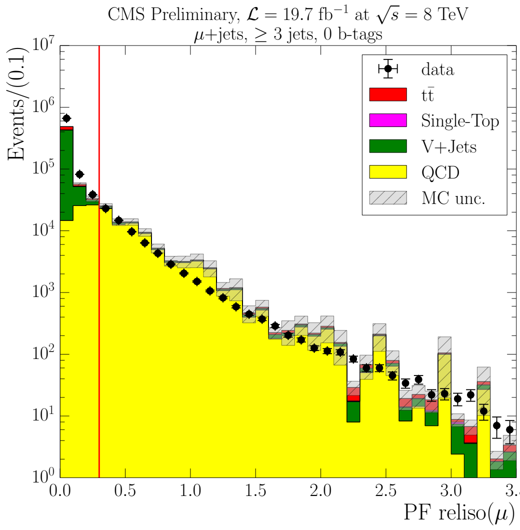
Figure 6.5: Distribution of the electron $|\eta|$ for conversion selection (a), non-isolated electron selection (b) and the shape comparison of both QCD control regions in data (c) in electron plus jets events with at least four jets and 0 b-tagged jets.

selected sample are not known, the real QCD shape can differ considerably from the prediction. Therefore, the non-isolated region is used to estimate the systematic uncertainty due to the QCD shape choice. This systematic error is estimated by substituting the nominal QCD template from the conversion control region with that from the non-isolated region, and calculating the change in the final measurement. More information on systematic uncertainties is presented in Section 6.6.

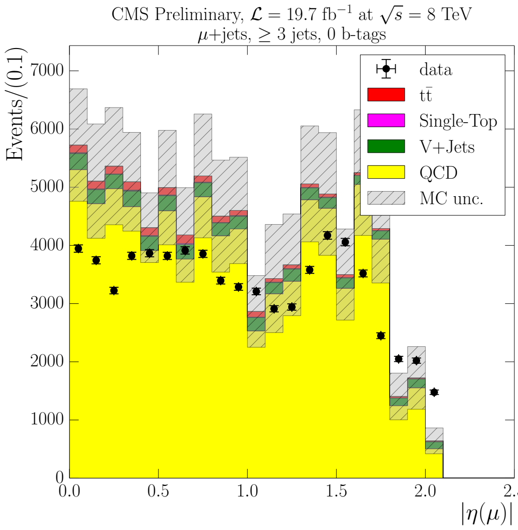
6.3.2 QCD multi-jet events with muons

For the QCD shape extraction in the muon plus jets channel, the inverted isolation cut is applied in the event selection: at least one muon with reliso > 0.3 is required, and events with isolated muons with reliso < 0.3 are vetoed. Similarly to the electron channel, only events with no b-tagged jets are taken into account. Additionally, to increase the available statistics, events with only three jets are also considered.

The relative isolation distribution with the cut value and the muon $|\eta|$ for the non-isolated region are shown on Figure 6.6. Clearly, the control region is dominated by QCD events and generally there is a reasonable agreement between data and simulation. The contamination from other processes is subtracted using MC.



(a)



(b)

Figure 6.6: Muon relative isolation with no isolation cut (a) and muon $|\eta|$ (b) for non-isolated (reliso > 0.3) selection in muon plus jets events with at least three jets and 0 b-tagged jets. The red line on plot (a) shows the reliso cut value of 0.3.

Table 6.5: Number of expected and observed events from MC and data in the electron channel before the fitting process. The uncertainties shown are purely statistical.

Selection step	$\mathrm{t\bar{t}} + $ **jets**	W + **jets**	Z + **jets**	**Single top**	**QCD**	**Sum MC**	**Data**
Preselection	1346476 ± 1028	1727900 ± 1138	380944 ± 234	169689 ± 262	130308513 ± 475118	133933524 ± 475120	13042702
Event cleaning/HLT	385009 ± 553	516544 ± 600	147515 ± 143	37308 ± 127	3854386 ± 83441	4940764 ± 83445	5846672
One isolated electron	342394 ± 522	446691 ± 548	100243 ± 117	33000 ± 120	578895 ± 29802	1501225 ± 29812	1688811
Muon veto	323759 ± 508	446579 ± 548	99805 ± 117	32301 ± 118	578860 ± 29802	1481306 ± 29812	1668851
Dilepton veto	319019 ± 504	446469 ± 548	73876 ± 101	32117 ± 118	578821 ± 29802	1450305 ± 29812	1628009
Conversion veto	310863 ± 498	430191 ± 538	70929 ± 99	31287 ± 116	321292 ± 21826	1164563 ± 21839	1396638
≥ 1 jets	310863 ± 498	430179 ± 538	70928 ± 99	31287 ± 116	321292 ± 21826	1164550 ± 21839	1396638
≥ 2 jets	310838 ± 498	429195 ± 536	70723 ± 99	31279 ± 116	320506 ± 21819	1162543 ± 21831	1396506
≥ 3 jets	306406 ± 494	386892 ± 493	63653 ± 90	29953 ± 114	226938 ± 16321	1013843 ± 16337	1215535
≥ 4 jets	174701 ± 373	76646 ± 179	13323 ± 35	9765 ± 67	44203 ± 4340	318640 ± 4361	351194
≥ 1 b-tagged jets	148287 ± 341	11280 ± 69	2152 ± 14	7675 ± 58	9304 ± 2055	178700 ± 2085	182481
≥ 2 b-tagged jets	70103 ± 228	1320 ± 23	337 ± 5	2900 ± 35	3035 ± 1789	77697 ± 1804	76379

6.4 Data/MC Comparison

In this section, the agreement between simulation and data for various kinematic distributions is reviewed after the final event selection. The event yields for all selection steps described in Section 6.2 are shown in Tables 6.5 and 6.6 for the electron and muon channels, respectively. All background events are normalised to the number of expected events for the total integrated luminosity of 19.7 fb^{-1}.

The QCD multi-jet background events shown in plots are derived using the data-driven methods explained previously. Figure 6.7 shows the lepton p_{T} and $|\eta|$ distributions for both electrons and muons. In Figure 6.8, missing transverse energy distributions are shown, also including the logarithmic scale for higher values of $E_{\mathrm{T}}^{\mathrm{miss}}$. Figure 6.9 shows the azimuthal angle of $E_{\mathrm{T}}^{\mathrm{miss}}$ ($\phi(E_{\mathrm{T}}^{\mathrm{miss}})$) and jet multiplicity distributions for both channels. The $\mathrm{t\bar{t}}$ uncertainty shown in the plots refers to the theoretical uncertainty on the NNLO calculation of the total $\mathrm{t\bar{t}}$ cross section at 8 TeV [51] used in this analysis.

Overall, good agreement is found between data and Monte Carlo simulation in all kinematic distributions. One should note that the comparison shown is made before the fitting process, which is described in Section 6.5.

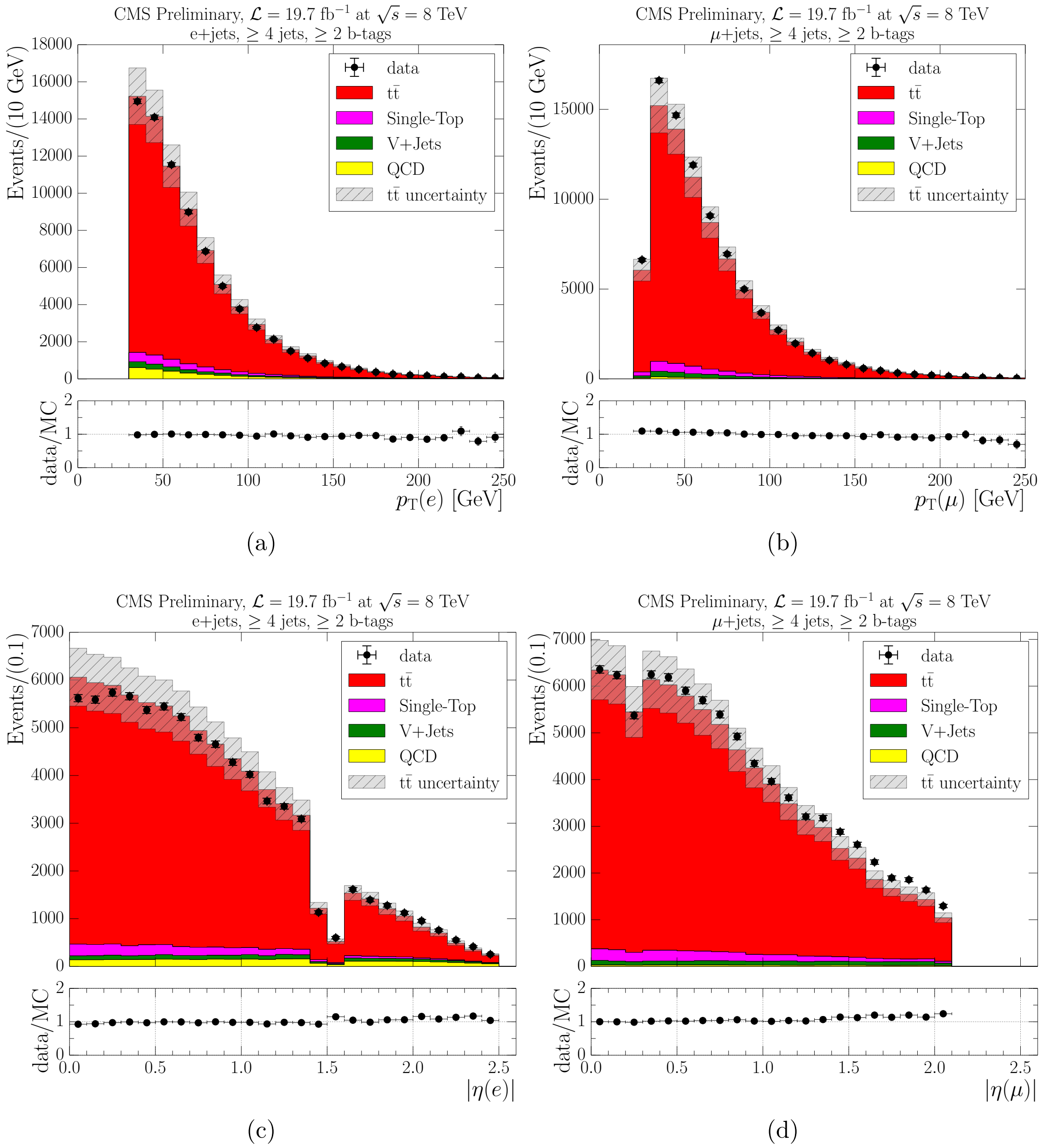


Figure 6.7: Data/MC comparison plots after the final event selection: electron p_T (a), muon p_T (b), electron $|\eta|$ (c) and muon $|\eta|$ (d). Left-hand plots: electron plus jets selection, right-hand plots: muon plus jets selection.

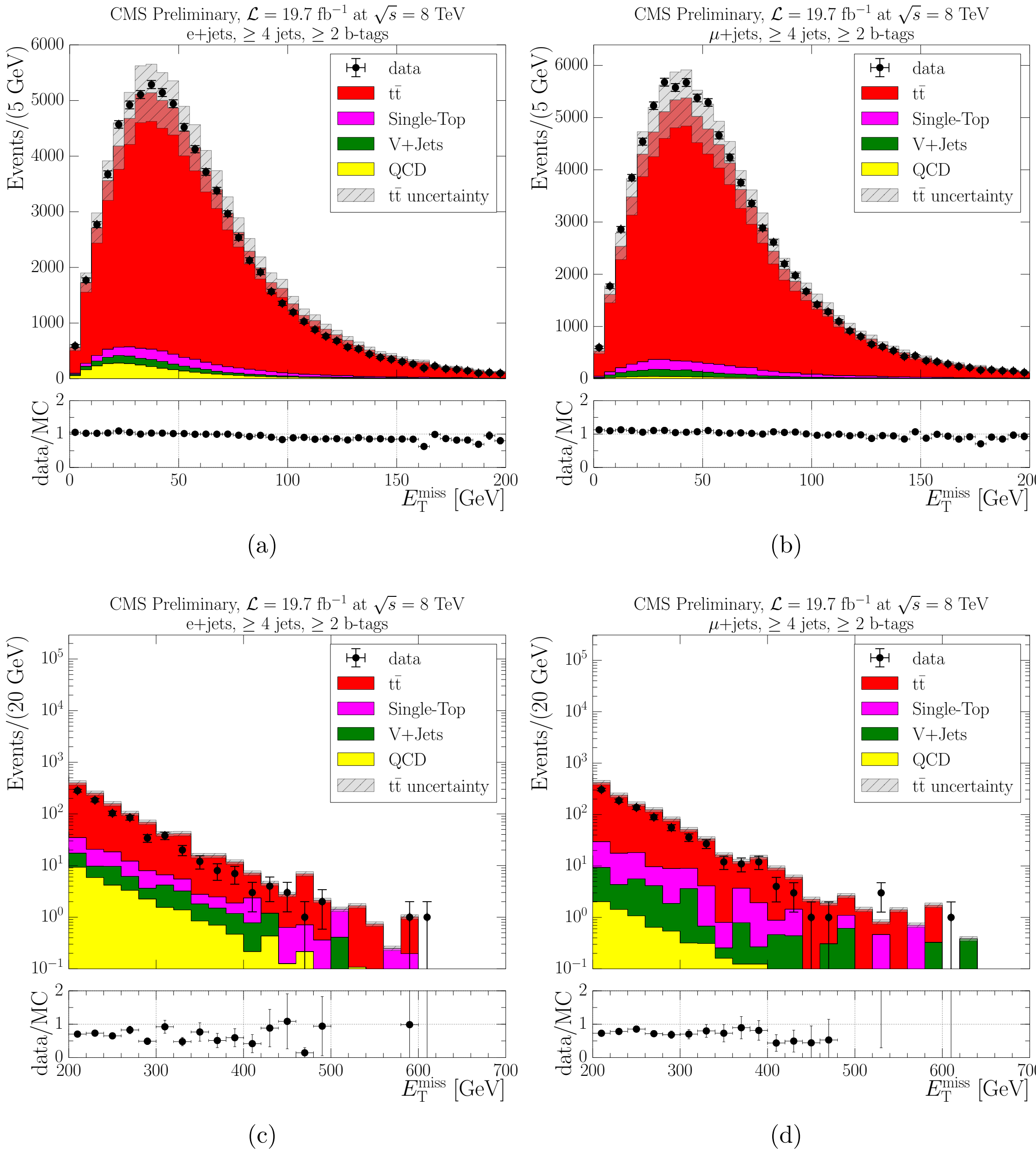


Figure 6.8: Data/MC comparison plots of $E_{\mathrm{T}}^{\mathrm{miss}}$ (a, b) and logarithmic view of $E_{\mathrm{T}}^{\mathrm{miss}}$ (c, d) after the final event selection. Left-hand plots: electron plus jets selection, right-hand plots: muon plus jets selection.

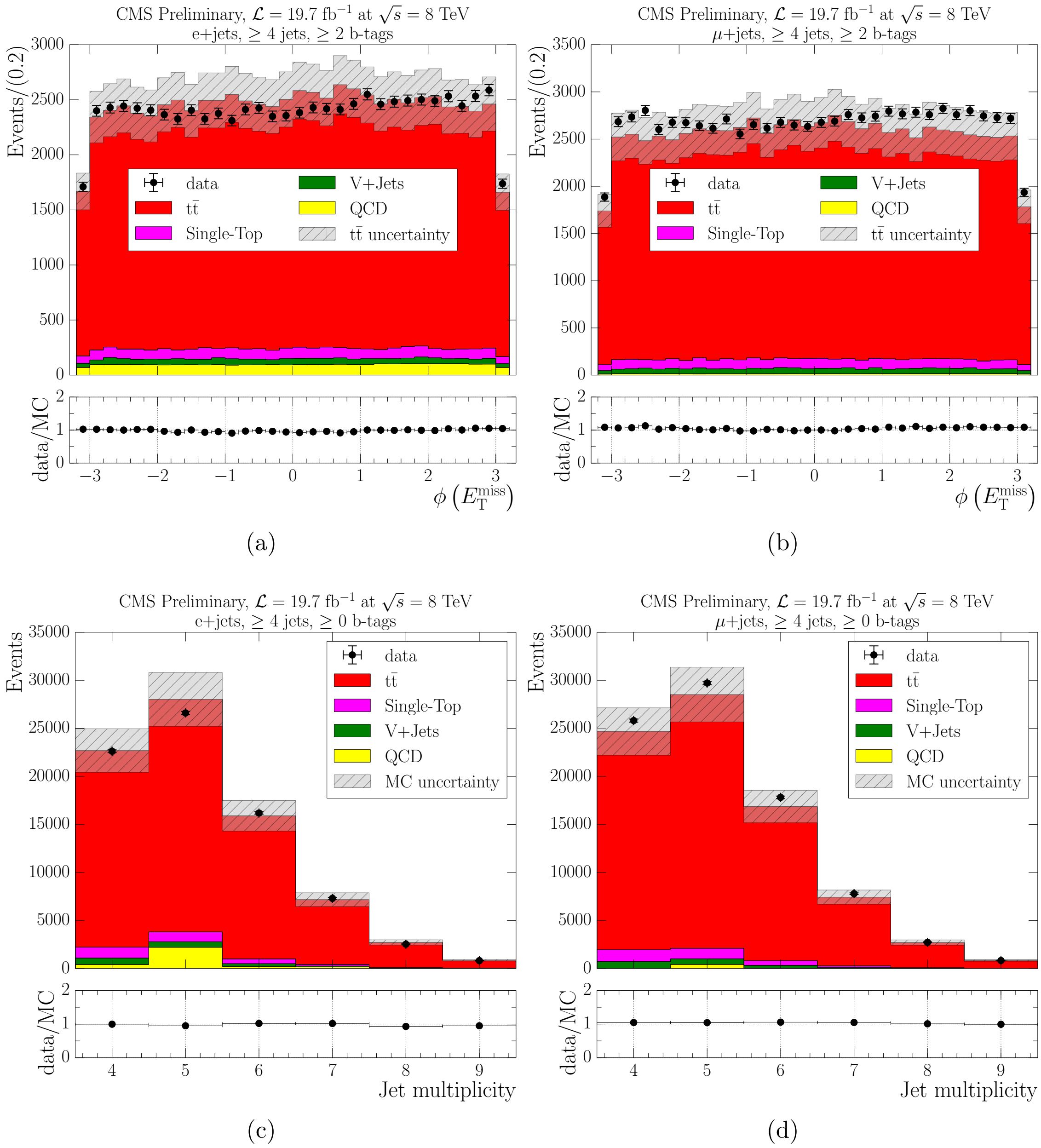


Figure 6.9: Data/MC comparison plots of $\phi(E_T^{miss})$ (a, b) and jet multiplicity (c, d) after the final event selection. Left-hand plots: electron plus jets selection, right-hand plots: muon plus jets selection.

Table 6.6: Number of expected and observed events from MC and data in the muon channel before the fitting process. The uncertainties shown are purely statistical.

Selection step	$\mathrm{t\bar{t}}+$**jets**	$\mathrm{W}+$**jets**	$\mathrm{Z}+$**jets**	**Single top**	**QCD**	**Sum MC**	**Data**
Preselection	1346476 ± 1028	1727900 ± 1138	380944 ± 234	169689 ± 262	104079124 ± 236784	107704135 ± 236789	20284215
Event cleaning/HLT	443056 ± 581	644505 ± 688	148143 ± 142	44727 ± 135	1664036 ± 33747	2944468 ± 33759	3063569
One isolated muon	360897 ± 521	473825 ± 556	88283 ± 106	35546 ± 121	83535 ± 5833	1042088 ± 5885	1327738
Second muon veto	353646 ± 516	473776 ± 556	53263 ± 82	35260 ± 120	82863 ± 5823	998811 ± 5874	1254896
Electron veto	336474 ± 503	473631 ± 556	52935 ± 82	34633 ± 119	82841 ± 5823	980516 ± 5873	1237495
≥ 1 jets	336474 ± 503	473622 ± 556	52934 ± 82	34633 ± 119	82841 ± 5823	980507 ± 5873	1237495
≥ 2 jets	336457 ± 503	472377 ± 554	52768 ± 81	34620 ± 119	81211 ± 5777	977436 ± 5827	1237428
≥ 3 jets	332863 ± 501	425334 ± 507	47493 ± 74	33146 ± 117	33505 ± 2726	872343 ± 2822	1108272
≥ 4 jets	188466 ± 377	83248 ± 184	10144 ± 29	10556 ± 67	7006 ± 1155	299422 ± 1231	340786
≥ 1 b-tagged jets	160223 ± 344	12341 ± 71	1690 ± 12	8322 ± 59	3763 ± 910	186341 ± 977	196667
≥ 2 b-tagged jets	76068 ± 231	1465 ± 24	248 ± 4	3096 ± 35	481 ± 413	81361 ± 476	85028

6.5 Differential Cross Section Measurement

6.5.1 Primary variables

In this analysis, the normalised differential cross section of $\mathrm{t\bar{t}}$ production is studied with respect to five primary variables:

- $E_{\mathrm{T}}^{\mathrm{miss}}$, or magnitude of missing transverse momentum;
- H_{T}, or scalar sum of jet transverse momenta;
- S_{T}, or scalar sum of the transverse momenta of all objects in the event;
- $M_{\mathrm{T}}^{\mathrm{W}}$, or transverse mass of the leptonically decaying W boson;
- $p_{\mathrm{T}}^{\mathrm{W}}$, or transverse momentum magnitude of the leptonically decaying W boson.

These observables are often called event-level (or global) variables due to the fact that no high-level kinematic reconstruction (e.g. that of a $\mathrm{t\bar{t}}$ pair) is required to calculate them. A lower level of complexity makes these variables less prone to inefficiencies, systematic errors or biases arising from various kinematic reconstruction techniques. Hence, global variables can provide better sensitivity in comparison of different theoretical models or generator tunes.

$\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is commonly referred to as the missing transverse energy, which is equivalent to missing transverse momentum under assumption that the missing particles contributing to $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ are massless. $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative of the vector sum of the transverse momenta of all reconstructed particles in an event. Therefore, its direction is opposite to the observed sum of transverse momentum, and its magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$ is given by:

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\left[\left(\sum_i p_x^i\right)^2 + \left(\sum_i p_y^i\right)^2\right]^{\frac{1}{2}}$$

where the sums are calculated over all measured particles in the event.

H_{T} is defined as the scalar sum of the transverse momenta of all reconstructed jets in the event:

$$H_{\mathrm{T}} = \sum_{\text{all jets}} p_{\mathrm{T}}^{\text{jet}}$$

Here all jets are required to have a $p_{\mathrm{T}} > 20$ GeV as described in Section 6.2.

S_{T} represents the sum of H_{T}, $E_{\mathrm{T}}^{\text{miss}}$ and the magnitude of p_{T} of the single isolated lepton:

$$S_{\mathrm{T}} = H_{\mathrm{T}} + E_{\mathrm{T}}^{\text{miss}} + p_{\mathrm{T}}^{\text{lepton}}$$

$p_{\mathrm{T}}^{\mathrm{W}}$ refers to the magnitude of transverse momentum of the leptonically decaying W boson, which is calculated from the single isolated lepton and $E_{\mathrm{T}}^{\text{miss}}$ assumed to originate from the $\mathrm{t\bar{t}}$ decay:

$$p_{\mathrm{T}}^{\mathrm{W}} = \sqrt{(p_x^{\text{lepton}} + p_x^{\text{miss}})^2 + (p_y^{\text{lepton}} + p_y^{\text{miss}})^2}$$

Finally, $M_{\mathrm{T}}^{\mathrm{W}}$ is defined as the transverse mass of the leptonically decaying W boson, also using the single isolated lepton and $E_{\mathrm{T}}^{\text{miss}}$:

$$M_{\mathrm{T}}^{\mathrm{W}} = \sqrt{(E_{\mathrm{T}}{}^{\text{lepton}} + E_{\mathrm{T}}^{\text{miss}})^2 - p_{\mathrm{T}}^{\mathrm{W}\,2}}$$

where $E_{\mathrm{T}}{}^{\text{lepton}}$ denotes the transverse energy of the lepton.

6.5.2 Choice of binning

The choice of binning for primary variables is an important part of a differential cross section measurement. It is directly influenced by available statistics and detector resolution. An ideal measurement would benefit from the cross section calculation in as many bins as possible. However, insufficient statistics leads to high statistical uncertainty in a particular bin, whereas limited detector resolution causes substantial migration effects between bins. Bin migration happens if an event originating from one bin appears in a different bin after reconstruction. Unreasonable binning choice would lead to a high number of such events, which can potentially compromise the cross section measurement.

To quantify the bin migration, purity and stability variables are introduced, defined as:

$$p_i = \frac{N_i^{\text{rec\&gen}}}{N_i^{\text{rec}}} \tag{6.1}$$

$$s_i = \frac{N_i^{\text{rec\&gen}}}{N_i^{\text{gen}}} \tag{6.2}$$

where $N_i^{\mathrm{rec\&gen}}$ is the number of events both generated and reconstructed in the same bin, and N_k^{rec} (N_i^{gen}) is the number of reconstructed (generated) events within a bin.

The purity p_i quantifies migration into a bin i, whereas the stability s_i is sensitive to migration out of a bin. The binning is chosen in such a way that both purity and stability are above 0.5, or do not fall far below this value. This process is performed for both electron and muon channels, and common binning is derived in order to be able to combine both channels in the final cross section measurement.

Figure 6.10 shows the two-dimensional distributions of reconstructed versus generated $E_{\mathrm{T}}^{\mathrm{miss}}$, which were used to calculate purity and stability. The distributions were obtained with the nominal (MadGraph + PYTHIA) signal $\mathrm{t\bar{t}}$ sample. Red lines represent the bin edges for the chosen binning. Tables 6.7 and 6.8 show the numerical values of purity and stability, as well as the selected binning for $E_{\mathrm{T}}^{\mathrm{miss}}$ variable. Results for other primary variables can be found in Appendix B.

Table 6.7: Stability and purity of $E_{\mathrm{T}}^{\mathrm{miss}}$ bins in the electron channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$0 < E_{\mathrm{T}}^{\mathrm{miss}} < 25$	$25 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 45$	$45 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 75$	$75 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 100$	$100 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 150$	$E_{\mathrm{T}}^{\mathrm{miss}} \geq 150$
events	10337	17982	19503	12594	7220	2832
purity	0.52	0.48	0.44	0.42	0.54	0.69
stability	0.42	0.41	0.47	0.52	0.66	0.86

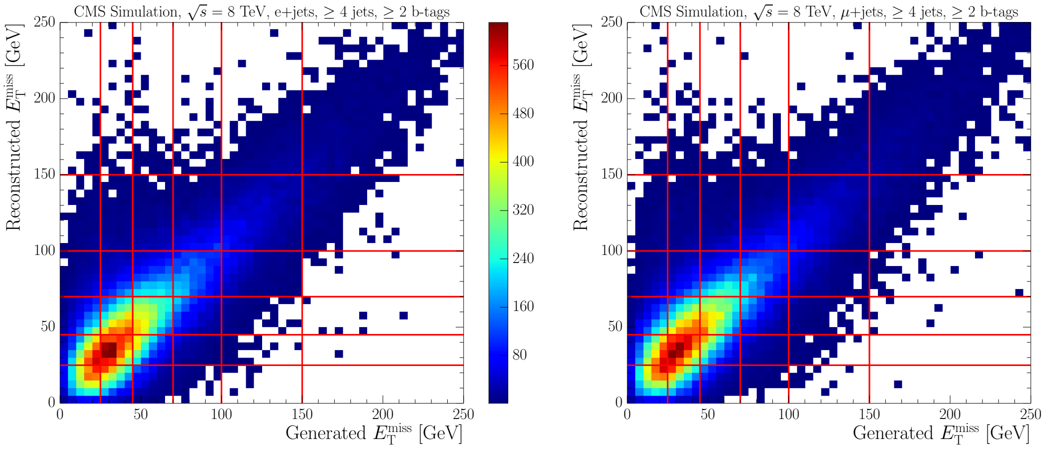
Table 6.8: Stability and purity of $E_{\mathrm{T}}^{\mathrm{miss}}$ bins in the muon channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$0 < E_{\mathrm{T}}^{\mathrm{miss}} < 25$	$25 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 45$	$45 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 75$	$75 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 100$	$100 \leq E_{\mathrm{T}}^{\mathrm{miss}} < 150$	$E_{\mathrm{T}}^{\mathrm{miss}} \geq 150$
events	11506	20348	22594	14413	8417	3274
purity	0.52	0.49	0.45	0.42	0.52	0.69
stability	0.43	0.42	0.47	0.5	0.66	0.87

6.5.3 Fitting procedure

To calculate the differential cross section as a function of a primary variable, the number of signal $\mathrm{t\bar{t}}$ events has to be measured in each primary variable bin. For this purpose, a template fit of the observed distribution of lepton pseudorapidity ($|\eta_l|$) is performed in each bin in order to estimate the number of $\mathrm{t\bar{t}}$ events by subtracting the primary backgrounds. Three separate normalised templates are produced for the input of the template fit:

- $|\eta_l|$ distribution of top-like events, i.e. combined template of $\mathrm{t\bar{t}}$ and single top MC;
- $|\eta_l|$ distribution of V+jets events, i.e. combined W+jets and Z+jets templates;



(a) (b)

Figure 6.10: Reconstructed versus generated E_T^{miss} for electron plus jets (a) and muon plus jets events (b).

- $|\eta_l|$ shape of QCD events, estimated from data (Section 6.3).

The combination of W+jets and Z+jets templates is justified by the similarity of the shapes of distributions, as well as by limited MC statistics. Figures 6.11 and 6.12 show templates for the E_T^{miss} variable for electron and muon channels, respectively. Templates for other primary variables are attached in Appendix C.

The template fit is performed independently in each primary variable bin by minimising the log-likelihood defined as:

$$LL(\lambda_i, d_i) = -2\,\log\left(\prod_i \frac{\lambda_i^{d_i}\cdot e^{-\lambda_i}}{d_i!}\right) = -2\sum_i \log\left(\frac{\lambda_i^{d_i}\cdot e^{-\lambda_i}}{d_i!}\right), \tag{6.3}$$

where λ_i is the sum of expected events from all the templates d_i is the number of data events in each bin i, and summation is performed over all bins i of the $|\eta_l|$ distribution. The expected lepton pseudorapidity (array of λ_i) is modelled by normalising the templates $\theta_{i\,j}$ by normalisation factors N_j:

$$\lambda_i = \sum_j N_j \theta_{i\,j}, \text{ where } \sum_i \theta_{i\,j} = 1 \text{ for every process } j. \tag{6.4}$$

The minimisation of the log-likelihood is achieved by varying the fit parameters, i.e. normalisation factors for each template. Therefore, the template fit strives for equality of λ_i and d_i in all lepton pseudorapidity bins. The initial normalisation values for the templates are taken from the expected number of events for $\mathrm{t\bar{t}}$, single top, W + jets and Z + jets as per Table 6.1. Although the QCD template shape is data-driven, the initial normalisation for it is also taken from Monte Carlo simulation (Tables 6.2 and 6.3).

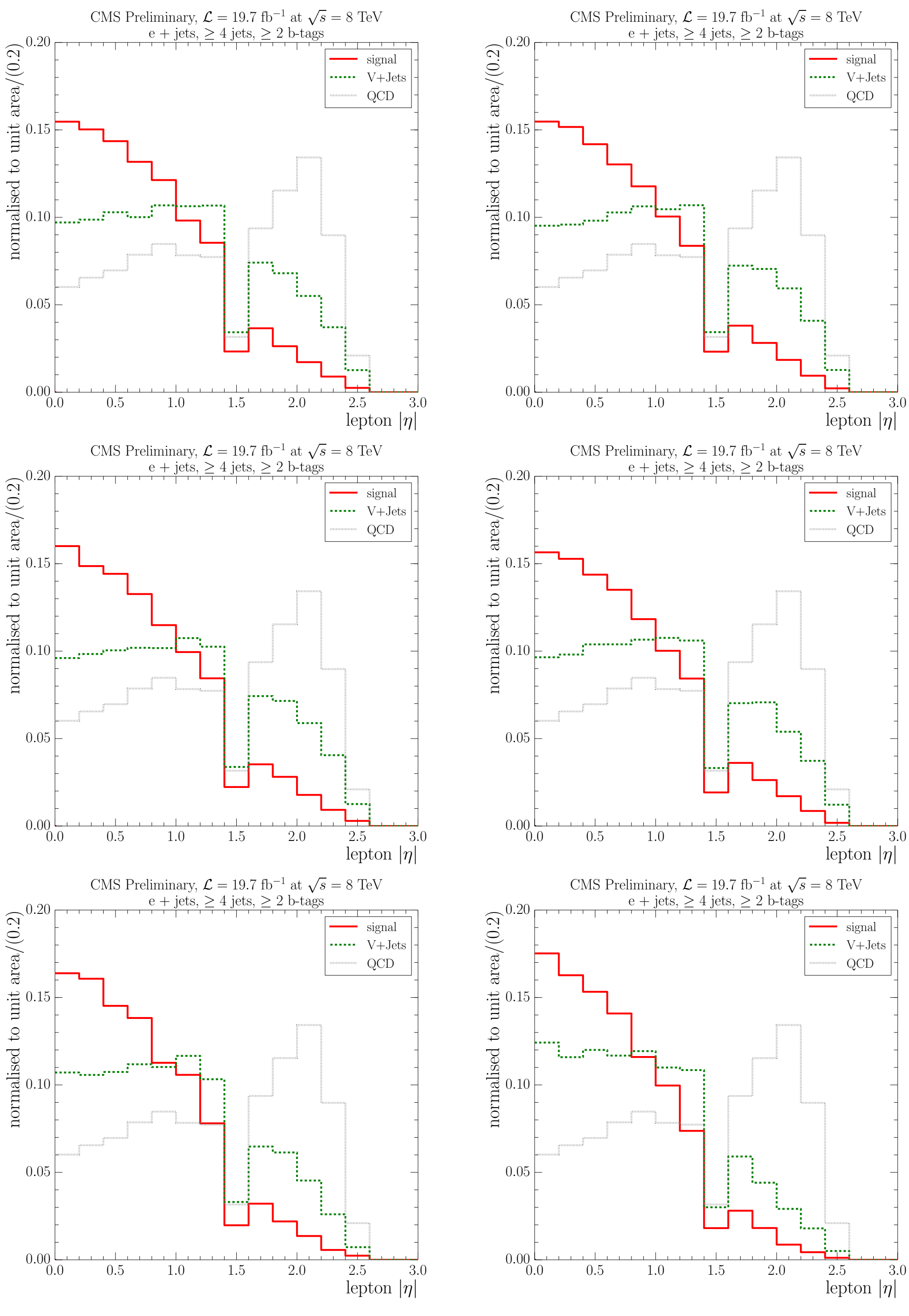


Figure 6.11: Electron $|\eta|$ templates for the fit in different bins of $E_{\mathrm{T}}^{\mathrm{miss}}$, from top left to bottom right: 0–25 GeV, 25–45 GeV, 45–70 GeV, 70–100 GeV, 100–150 GeV and ≥ 150 GeV.

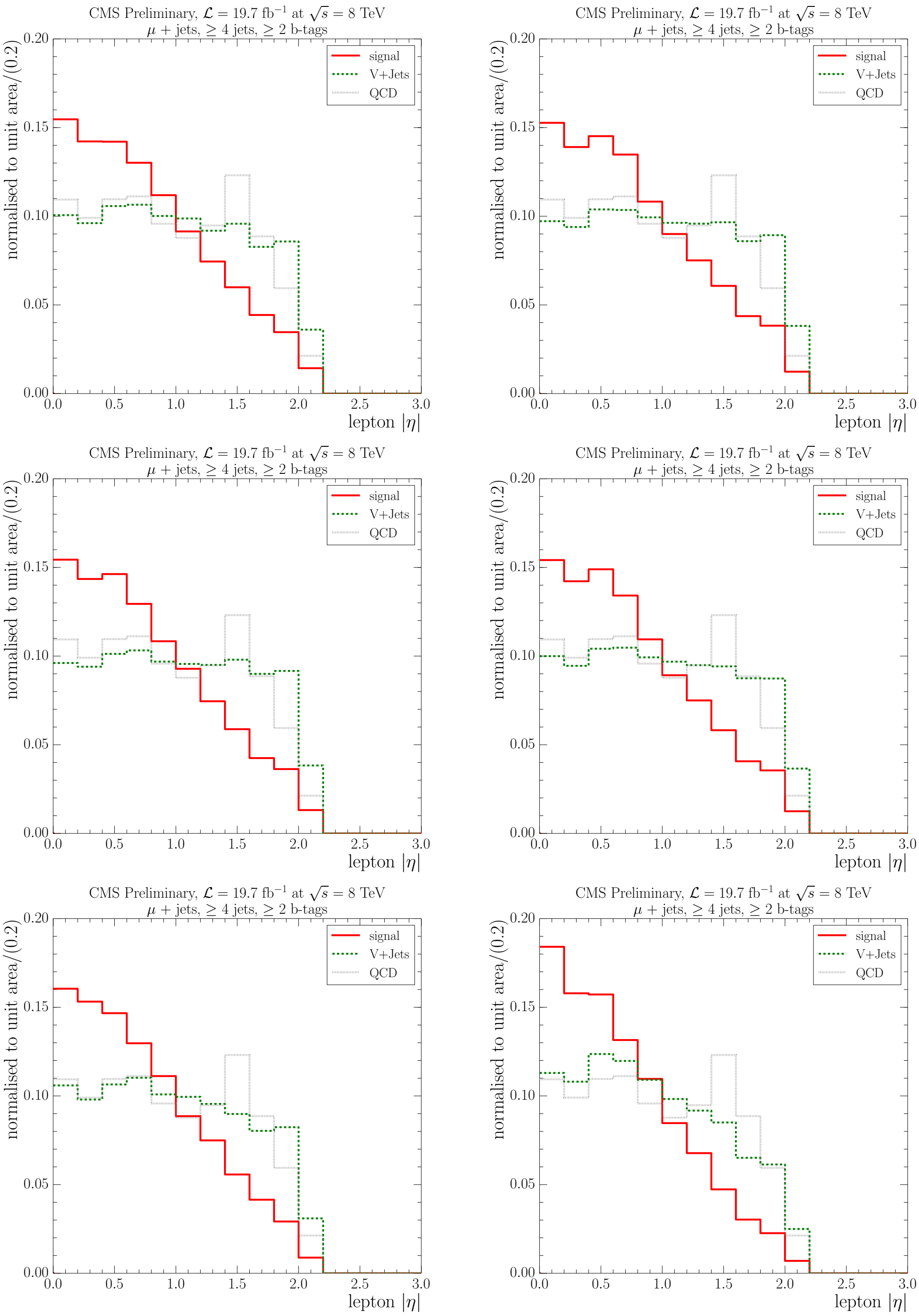


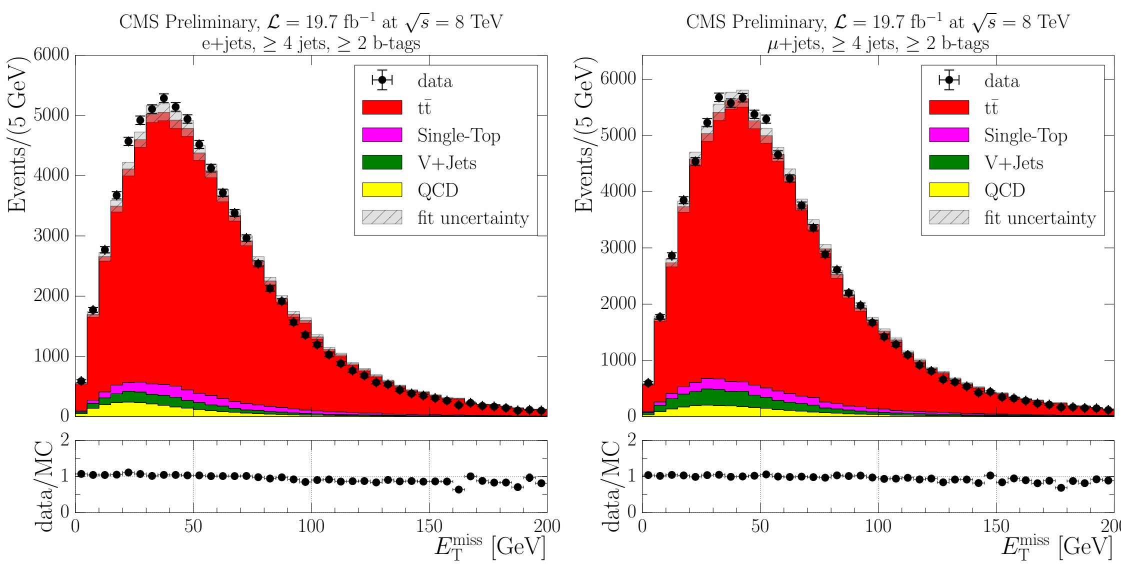
Figure 6.12: Muon $|\eta|$ templates for the fit in different bins of $E_{\mathrm{T}}^{\mathrm{miss}}$, from top left to bottom right: 0–25 GeV, 25–45 GeV, 45–70 GeV, 70–100 GeV, 100–150 GeV and $\geq$ 150 GeV.

The similarity of QCD and V + jets templates, particularly for muon plus jets events (Figure 6.12), can lead to unphysical bin-to-bin fluctuations of these backgrounds. To mitigate such behaviour, the following conservative Gaussian constraints were added to the likelihood function of the fit:

$$(N^{\text{fit}}_{\text{V+jets}} - N^{\text{MC}}_{\text{V+jets}})^2/(0.5 \times N^{\text{MC}}_{\text{V+jets}})^2 \tag{6.5}$$
$$(N^{\text{fit}}_{\text{QCD}} - N^{\text{MC}}_{\text{QCD}})^2/(2 \times N^{\text{MC}}_{\text{QCD}})^2 \tag{6.6}$$

The first term represents the constraint of the number of V+jets events to within 50 % of that predicted by theory, whereas the second term is a 200 % constraint for QCD events.

Figure 6.13 shows the control plots of data/MC comparison for the $E_{\mathrm{T}}^{\mathrm{miss}}$ variable in both electron and muon channels. Here the normalisation for signal and backgrounds is taken from the fitted results. Similar plots for H_{T} and S_{T} are shown in Figure 6.14, whereas Figure 6.15 shows control plots for $p_{\mathrm{T}}^{\mathrm{W}}$ and $M_{\mathrm{T}}^{\mathrm{W}}$. Overall, the data and simulation agree within the fit uncertainty. The observed discrepancies can be attributed to the known issue of the event generators incorrectly modelling the p_{T} spectrum of the top quarks [131, 132]. This effect is accounted for as a source of systematic uncertainty, as explained in Section 6.6.



(a) (b)

Figure 6.13: Data/MC comparison plots of $E_{\mathrm{T}}^{\mathrm{miss}}$ for electron plus jets (a) and muon plus jets (b) events after the final event selection using the normalisation from the fitted results.

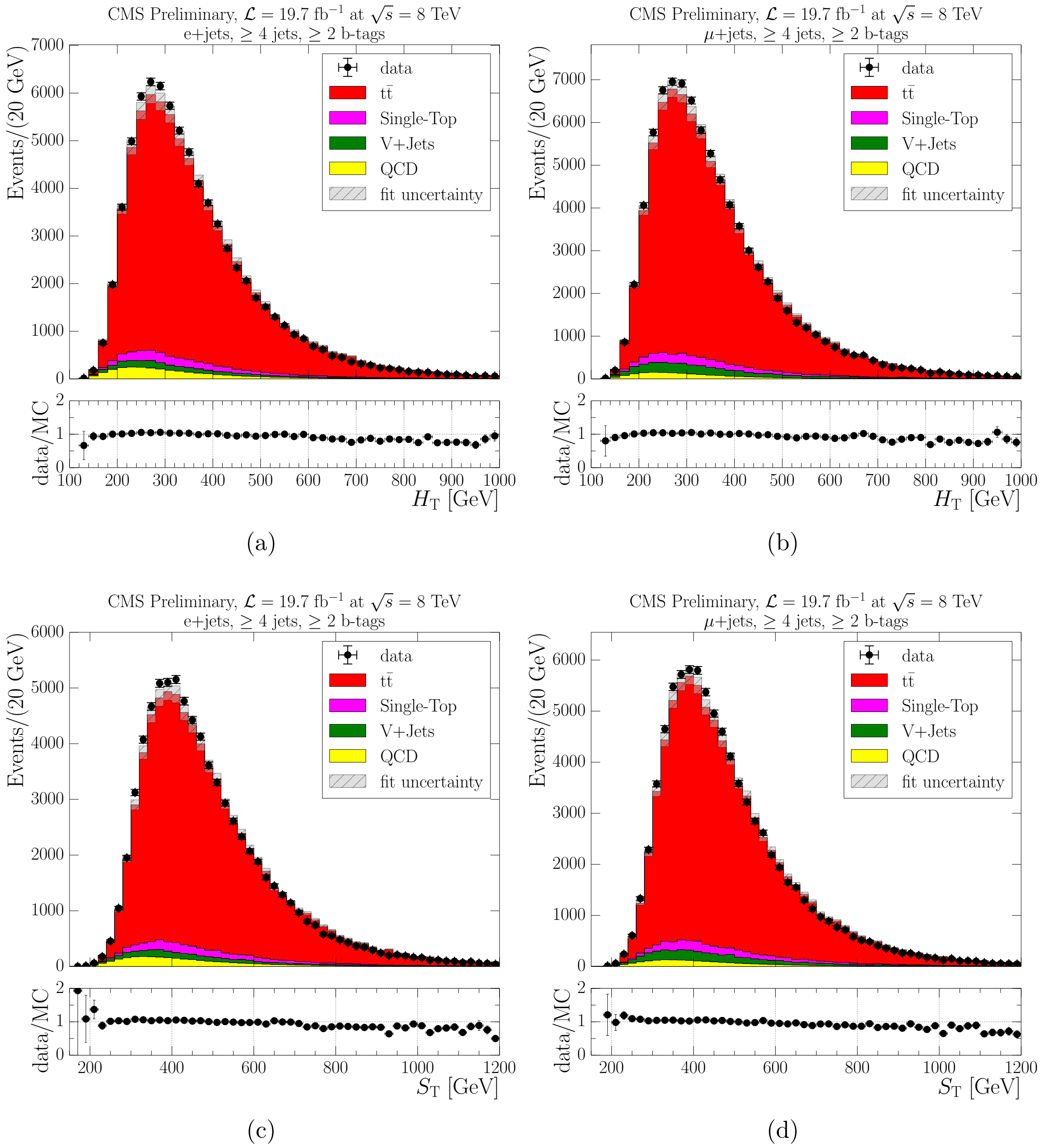


Figure 6.14: Data/MC comparison plots of H_T (a, b) and S_T (c, d) after the final event selection using the normalisation from the fitted results. Left-hand plots: electron plus jets selection, right-hand plots: muon plus jets selection.

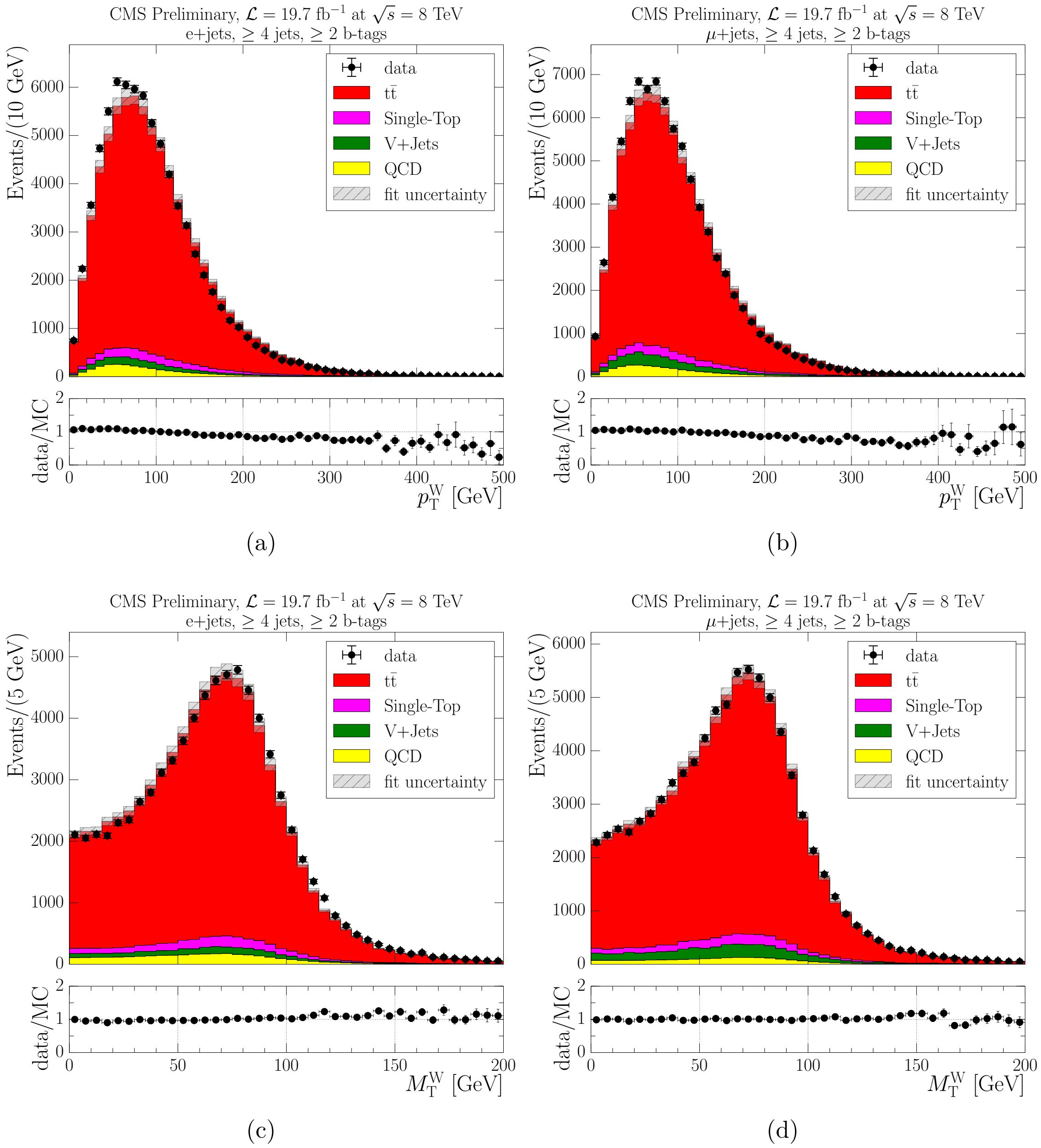


Figure 6.15: Data/MC comparison plots of p_T^W (a, b) and M_T^W (c, d) after the final event selection using the normalisation from the fitted results. Left-hand plots: electron plus jets selection, right-hand plots: muon plus jets selection.

The output of the fit yields the number of top-like events in each bin, which includes both $t\bar{t}$ and single top processes since the lepton pseudorapidity distribution is very similar between them. The number of $t\bar{t}$ events is extracted from the signal fit result by using the expected ratio of $t\bar{t}$ and single top events before the fitting process:

$$N^{\text{fit}}_{t\bar{t}} = N^{\text{fit}}_{\text{signal}} \times \frac{N^{\text{exp}}_{t\bar{t}}}{N^{\text{exp}}_{\text{signal}}}, \tag{6.7}$$

where $N^{\text{fit}}_{t\bar{t}}$ ($N^{\text{fit}}_{\text{signal}}$) is the number of $t\bar{t}$ ($t\bar{t}$ + single top) events as found by the fit, and $N^{\text{exp}}_{t\bar{t}}$ ($N^{\text{exp}}_{\text{signal}}$) is the number of $t\bar{t}$ ($t\bar{t}$ + single top) events as predicted by theory.

6.5.4 Unfolding

Measurements of physical observables are inevitably distorted by various detector effects, including limited detector resolution, experimental acceptance and selection inefficiency. Therefore, it is often very challenging to compare the results with theory predictions, as well as other experiments. To tackle this issue, unfolding can be used to estimate true underlying distributions of measured physical observables.

In order to describe the concept of unfolding, let us consider the distribution of a measured observable as a vector b. The measurement of this observable can be usually simulated by using Monte Carlo techniques according to some theoretical model and detector simulation. Let x_0 be the "true" generated distribution, i.e. unaffected by detector effects, and b_0 the corresponding measured distribution obtained by performing reconstruction in simulation:

$$\hat{A}x_0 = b_0. \tag{6.8}$$

Here $\hat{A}$ is the response matrix of the detector, representing all aforementioned detector effects. Therefore, the given measured observable b and corresponding true distribution x are similarly related:

$$\hat{A}x = b. \tag{6.9}$$

Attempting to solve this system of equations by direct inversion of the matrix $\hat{A}$ usually results in rapidly oscillating, unacceptable solutions. Therefore, a more sophisticated approach is needed. The process of finding the true distribution x is referred to as an unfolding, and in this particular analysis, Singular Value Decomposition (SVD) method of unfolding [133] is used.

The SVD unfolding incorporates the following factorisation of the response matrix:

$$\hat{A} = USV^T, \tag{6.10}$$

where S is a diagonal matrix with non-negative elements (called singular values of matrix $\hat{A}$), and U and V are orthogonal matrices (called the left and right singular vectors). The inverse of the response matrix then takes the form:

$$\hat{A}^{-1} = VS^{-1}U^{T}. \tag{6.11}$$

Such factorisation allows simple manipulation of the response matrix and its inverse, which is particularly useful for ill-determined cases of near-degenerate matrices when singular values of $\hat{A}$ are close to zero. Additionally, if statistical errors of the measurement are substantial, it can result in amplification of essentially random components in the solution. These issues can be resolved by implementing regularisation in the unfolding procedure. The regularisation parameter k is used to select the number of statistically significant equations [133]. The value of k is determined by analysing the vector d obtained by rotating the measured distribution:

$$d = U^{T} \times b \tag{6.12}$$

For a reasonably smooth measured distribution (an *a priori* knowledge used for regularisation), only the first few terms of the decomposition are expected to be significant, with the rest of the terms corresponding to quickly oscillating vectors. This implies that the distribution d_i of the vector d components is expected to exponentially fall towards a Gaussian-distributed random values for large i. The number of first few significant components ($|d_i| \gtrsim 1$) is therefore chosen as a regularisation parameter (see Figure 6.17, (c)).

In this work, the RooUnfold package [134] is used to implement the SVD method of data unfolding. The unfolded $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the generated p_{T} of the neutrino produced in the semileptonic $\mathrm{t\bar{t}}$ decay. The underlying H_{T} (and therefore S_{T}) distribution is calculated by applying the full jet clustering reconstruction described in Section 3.5.3 to the generated particles, with the $p_{\mathrm{T}} > 20$ GeV requirement for the generated jets. $M_{\mathrm{T}}^{\mathrm{W}}$ and $p_{\mathrm{T}}^{\mathrm{W}}$ variables are constructed from generated $E_{\mathrm{T}}^{\mathrm{miss}}$ and leptons, as shown in Section 6.5.1. The response matrices are built from true and measured distributions of primary variables, based on $\mathrm{t\bar{t}}$ Monte Carlo samples.

To test and validate the performance of the unfolding, a set of 300 pseudo-experiments (also known as toy MC) was created by generating a random number of events in each bin of a primary variable according to a Poisson distribution with the mean around the initial central value. This was performed for both true and measured distributions in the unfolding, with the response matrix recalculated accordingly. Then the unfolding procedure is performed for the measured distribution in each pseudo-experiment by using the SVD decomposition of a response matrix from every other pseudo-experiment, making a total of $300 \times 299 = 89700$ different measurements. For each combination, a pull is calculated:

$$p_i = \frac{N_i^{\mathrm{unfolded}} - N_i^{\mathrm{truth}}}{\sigma_i}, \tag{6.13}$$

where σ_i is the combined statistical and unfolding uncertainty in bin i. The pull distributions for all bins of the $E_{\mathrm{T}}^{\mathrm{miss}}$ variable are shown in Figure 6.16. Clearly, there is no significant bias in the unfolding procedure, as all the distributions are centred around zero. The width of the distributions is close to unity, meaning that the errors are reliably calculated. The unfolding performance was similarly validated for other primary variables.

To further check the robustness of the SVD approach, a series of closure tests were performed for all possible regularisation parameters in the unfolding of all primary variables. A closure test is carried out by running the unfolding on the number of $\mathrm{t\bar{t}}$ events in the reconstructed MC sample used to obtain the response matrix. For a correct procedure, the unfolded values should correspond exactly to the true generated number of events obtained from the same MC sample. Figure 6.17 shows the results of these tests for $E_{\mathrm{T}}^{\mathrm{miss}}$ variable in the electron channel, suggesting that the unfolding performs as expected for two different choices of binning.

It can also be seen that non-optimal binning leads to an increased unfolding uncertainty. The regularisation parameters k were chosen by fitting the d_i distribution (mentioned above) to an exponential. The optimal k-value for the standard $E_{\mathrm{T}}^{\mathrm{miss}}$ binning obtained from plot (c) on Figure 6.17 is 3, whereas the increased number of bins in plot (d) leads to a k-value of 7. Plots (a) and (b) show that for these k-values, the unfolding uncertainty is significantly larger for the increased number of $E_{\mathrm{T}}^{\mathrm{miss}}$ bins compared to the standard binning chosen as described in Section 6.5.2. This can be explained by the increased migration effects between smaller bins, and therefore less stable unfolding procedure.

6.5.5 Differential cross section calculation

The unfolded number of $\mathrm{t\bar{t}}$ events $N_{\mathrm{t\bar{t}}}^{k}$ can be converted to a partial cross section of $\mathrm{t\bar{t}}$ production in each bin k of a given variable:

$$\Delta\sigma_{\mathrm{t\bar{t}}}^{k} = \frac{N_{\mathrm{t\bar{t}}}^{k}}{\mathrm{BR} \times \mathcal{L}}, \tag{6.14}$$

where $\mathcal{L}$ is the total integrated luminosity and BR is the theoretical branching ratio of semileptonic $\mathrm{t\bar{t}}$ decay. These values cancel in the normalisation of the cross section, which is performed by division by the bin width $\Delta\mathrm{X}^{k}$ of a given variable X,

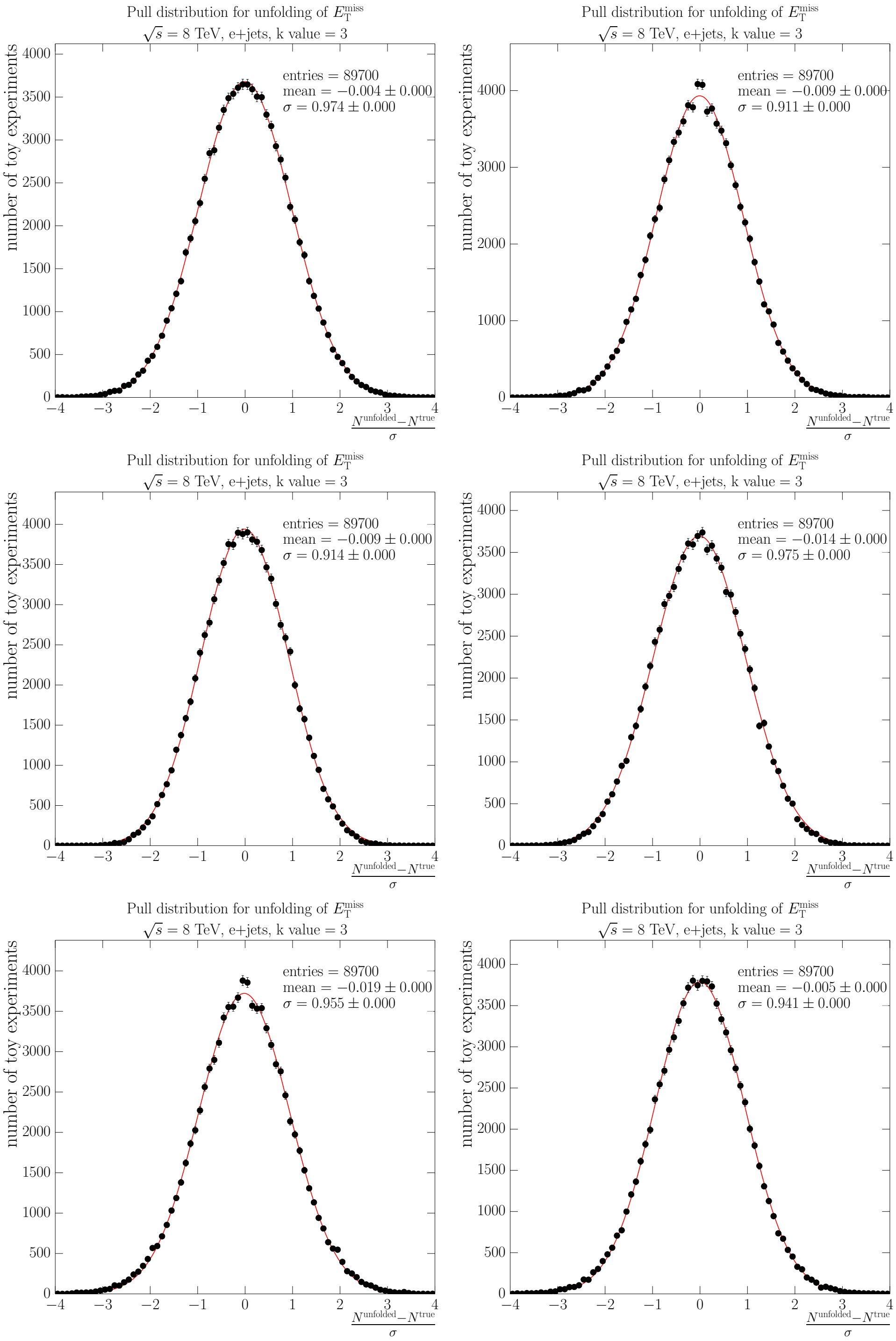
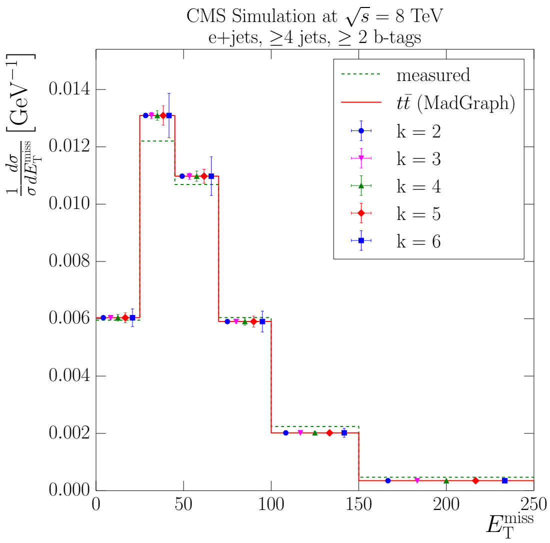
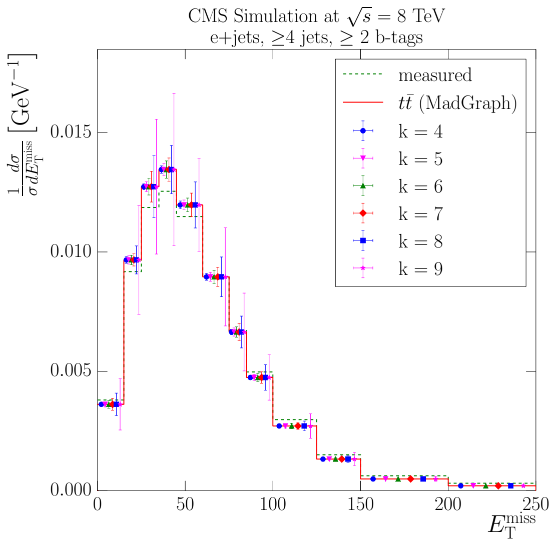


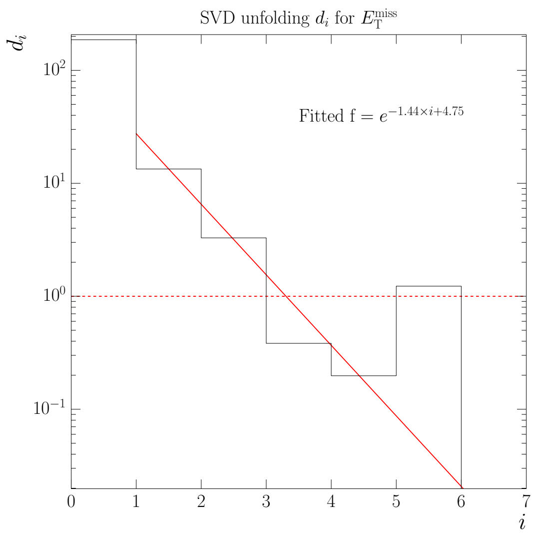
Figure 6.16: Pull distributions for the SVD unfolding method for all $E_{\mathrm{T}}^{\mathrm{miss}}$ bins, from top left to bottom right: 0 GeV to 25 GeV, 25 GeV to 45 GeV, 45 GeV to 70 GeV, 70 GeV to 100 GeV, 100 GeV to 150 GeV and $\geq$ 150 GeV.



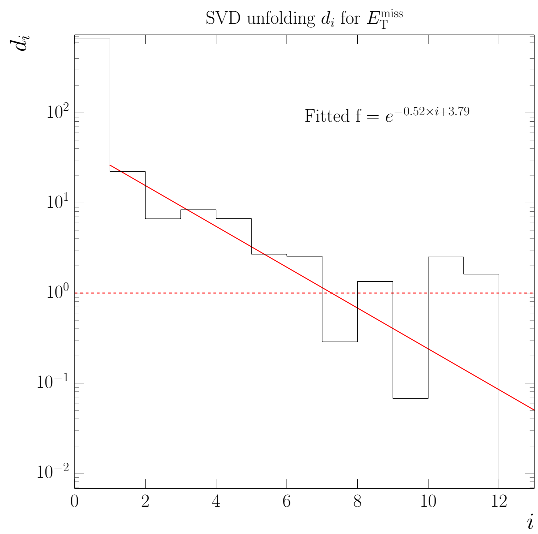
(a)



(b)



(c)



(d)

Figure 6.17: Unfolding performance tests for $E_{\mathrm{T}}^{\mathrm{miss}}$. Plots (a) and (b) show the results of closure tests performed by unfolding the reconstructed MadGraph distribution (shown as 'measured') back to the generated values. Plots (c) and (d) show the corresponding d_i distributions used to choose the optimal regularisation parameter. (a) and (c) represent the standard binning chosen in the analysis; (b) and (d) correspond to the doubled number of $E_{\mathrm{T}}^{\mathrm{miss}}$ bins.

yielding the differential cross section:

$$\frac{\mathrm{d}\sigma^k_{\mathrm{t\bar{t}}}}{\mathrm{dX}} = \frac{\Delta\sigma^k_{\mathrm{t\bar{t}}}}{\Delta\mathrm{X}^k}. \tag{6.15}$$

Finally, the normalised differential cross section is obtained by using the sum of the partial differential cross sections:

$$\frac{1}{\sigma^{\mathrm{tot}}_{\mathrm{t\bar{t}}}}\frac{\mathrm{d}\sigma^k_{\mathrm{t\bar{t}}}}{\mathrm{dX}} = \frac{1}{\sum\limits_i \Delta\sigma^i_{\mathrm{t\bar{t}}}}\frac{\mathrm{d}\sigma^k_{\mathrm{t\bar{t}}}}{\mathrm{dX}} = \frac{1}{\Delta\mathrm{X}^k}\frac{N^k_{\mathrm{t\bar{t}}}}{\sum\limits_i N^i_{\mathrm{t\bar{t}}}}. \tag{6.16}$$

6.6 Systematic Uncertainties

The sources of systematic uncertainties considered in this analysis are largely the same as those of the top quark mass analysis (Section 5.6). Similarly, the errors are estimated by varying the input quantities according to their theoretical or experimental variations, and measuring the change in the final result. All systematic errors are added to the fitting and unfolding errors in quadrature. In the total error calculation, all systematic uncertainties are set to be symmetric, conservatively taking the larger of the upward and downward variations.

The full lists of systematic uncertainties for the normalised differential cross section measurement with respect $E_{\mathrm{T}}^{\mathrm{miss}}$ variable is presented in Table 6.9, and the rest of the primary variables can be found in Appendix D. Rate-changing systematic uncertainties, such as luminosity uncertainty and single top/$\mathrm{t\bar{t}}$ production cross sections, expectedly have very a low impact on all measurements. Typically, the dominating systematic uncertainties are due to the factorisation scale and matching threshold variations in both V+jets and $\mathrm{t\bar{t}}$ events (shown in Table 5.4), mainly driven by low statistics available in corresponding Monte Carlo samples. Unfortunately, this leads to substantial errors in fit results.

Jet energy scale is also amongst the dominating uncertainties due to the fact that it has a direct influence on the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement as the jet energy corrections are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$ (as explained in Section 3.5.4). H_{T} and S_{T} variables are also affected by this systematic uncertainty, since they directly depend on the jet transverse momenta.

A systematic uncertainty due to the hadronisation model is estimated by replacing the standard MadGraph $\mathrm{t\bar{t}}$ signal sample with MC@NLO signal sample. This systematic variation has a particularly strong effect on H_{T} and S_{T} variables, since they are sensitive to the hadronic activity in the event.

Differential cross section measurement with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$-related variables can be affected by the unclustered energy variation, which studies the impact of fluctuations in the calorimeter energy depositions that are not clustered in jets.

This systematic uncertainty has a moderate (≈ 1 %) effect on distributions using $E_{\mathrm{T}}^{\mathrm{miss}}$.

The $p_{\mathrm{T}}(t, \bar{t})$ reweighting systematic variation accounts for the known issue of the event generators incorrectly modelling the p_{T} spectrum of the top quarks. This issue was discovered in early differential cross section measurements [53], when it was found that the data distribution of the top quark transverse momentum is softer in the tail region than the MC prediction from most MC generators, including MadGraph, POWHEG and MC@NLO. The results were confirmed by a range of analyses studying different $\mathrm{t\bar{t}}$ decay modes. A set of scale factors were derived in order to restore the data/MC agreement. These scale factors are not applied for the central measurements in this analysis, but investigated as a systematic variation. It was found that this variation has a very moderate effect on event-level distributions, with relative changes in the differential cross section below 1 %.

6.7 Results

The template fit explained in Section 6.5.3 was performed for both electron and muon channels, yielding two distributions of signal $\mathrm{t\bar{t}}$ events. These distributions are then unfolded using the SVD unfolding method described in Section 6.5.4, and then summed together to obtain the total distribution of $\mathrm{t\bar{t}}$ events in semileptonic production mode. Finally, the normalised cross section is calculated as outlined in Section 6.5.5.

The final combined results are shown in Figures 6.18–6.22 for all primary variables. The unfolded data are compared with predictions by MadGraph+PYTHIA, POWHEG + PYTHIA and MC@NLO + HERWIG event generators. Results are also compared with theoretical predictions showing the effects due to variations of modelling parameters, particularly matching threshold and normalisation scale. To allow easier interpretation of data, additional box graphs underneath main plots show the ratios between theory predictions and data. In all figures, the inner error bars (gray bands on ratio plots) represent the combined statistical and unfolding uncertainties, whereas the outer error bars (yellow bands on ratio plots) include the uncertainty due to systematic variations added in quadrature. Naturally, the systematic errors due to matching threshold and normalisation scale choice are excluded from comparison with variations due to these parameters, and the hadronisation systematic is excluded from the comparison with different generators.

Table 6.9: Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$ variable (combination of electron and muon channels). Dominating uncertainties are emphasised in bold.

Uncertainty source	0–25 GeV	25–45 GeV	45–70 GeV	70–100 GeV	100–150 GeV	$\geq$ 150 GeV
Luminosity $+1\sigma$ (%)	-0.03	-0.01	0.01	0.02	0.01	-0.01
Luminosity -1σ (%)	0.03	0.01	-0.01	-0.02	-0.01	0.00
Single top cross section $+1\sigma$ (%)	0.00	0.00	0.01	0.00	-0.01	-0.04
Single top cross section -1σ (%)	-0.00	-0.00	-0.01	-0.00	0.01	0.04
$t\bar{t}$ cross section $+1\sigma$ (%)	-0.00	-0.00	-0.01	-0.00	0.02	0.05
$t\bar{t}$ cross section -1σ (%)	0.00	0.00	0.01	0.00	-0.02	-0.05
b-tagging efficiency $+1\sigma$ (%)	0.01	0.02	0.01	-0.02	-0.04	-0.05
b-tagging efficiency -1σ (%)	0.13	0.06	-0.05	-0.09	-0.07	-0.03
b-tagging mis-tag rate $+1\sigma$ (%)	0.04	0.03	-0.00	-0.04	-0.06	-0.06
b-tagging mis-tag rate -1σ (%)	0.09	0.04	-0.03	-0.06	-0.04	-0.01
Jet energy resolution $+1\sigma$ (%)	0.05	0.04	-0.01	-0.06	-0.07	-0.05
Jet energy resolution -1σ (%)	0.08	0.04	-0.04	-0.06	-0.04	0.01
Jet energy scale $+1\sigma$ (%)	**-1.72**	**-1.02**	**0.11**	**1.29**	**2.26**	**3.01**
Jet energy scale -1σ (%)	**1.75**	**1.11**	**-0.02**	**-1.36**	**-2.57**	**-3.47**
Pile-up $+1\sigma$ (%)	0.09	0.03	-0.03	-0.06	-0.03	0.02
Pile-up -1σ (%)	0.07	0.04	-0.02	-0.06	-0.07	-0.08
QCD shape uncertainty (%)	-0.85	-0.34	0.24	0.53	0.57	0.53
hadronisation uncertainty (%)	-1.15	0.33	0.24	-0.26	0.48	0.85
$p_{\mathrm{T}}(t,\bar{t})$ reweighting (%)	0.44	0.26	0.00	-0.27	-0.66	-1.12
$t\bar{t}$ (matching up) (%)	-0.41	-0.48	-0.56	0.31	2.65	2.06
$t\bar{t}$ (matching down) (%)	-0.57	0.01	-0.03	0.16	0.79	-0.18
$t\bar{t}$ (Q^2 up) (%)	**-1.20**	**-0.48**	**-0.27**	**1.06**	**1.61**	**2.24**
$t\bar{t}$ (Q^2 down) (%)	**1.14**	**0.58**	**-0.90**	**-0.80**	**0.57**	**0.18**
V+jets (matching up) (%)	0.80	0.65	0.08	-0.88	-1.45	-1.45
V+jets (matching down) (%)	0.30	0.18	-0.17	-0.44	0.03	0.86
V+jets (Q^2 up) (%)	**1.89**	**0.82**	**-0.56**	**-1.37**	**-1.25**	**-0.32**
V+jets (Q^2 down) (%)	**0.12**	**-0.19**	**-0.35**	**-0.03**	**0.85**	**2.02**
Electron energy $+1\sigma$ (%)	-0.12	0.02	0.12	0.02	-0.17	-0.33
Electron energy -1σ (%)	0.09	0.03	-0.08	-0.08	0.09	0.25
Muon energy $+1\sigma$ (%)	-0.02	0.02	0.04	-0.01	-0.07	-0.12
Muon energy -1σ (%)	0.01	-0.02	-0.04	0.01	0.09	0.18
Tau energy $+1\sigma$ (%)	0.04	-0.00	-0.03	-0.01	0.03	0.05
Tau energy -1σ (%)	-0.02	0.00	0.01	0.00	-0.01	-0.01
Unclustered energy $+1\sigma$ (%)	-1.10	-0.70	0.05	0.88	1.54	1.96
Unclustered energy -1σ (%)	1.00	0.67	-0.00	-0.85	-1.50	-1.86
Total (%)	3.65	2.04	1.58	2.80	4.99	5.63

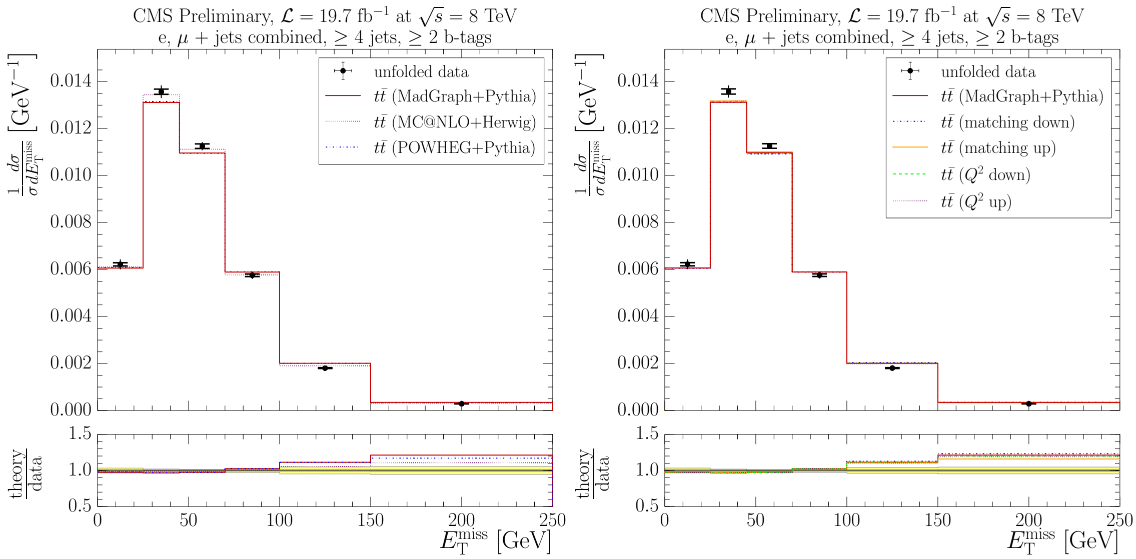


Figure 6.18: Normalised semileptonic $t\bar{t}$ differential cross section with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$. The data are compared to predictions of three different MC generators (left) and theoretical predictions showing the variation due to modelling uncertainties (right). The inner error bars represent the combined statistical and unfolding uncertainties, the outer error bars include the systematic uncertainties added in quadrature.

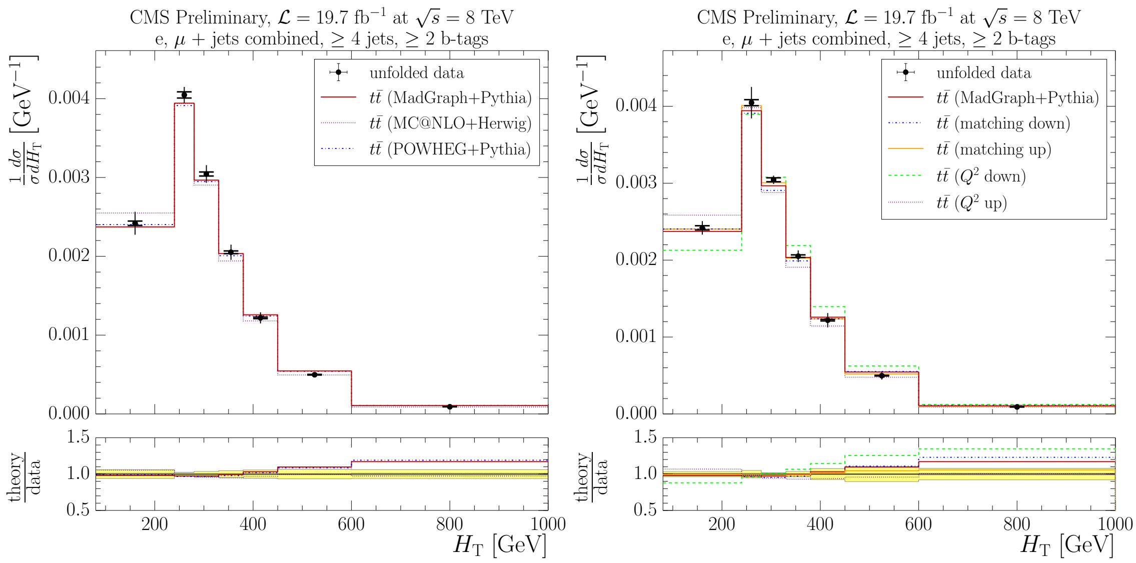


Figure 6.19: Normalised semileptonic $t\bar{t}$ differential cross section with respect to H_{T}. The data are compared to predictions of three different MC generators (left) and theoretical predictions showing the variation due to modelling uncertainties (right). The inner error bars represent the combined statistical and unfolding uncertainties, the outer error bars include the systematic uncertainties added in quadrature.

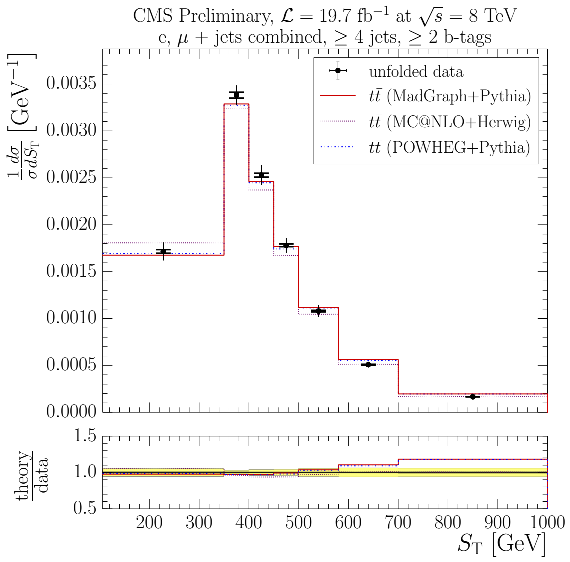


Figure 6.20: Normalised semileptonic $t\bar{t}$ differential cross section with respect to S_T. The data are compared to predictions of three different MC generators (left) and theoretical predictions showing the variation due to modelling uncertainties (right). The inner error bars represent the combined statistical and unfolding uncertainties, the outer error bars include the systematic uncertainties added in quadrature.

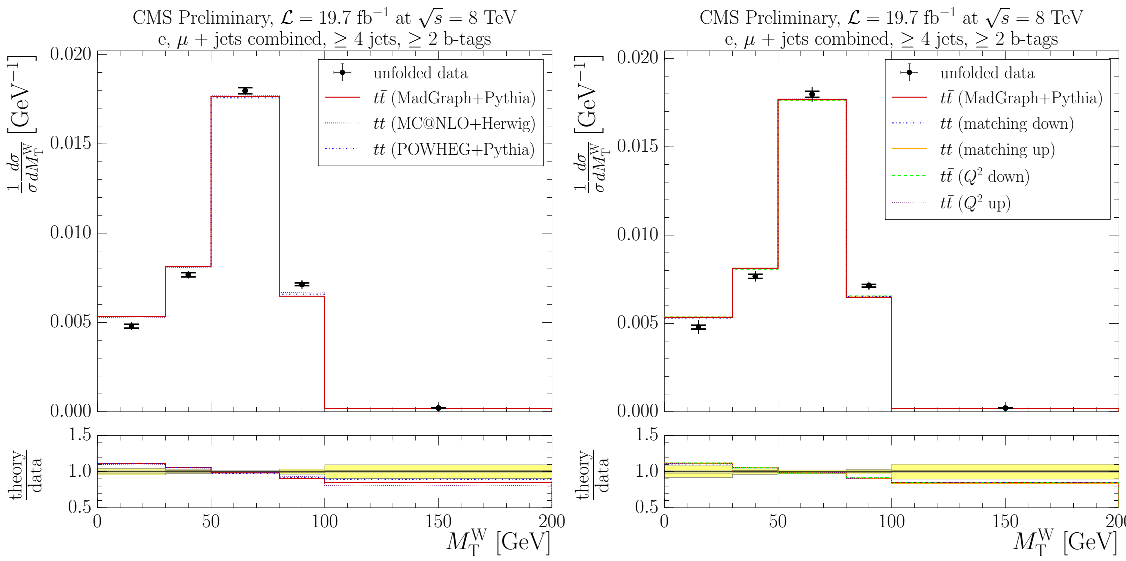


Figure 6.21: Normalised semileptonic $t\bar{t}$ differential cross section with respect to M_T^W. The data are compared to predictions of three different MC generators (left) and theoretical predictions showing the variation due to modelling uncertainties (right). The inner error bars represent the combined statistical and unfolding uncertainties, the outer error bars include the systematic uncertainties added in quadrature.

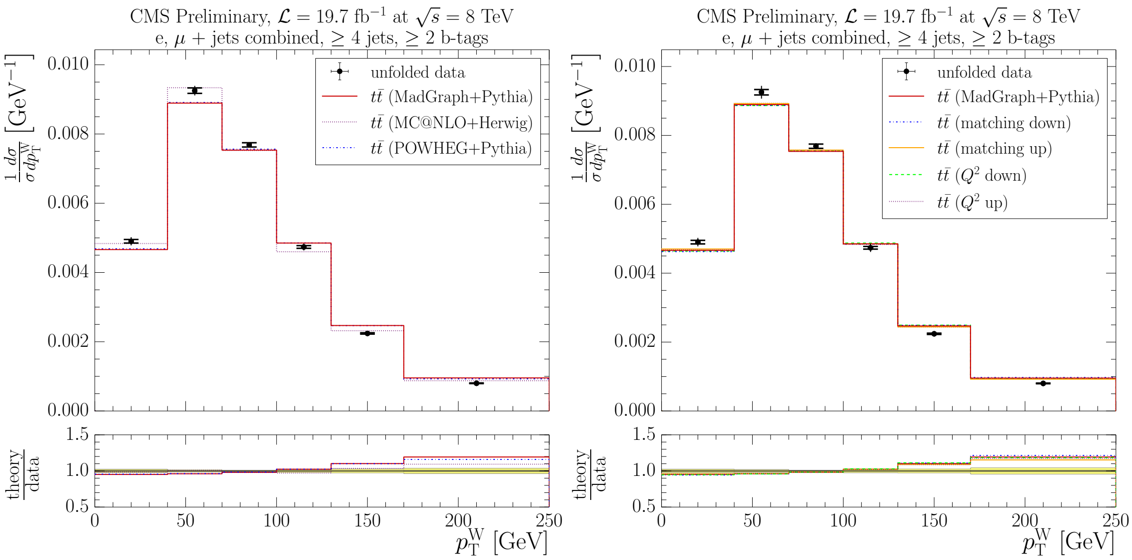


Figure 6.22: Normalised semileptonic $t\bar{t}$ differential cross section with respect to p_T^W. The data are compared to predictions of three different MC generators (left) and theoretical predictions showing the variation due to modelling uncertainties (right). The inner error bars represent the combined statistical and unfolding uncertainties, the outer error bars include the systematic uncertainties added in quadrature.

In general, it was found that the data are better described by the MC@NLO + HERWIG generator, which expectedly has a higher simulation accuracy compared to the leading order generators. For most of the primary variables, namely E_T^{miss}, S_T, H_T and p_T^W the data spectrum appears to be slightly softer than ones predicted by MadGraph + PYTHIA and POWHEG + PYTHIA simulations.

6.8 Summary

A differential cross section measurement of top quark pair production with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$ and other global variables was performed in proton-proton collisions at a centre of mass energy of $\sqrt{s} = 8$ TeV, using the full 2012 LHC dataset corresponding to integrated luminosity of 19.7 fb^{-1}, collected by the CMS experiment. The analysis selected events with a single isolated highly-energetic electron or muon, which is assumed to come from one of the W bosons in top quark and anti-quark decay. The shape of QCD multi-jet events was estimated using data-driven techniques. A template fit method was used to determine the sample composition and the number of $\mathrm{t\bar{t}}$ events, which was corrected for misidentification, detector resolution and acceptance effects using the SVD unfolding technique. The results from both semileptonic decay channels were combined and compared with different Monte Carlo generators, as well as with variations in theoretical predictions due to modelling uncertainties. No significant deviations from the Standard Model predictions were observed.

7. Conclusions

This thesis has covered the author's and collaborators' research carried out in the top quark sector of particle physics. Two major analysis contributions, namely the top quark mass measurement in the electron plus jets channel, and the $\mathrm{t\bar{t}}$ differential cross section measurement in the electron and muon plus jets channels, were presented in detail. Additionally, the service work performed by the author for the CMS collaboration – the high-level trigger development for top quark physics – was also covered. The triggers developed were used not only in the mass analysis presented, but also in other CMS top quark physics analyses using the electron plus jets $\mathrm{t\bar{t}}$ events collected at a centre of mass energy of $\sqrt{s} = 7$ TeV.

The top quark mass measurement was performed using the full 2011 dataset recorded by the CMS detector with a total integrated luminosity of approximately 5.0 fb^{-1}. The analysis served as a complementary cross-check of the main CMS top quark mass measurement in the lepton plus jets channel, which remains the most precise single measurement of the top quark mass to date. The mass extraction technique and the kinematic fit procedure, common between two analyses, were described in detail. The analyses results are consistent with each other, as well as with the latest combination result from the Tevatron and the LHC experiments.

The differential cross section measurement of top quark pair production with respect to event-level distributions is based on 2012 data collected by CMS at a centre of mass energy of $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 19.7 fb^{-1}. The analysis selected semileptonic $\mathrm{t\bar{t}}$ events with a single isolated highly-energetic electron or muon, and at least four jets in the final state. The number of $\mathrm{t\bar{t}}$ events was determined using a template fit method, and corrected for misidentification, detector resolution and acceptance effects using the SVD unfolding technique. The results from both the electron and muon plus jets channels were combined and compared with different predictions of Monte Carlo generators, as well as with a set of variations in theoretical predictions due to modelling uncertainties. In general, the results were found to be consistent with the Standard Model predictions.

The cross section measurement can be improved in a few different ways. Recent studies have shown that the binning choice for all primary variables can be enhanced by exploiting a more sophisticated fit requiring sufficient statistics and acceptable migration between bins. Also, addition of other variables in the template fit – for example, the M3 variable, i.e. the invariant mass of the three jets yielding the highest vectorial sum of their transverse momenta – can improve the differentiating power between the QCD and W/Z + jets backgrounds.

The normalised differential cross section measurement based on 7 TeV data [124] (not covered in this thesis) was only performed with respect to the missing transverse energy. Currently, work is under way to improve the 7 TeV measurement by including other event-level variables and using the latest definitions of reconstructed objects. This would allow the measurement of the ratio of 8 TeV and 7 TeV cross section results, potentially cancelling more systematic uncertainties and providing more sensitivity to deviations from the Standard Model, as well as increasing the discriminating power between MC generators. The results are due to be published later this year, upon the completion of the current analysis work. All these differential cross section measurements will be enhanced using 2015 LHC data at a centre of mass energy of 13 TeV.

The measurements shown in this thesis represent the author's contribution to the high-statistics precision measurements of top quark physics currently being made at the LHC. Such measurements provide detailed checks of the Standard Model as well as improved background estimates in searches for new physics signatures.

A. High-level triggers in 2011 data

A full list of electron triggers including their version numbers used in the top quark mass analysis (Chapter 5):

- HLT-Ele25-CaloIdVT-TrkIdT-CentralTriJet30-v1

  for run number $\leq$ 161216

- HLT-Ele25-CaloIdVT-TrkIdT-CentralTriJet30-v2

  for run number $\leq$ 163269

- HLT-Ele25-CaloIdVT-TrkIdT-TriCentralJet30-v3

  for run number $\leq$ 165969

- HLT-Ele25-CaloIdVT-CaloIsoT-TrkIdT-TrkIsoT-TriCentralJet30-v1

  for run number $\leq$ 166967

- HLT-Ele25-CaloIdVT-CaloIsoT-TrkIdT-TrkIsoT-TriCentralJet30-v2

  for run number $\leq$ 167913

- HLT-Ele25-CaloIdVT-CaloIsoT-TrkIdT-TrkIsoT-TriCentralJet30-v4

  for run number $\leq$ 173235

- HLT-Ele25-CaloIdVT-CaloIsoT-TrkIdT-TrkIsoT-TriCentralJet30-v5

  for run number $\leq$ 178380

- HLT-Ele25-CaloIdVT-CaloIsoT-TrkIdT-TrkIsoT-TriCentralPFJet30-v2

  for run number $\leq$ 179889

- HLT-Ele25-CaloIdVT-CaloIsoT-TrkIdT-TrkIsoT-TriCentralPFJet30-v3

  for run number $\leq$ 180252

B. Choice of binning for primary variables

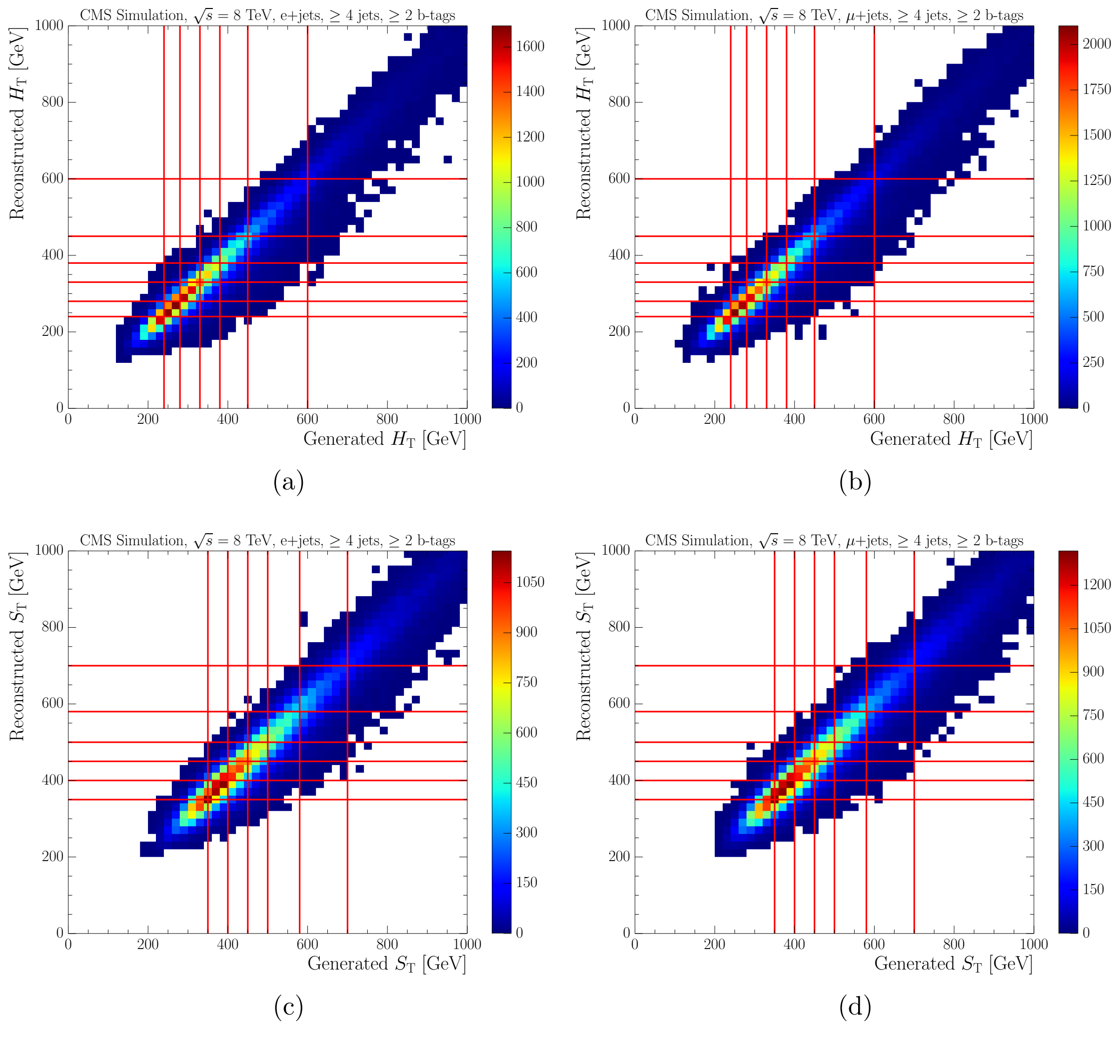


Figure B.1: Reconstructed versus generated H_T (a, b) and S_T (c, d) for electron plus jets (left) and muon plus jets events (right).

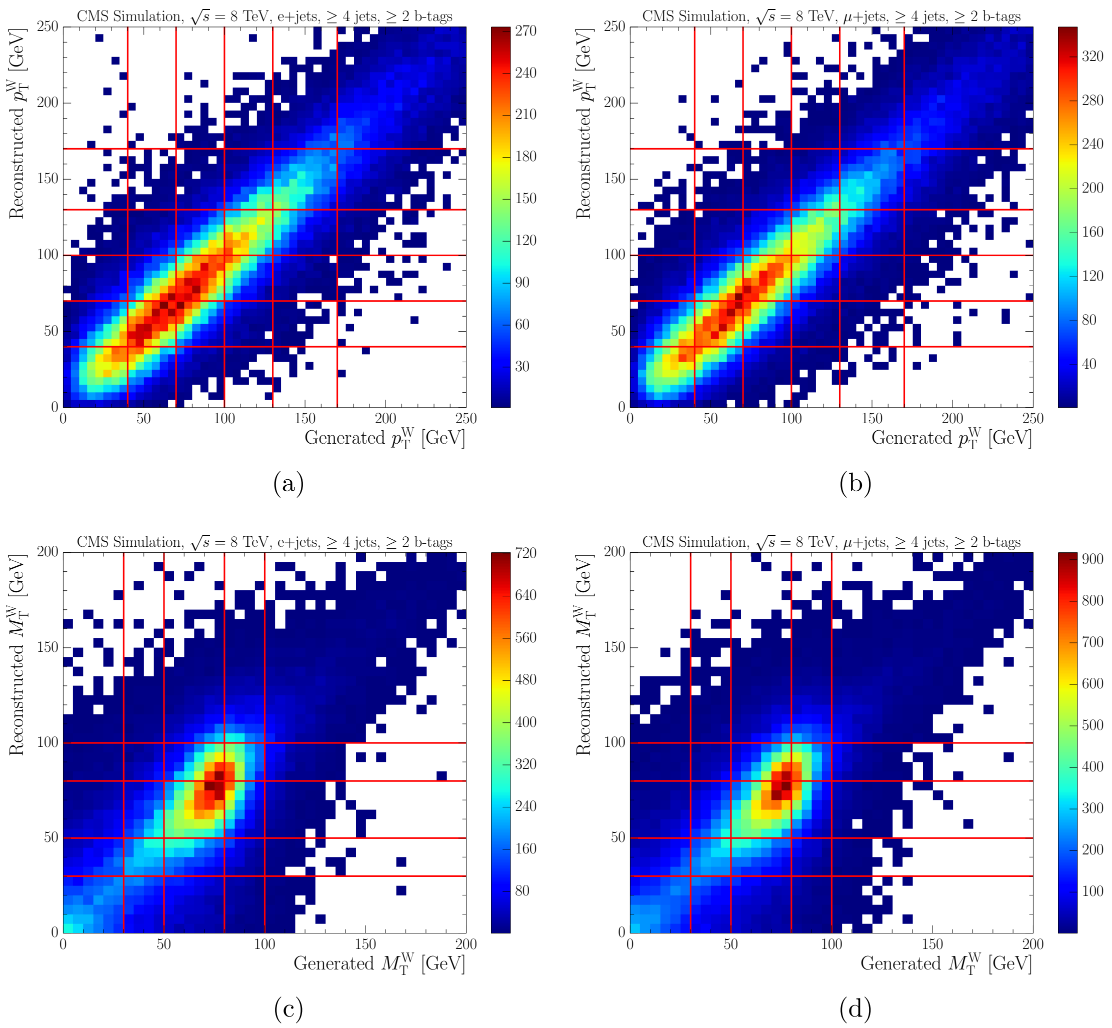


Figure B.2: Reconstructed versus generated $p_{\mathrm{T}}^{\mathrm{W}}$ (a, b) and $M_{\mathrm{T}}^{\mathrm{W}}$ (c, d) for electron plus jets (left) and muon plus jets events (right).

Table B.1: Stability and purity of H_{T} bins in the electron channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$80 < H_{\mathrm{T}} < 240$	$240 \leq H_{\mathrm{T}} < 280$	$280 \leq H_{\mathrm{T}} < 330$	$330 \leq H_{\mathrm{T}} < 380$	$380 \leq H_{\mathrm{T}} < 450$	$450 \leq H_{\mathrm{T}} < 600$	$H_{\mathrm{T}} \geq 600$
events	10601	12569	14372	12397	10893	10721	5609
purity	0.82	0.62	0.63	0.62	0.68	0.81	0.9
stability	0.76	0.63	0.65	0.64	0.68	0.8	0.9

Table B.2: Stability and purity of H_{T} bins in the muon channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$80 < H_{\mathrm{T}} < 240$	$240 \leq H_{\mathrm{T}} < 280$	$280 \leq H_{\mathrm{T}} < 330$	$330 \leq H_{\mathrm{T}} < 380$	$380 \leq H_{\mathrm{T}} < 450$	$450 \leq H_{\mathrm{T}} < 600$	$H_{\mathrm{T}} \geq 600$
events	12337	14821	16747	14268	12258	12119	5857
purity	0.82	0.63	0.64	0.63	0.67	0.81	0.9
stability	0.76	0.64	0.65	0.65	0.68	0.81	0.89

Table B.3: Stability and purity of S_{T} bins in the electron channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$106 < S_{\mathrm{T}} < 350$	$350 \leq S_{\mathrm{T}} < 400$	$400 \leq S_{\mathrm{T}} < 450$	$450 \leq S_{\mathrm{T}} < 500$	$500 \leq S_{\mathrm{T}} < 580$	$580 \leq S_{\mathrm{T}} < 700$	$S_{\mathrm{T}} \geq 700$
events	11183	13246	12073	10709	11317	9166	8373
purity	0.83	0.6	0.53	0.53	0.63	0.71	0.87
stability	0.73	0.6	0.55	0.55	0.66	0.73	0.91

Table B.4: Stability and purity of S_{T} bins in the muon channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$106 < S_{\mathrm{T}} < 350$	$350 \leq S_{\mathrm{T}} < 400$	$400 \leq S_{\mathrm{T}} < 450$	$450 \leq S_{\mathrm{T}} < 500$	$500 \leq S_{\mathrm{T}} < 580$	$580 \leq S_{\mathrm{T}} < 700$	$S_{\mathrm{T}} \geq 700$
events	13593	15717	13878	12307	12732	10135	8875
purity	0.83	0.61	0.54	0.54	0.64	0.71	0.87
stability	0.74	0.61	0.55	0.56	0.66	0.74	0.9

Table B.5: Stability and purity of $p_{\mathrm{T}}^{\mathrm{W}}$ bins in the electron channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$0 < p_{\mathrm{T}}^{\mathrm{W}} < 40$	$40 \leq p_{\mathrm{T}}^{\mathrm{W}} < 70$	$70 \leq p_{\mathrm{T}}^{\mathrm{W}} < 100$	$100 \leq p_{\mathrm{T}}^{\mathrm{W}} < 130$	$130 \leq p_{\mathrm{T}}^{\mathrm{W}} < 170$	$p_{\mathrm{T}}^{\mathrm{W}} \geq 170$
events	10005	15849	16053	12203	9532	7717
purity	0.64	0.54	0.52	0.5	0.56	0.77
stability	0.63	0.54	0.52	0.51	0.57	0.76

Table B.6: Stability and purity of $p_{\mathrm{T}}^{\mathrm{W}}$ bins in the muon channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$0 < p_{\mathrm{T}}^{\mathrm{W}} < 40$	$40 \leq p_{\mathrm{T}}^{\mathrm{W}} < 70$	$70 \leq p_{\mathrm{T}}^{\mathrm{W}} < 100$	$100 \leq p_{\mathrm{T}}^{\mathrm{W}} < 130$	$130 \leq p_{\mathrm{T}}^{\mathrm{W}} < 170$	$p_{\mathrm{T}}^{\mathrm{W}} \geq 170$
events	12379	18821	18330	13495	10260	8394
purity	0.67	0.55	0.52	0.5	0.55	0.76
stability	0.63	0.55	0.53	0.51	0.56	0.76

Table B.7: Stability and purity of $M_{\mathrm{T}}^{\mathrm{W}}$ bins in the electron channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$0 < M_{\mathrm{T}}^{\mathrm{W}} < 30$	$30 \leq M_{\mathrm{T}}^{\mathrm{W}} < 50$	$50 \leq M_{\mathrm{T}}^{\mathrm{W}} < 80$	$80 \leq M_{\mathrm{T}}^{\mathrm{W}} < 100$	$MT \geq 100$
events	11509	11633	26288	14824	8921
purity	0.57	0.35	0.65	0.4	0.36
stability	0.62	0.37	0.53	0.41	0.67

Table B.8: Stability and purity of $M_{\mathrm{T}}^{\mathrm{W}}$ bins in the muon channel for $\mathrm{t\bar{t}}$ MC events

bin, GeV	$0 < M_{\mathrm{T}}^{\mathrm{W}} < 30$	$30 \leq M_{\mathrm{T}}^{\mathrm{W}} < 50$	$50 \leq M_{\mathrm{T}}^{\mathrm{W}} < 80$	$80 \leq M_{\mathrm{T}}^{\mathrm{W}} < 100$	$MT \geq 100$
events	13209	13636	30669	16974	9179.3
purity	0.56	0.36	0.66	0.42	0.37
stability	0.64	0.38	0.54	0.42	0.66

C. Fitting templates

H_{T} variable

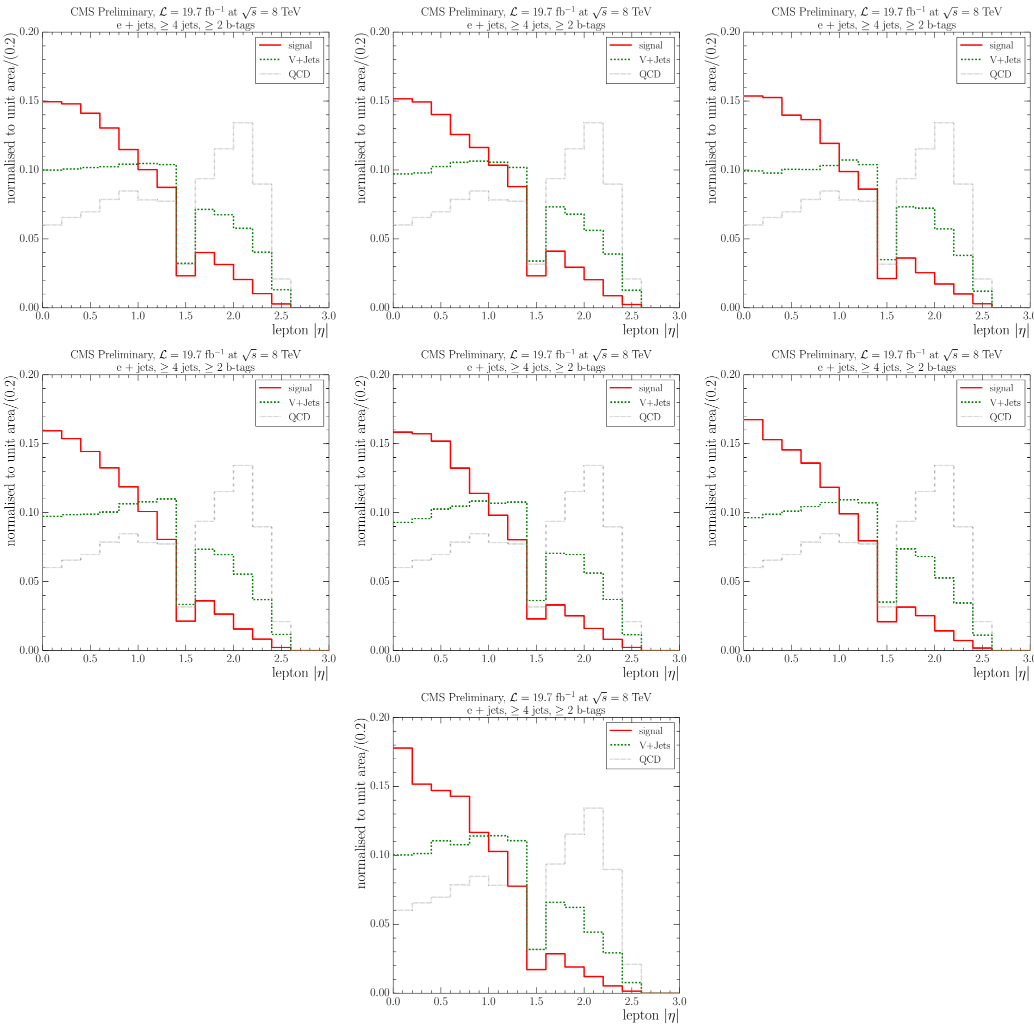


Figure C.1: Electron $|\eta|$ templates for the fit in different bins of H_{T}, from top left to bottom: 0–240 GeV, 240–280 GeV, 280–330 GeV, 330–380 GeV, 380–450 GeV, 450–600 GeV and ≥ 600 GeV.

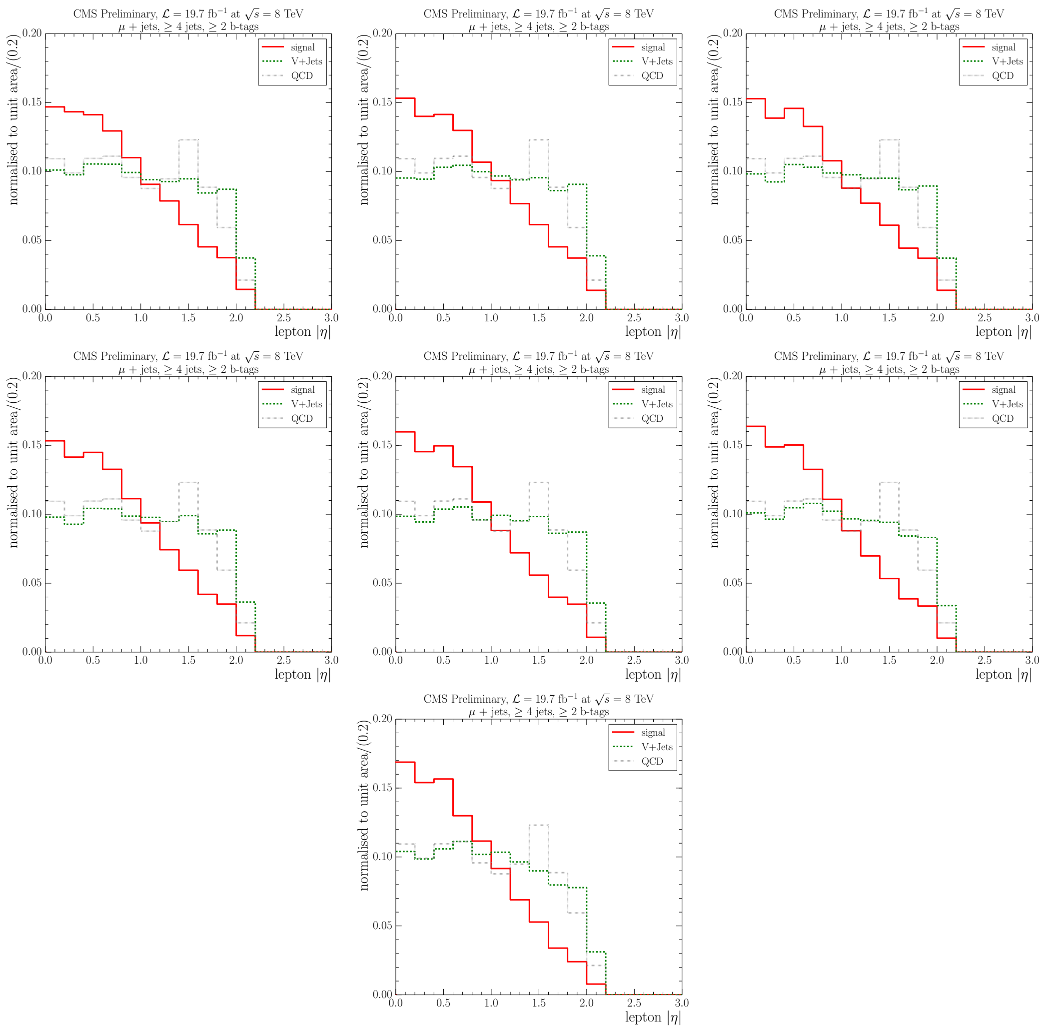


Figure C.2: Muon $|\eta|$ templates for the fit in different bins of H_{T}, from top left to bottom: 0–240 GeV, 240–280 GeV, 280–330 GeV, 330–380 GeV, 380–450 GeV, 450–600 GeV and $\geq$ 600 GeV.

S_T variable

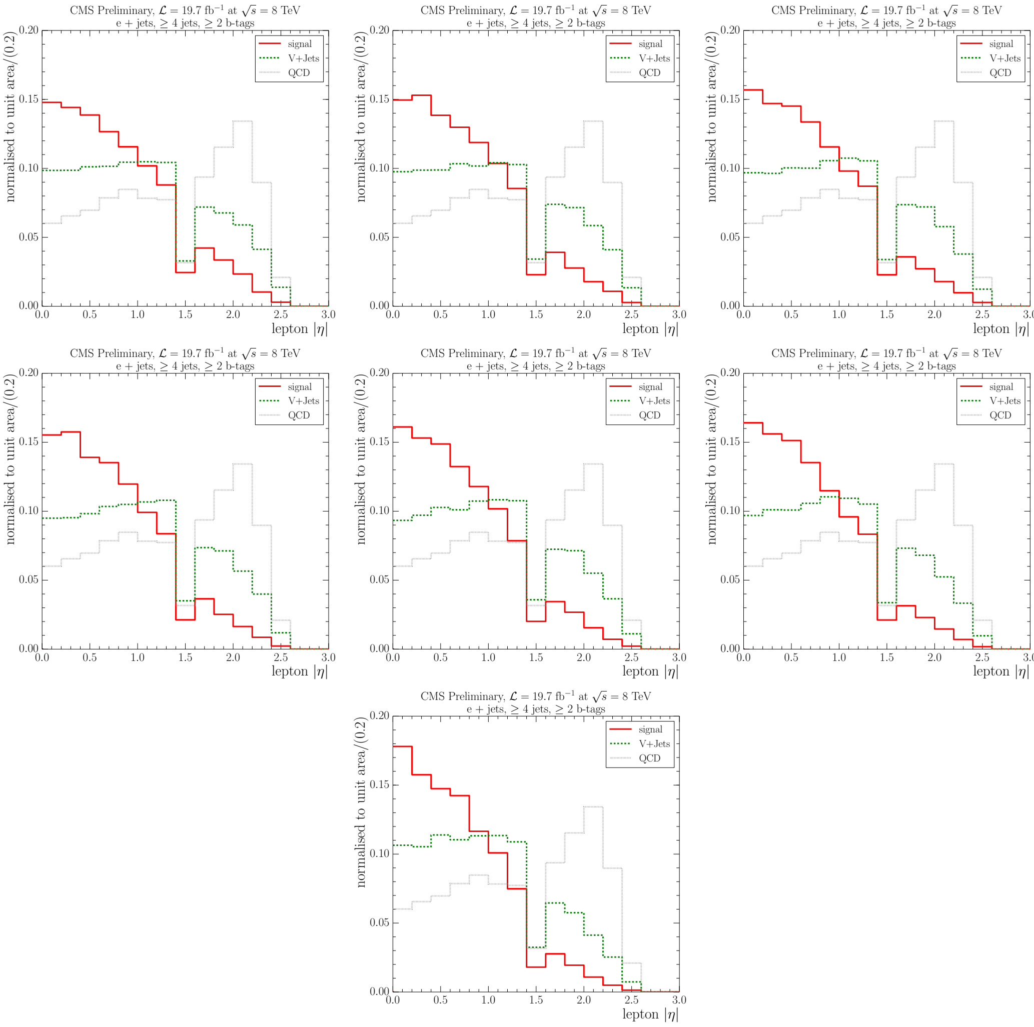


Figure C.3: Electron $|\eta|$ templates for the fit in different bins of S_T, from top left to bottom: 0–350 GeV, 350–400 GeV, 400–450 GeV, 450–500 GeV, 500–580 GeV, 580–700 GeV and $\geq$ 700 GeV.

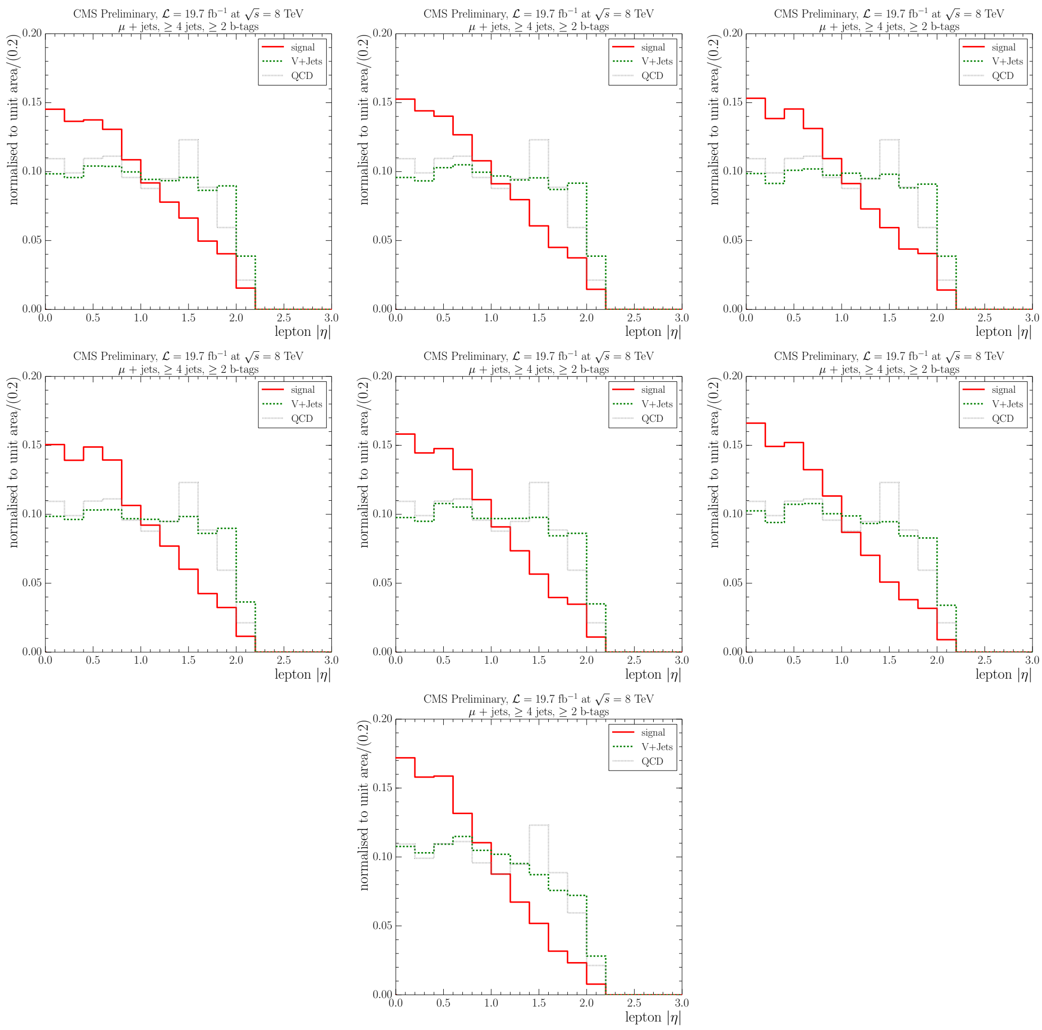


Figure C.4: Muon $|\eta|$ templates for the fit in different bins of S_{T}, from top left to bottom: 0–350 GeV, 350–400 GeV, 400–450 GeV, 450–500 GeV, 500–580 GeV, 580–700 GeV and $\geq$ 700 GeV.

p_T^W variable

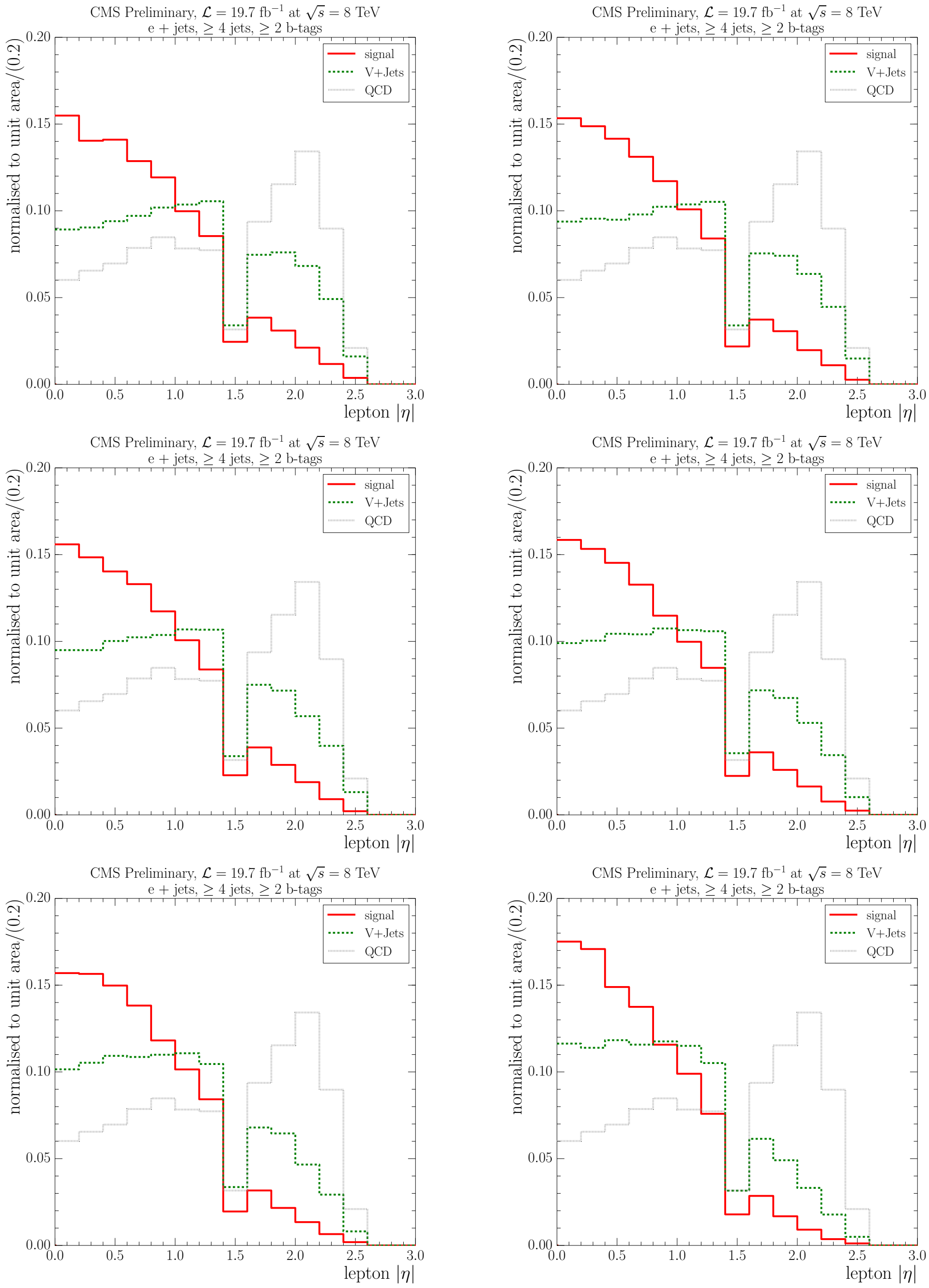


Figure C.5: Electron $|\eta|$ templates for the fit in different bins of p_T^W, from top left to bottom right: 0–40 GeV, 40–70 GeV, 70–100 GeV, 100–130 GeV, 130–170 GeV and ≥ 170 GeV.

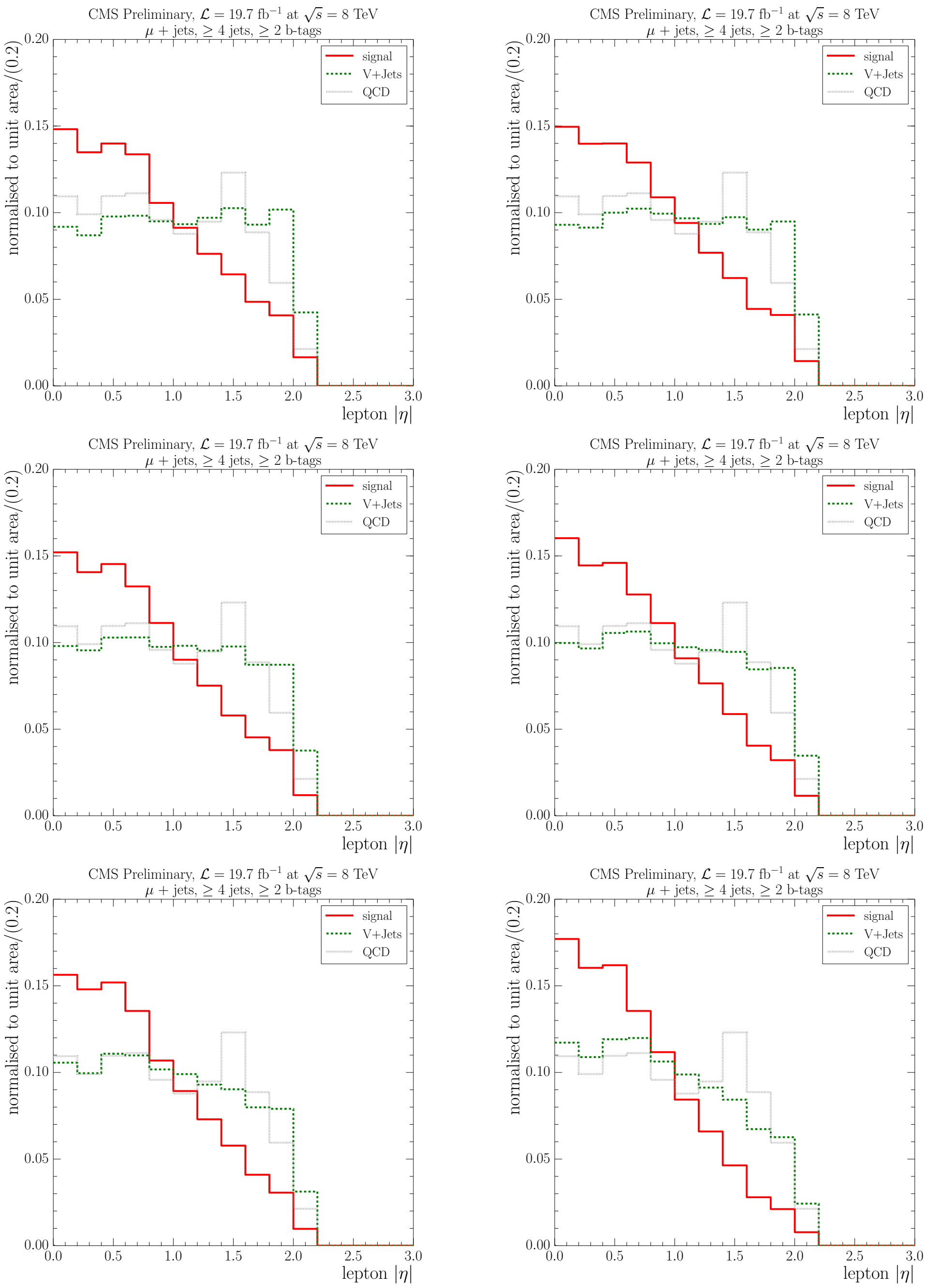


Figure C.6: Muon $|\eta|$ templates for the fit in different bins of $p_{\mathrm{T}}^{\mathrm{W}}$, from top left to bottom right: 0–40 GeV, 40–70 GeV, 70–100 GeV, 100–130 GeV, 130–170 GeV and ≥ 170 GeV.

$M_{\mathrm{T}}^{\mathrm{W}}$ variable

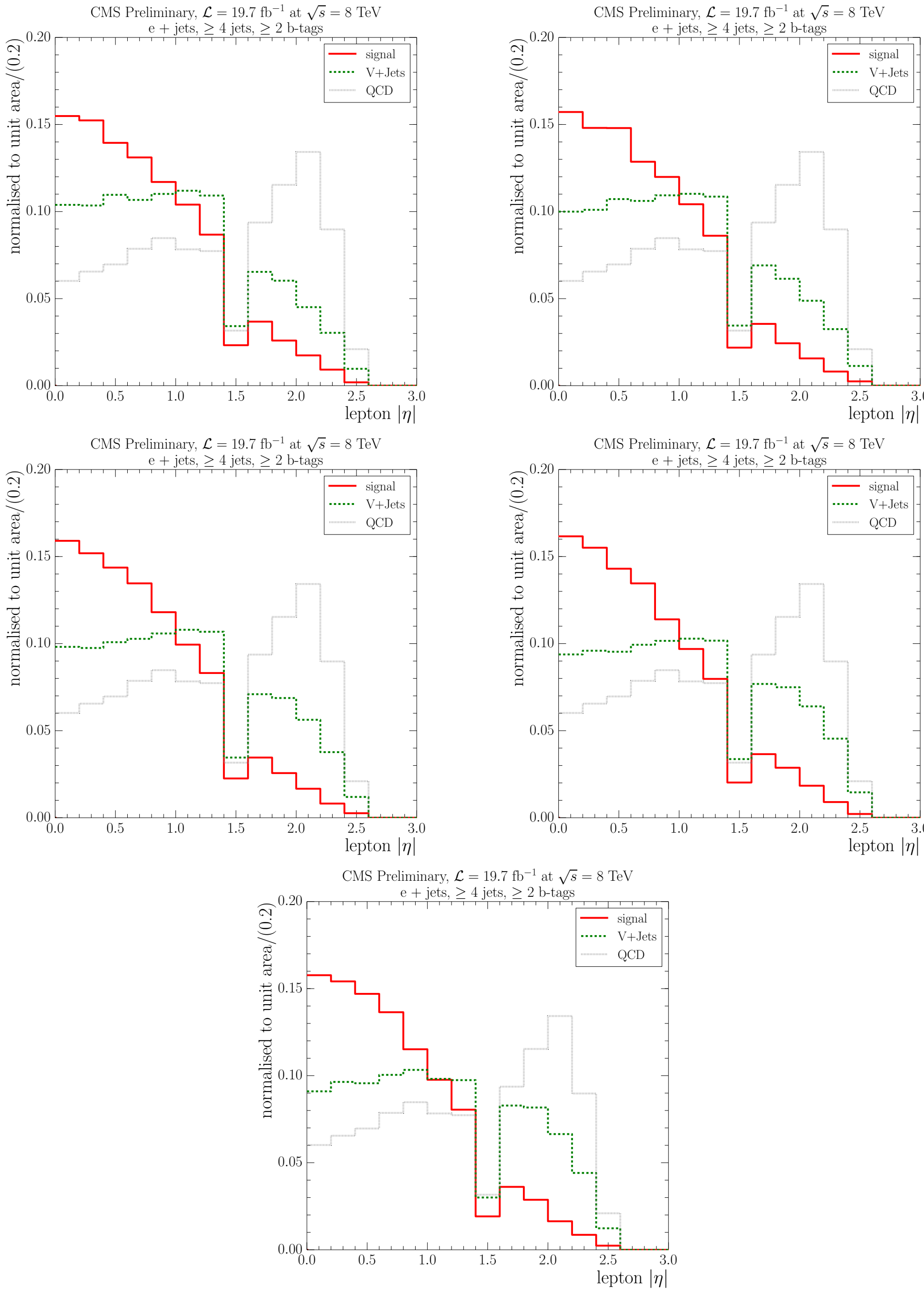


Figure C.7: Electron $|\eta|$ templates for the fit in different bins of $M_{\mathrm{T}}^{\mathrm{W}}$, from top left to bottom: 0–30 GeV, 30–50 GeV, 50–80 GeV, 80–100 GeV and ≥ 100 GeV.

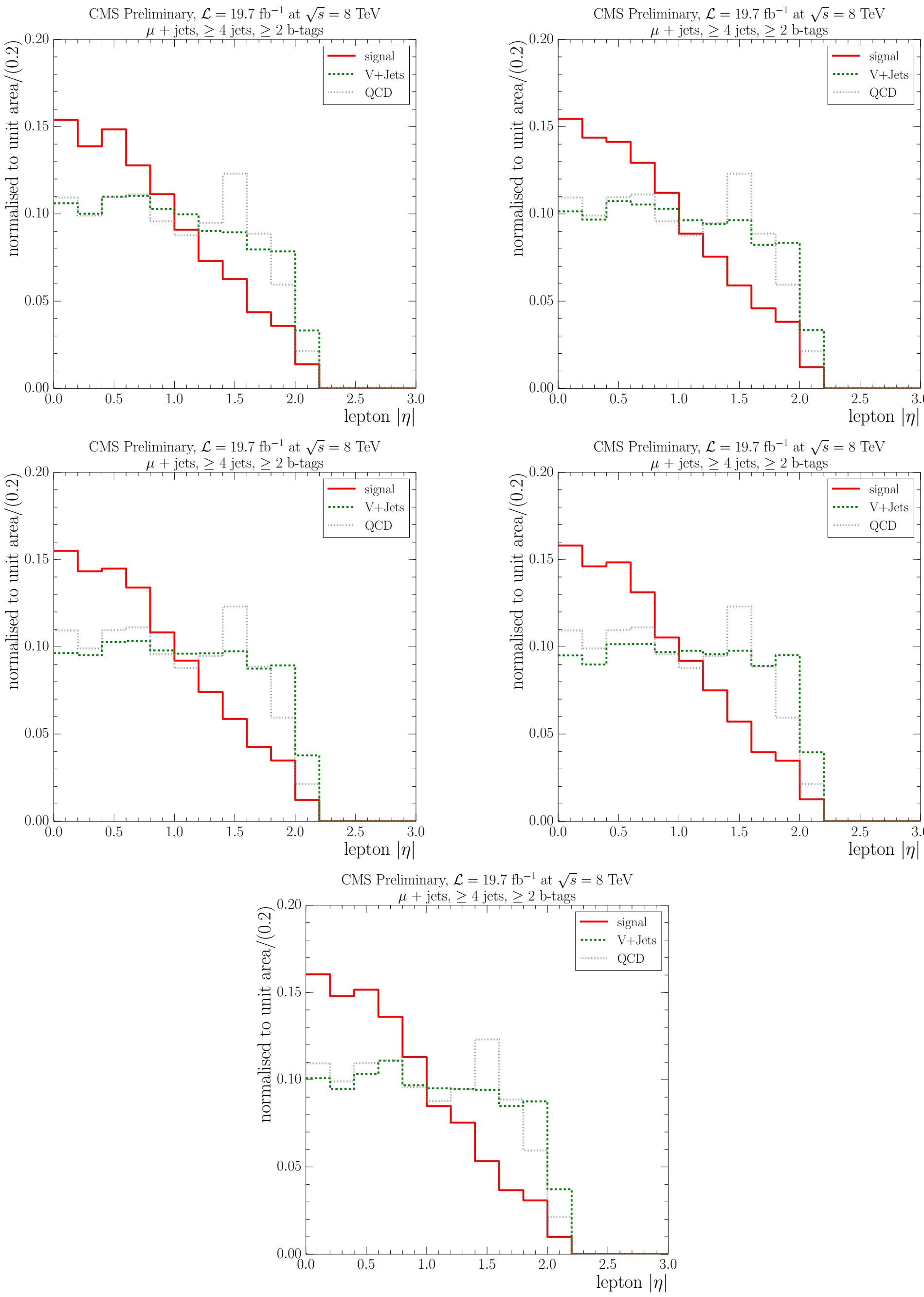


Figure C.8: Muon $|\eta|$ templates for the fit in different bins of $M_{\mathrm{T}}^{\mathrm{W}}$, from top left to bottom: 0–30 GeV, 30–50 GeV, 50–80 GeV, 80–100 GeV and ≥ 100 GeV.

D. Systematic uncertainties of the cross section measurement

Table D.1: Systematic uncertainties for the normalised $t\bar{t}$ cross section measurement with respect to H_T variable (combination of electron and muon channels). Dominating uncertainties are emphasised in bold.

Systematic	80–240 GeV	240–280 GeV	280–330 GeV	330–380 GeV	380–450 GeV	450–600 GeV	$\geq$ 600 GeV
Luminosity $+1\sigma$ (%)	-0.00	0.00	0.01	0.01	0.00	-0.01	-0.03
Luminosity -1σ (%)	0.00	-0.00	-0.01	-0.01	-0.00	0.01	0.03
Single top cross section $+1\sigma$ (%)	-0.01	-0.00	0.01	0.01	0.01	0.00	-0.00
Single top cross section -1σ (%)	0.01	0.00	-0.01	-0.01	-0.01	-0.00	0.01
$t\bar{t}$ cross section $+1\sigma$ (%)	0.01	0.00	-0.01	-0.01	-0.01	-0.00	0.00
$t\bar{t}$ cross section -1σ (%)	-0.01	-0.00	0.01	0.01	0.01	0.00	-0.01
b-tagging efficiency $+1\sigma$ (%)	0.00	0.02	0.03	0.01	-0.01	-0.06	-0.10
b-tagging efficiency -1σ (%)	0.02	-0.00	-0.02	-0.04	-0.03	0.01	0.05
b-tagging mis-tag rate $+1\sigma$ (%)	0.01	0.01	0.02	0.01	-0.02	-0.06	-0.09
b-tagging mis-tag rate -1σ (%)	0.01	-0.00	-0.02	-0.03	-0.02	0.01	0.04
Jet energy resolution $+1\sigma$ (%)	0.04	-0.00	-0.05	-0.04	-0.03	-0.03	-0.00
Jet energy resolution -1σ (%)	-0.03	-0.00	0.03	0.05	0.04	0.02	-0.03
Jet energy scale $+1\sigma$ (%)	**-1.39**	**-0.54**	**0.48**	**1.29**	**1.84**	**2.27**	**2.67**
Jet energy scale -1σ (%)	**2.74**	**1.04**	**-1.10**	**-2.67**	**-3.60**	**-4.22**	**-4.65**
Pile-up $+1\sigma$ (%)	0.06	0.02	-0.02	-0.07	-0.10	-0.08	-0.04
Pile-up -1σ (%)	-0.01	-0.01	0.01	0.03	0.03	0.00	-0.03
QCD shape uncertainty (%)	-0.26	-0.09	0.13	0.31	0.38	0.32	0.21
hadronisation uncertainty (%)	**-2.00**	**-4.84**	**-1.11**	**2.30**	**6.58**	**9.39**	**6.29**
$p_T(t,\bar{t})$ reweight (%)	0.12	0.05	-0.02	-0.08	-0.15	-0.24	-0.34
$t\bar{t}$ (matching up) (%)	-0.05	0.18	0.13	-0.32	0.48	0.38	-1.87
$t\bar{t}$ (matching down) (%)	**2.74**	**0.26**	**-1.42**	**-2.39**	**-3.53**	**-3.51**	**-2.16**
$t\bar{t}$ (Q^2 up) (%)	**3.94**	**-1.65**	**-2.70**	**-2.76**	**-2.55**	**-3.39**	**-2.61**
$t\bar{t}$ (Q^2 down) (%)	**-5.20**	**2.03**	**3.39**	**3.72**	**4.01**	**4.47**	**2.99**
V+jets (matching up) (%)	0.26	0.07	-0.12	-0.29	-0.29	-0.31	-0.48
V+jets (matching down) (%)	-0.43	-0.20	0.11	0.24	0.38	0.97	1.39
V+jets (Q^2 up) (%)	0.44	0.23	-0.16	-0.60	-0.80	-0.59	-0.28
V+jets (Q^2 down) (%)	-0.53	-0.35	0.13	0.70	0.99	0.80	0.79
Electron energy $+1\sigma$ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Electron energy -1σ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muon energy $+1\sigma$ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muon energy -1σ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tau energy $+1\sigma$ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tau energy -1σ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unclustered energy $+1\sigma$ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unclustered energy -1σ (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total (%)	6.34	5.45	3.84	5.24	8.76	11.45	8.83

Table D.2: Systematic uncertainties for the normalised $t\bar{t}$ cross section measurement with respect to S_T variable (combination of electron and muon channels). Dominating uncertainties are emphasised in bold.

Systematic	106–350 GeV	350–400 GeV	400–450 GeV	450–500 GeV	500–580 GeV	580–700 GeV	$\geq$ 700 GeV
Luminosity $+1\sigma$ (%)	-0.01	0.00	0.01	0.01	0.01	0.00	-0.01
Luminosity -1σ (%)	0.01	-0.00	-0.01	-0.01	-0.01	-0.00	0.01
Single top cross section $+1\sigma$ (%)	-0.00	0.00	0.01	0.01	0.00	-0.01	-0.02
Single top cross section -1σ (%)	0.00	-0.00	-0.01	-0.01	-0.00	0.01	0.02
$t\bar{t}$ cross section $+1\sigma$ (%)	0.00	-0.00	-0.01	-0.01	-0.00	0.01	0.02
$t\bar{t}$ cross section -1σ (%)	-0.00	0.00	0.01	0.01	0.00	-0.01	-0.03
b-tagging efficiency $+1\sigma$ (%)	0.00	0.01	0.02	0.01	-0.01	-0.04	-0.07
b-tagging efficiency -1σ (%)	0.03	-0.00	-0.03	-0.05	-0.04	-0.02	0.02
b-tagging mis-tag rate $+1\sigma$ (%)	0.00	0.01	0.02	0.02	-0.01	-0.04	-0.06
b-tagging mis-tag rate -1σ (%)	0.03	0.00	-0.03	-0.04	-0.04	-0.03	0.00
Jet energy resolution $+1\sigma$ (%)	0.09	0.02	-0.08	-0.13	-0.12	-0.09	-0.04
Jet energy resolution -1σ (%)	-0.04	0.01	0.03	0.03	0.04	0.03	0.02
Jet energy scale $+1\sigma$ (%)	**-1.12**	**-0.42**	**0.35**	**0.97**	**1.56**	**2.23**	**2.77**
Jet energy scale -1σ (%)	**2.18**	**0.77**	**-1.03**	**-2.38**	**-3.20**	**-3.71**	**-4.02**
Pile-up $+1\sigma$ (%)	0.06	0.01	-0.04	-0.08	-0.10	-0.09	-0.04
Pile-up -1σ (%)	-0.00	0.00	0.01	0.01	0.01	-0.02	-0.05
QCD shape uncertainty (%)	-0.23	-0.03	0.21	0.34	0.33	0.21	0.04
hadronisation uncertainty (%)	**-0.75**	**-5.16**	**-1.90**	**1.56**	**5.50**	**8.35**	**6.14**
$p_T(t,\bar{t})$ reweighting (%)	0.11	0.09	0.02	-0.09	-0.20	-0.30	-0.40
$t\bar{t}$ (matching down) (%)	**2.06**	**-0.22**	**-1.34**	**-1.88**	**-2.47**	**-2.39**	**-2.59**
$t\bar{t}$ (matching up) (%)	-0.29	0.20	0.58	-0.05	0.66	0.68	-1.57
$t\bar{t}$ (Q^2 down) (%)	**-5.02**	**2.83**	**3.98**	**3.42**	**4.58**	**4.66**	**2.73**
$t\bar{t}$ (Q^2 up) (%)	**3.39**	**-1.86**	**-2.59**	**-3.05**	**-2.18**	**-2.83**	**-2.89**
V+jets (matching down) (%)	-0.14	-0.08	-0.07	-0.10	0.07	0.54	1.01
V+jets (matching up) (%)	0.60	0.12	-0.43	-0.72	-0.75	-0.77	-0.86
V+jets (Q^2 down) (%)	-0.38	-0.36	0.01	0.55	0.80	0.88	0.96
V+jets (Q^2 up) (%)	0.53	0.19	-0.24	-0.57	-0.84	-1.00	-0.82
Electron energy -1σ (%)	0.02	-0.00	-0.04	-0.04	-0.02	0.01	0.01
Electron energy $+1\sigma$ (%)	0.01	0.02	0.01	0.00	-0.03	-0.07	-0.09
Muon energy -1σ (%)	-0.01	-0.01	-0.01	0.00	0.02	0.03	0.03
Muon energy $+1\sigma$ (%)	0.02	0.00	-0.01	-0.02	-0.03	-0.02	-0.01
Tau energy -1σ (%)	0.00	-0.00	-0.01	-0.01	-0.00	0.01	0.02
Tau energy $+1\sigma$ (%)	0.00	-0.00	-0.00	-0.01	-0.00	0.00	0.01
Unclustered energy -1σ (%)	0.20	0.09	-0.07	-0.23	-0.34	-0.37	-0.35
Unclustered energy $+1\sigma$ (%)	-0.20	-0.09	0.06	0.22	0.34	0.40	0.39
Total (%)	5.70	6.02	4.67	4.75	8.00	10.40	8.62

Table D.3: Systematic uncertainties for the normalised $t\bar{t}$ cross section measurement with respect to $M_{\mathrm{T}}^{\mathrm{W}}$ variable (combination of electron and muon channels). Dominating uncertainties are emphasised in bold.

Systematic	0–30 GeV	30–50 GeV	50–80 GeV	80–100 GeV	$\geq$ 100 GeV
Luminosity $+1\sigma$ (%)	-0.03	-0.02	0.00	0.03	0.04
Luminosity -1σ (%)	0.03	0.02	-0.00	-0.03	-0.05
Single top cross section $+1\sigma$ (%)	0.01	0.00	-0.00	-0.01	-0.01
Single top cross section -1σ (%)	-0.01	-0.00	0.00	0.01	0.01
$t\bar{t}$ cross section $+1\sigma$ (%)	-0.01	-0.00	0.00	0.01	0.01
$t\bar{t}$ cross section -1σ (%)	0.01	0.00	-0.00	-0.01	-0.01
b-tagging efficiency $+1\sigma$ (%)	-0.07	-0.03	0.02	0.04	0.05
b-tagging efficiency -1σ (%)	0.08	0.06	-0.01	-0.09	-0.16
b-tagging mis-tag rate $+1\sigma$ (%)	-0.05	-0.02	0.01	0.03	0.03
b-tagging mis-tag rate -1σ (%)	0.06	0.05	-0.00	-0.08	-0.14
Jet energy resolution $+1\sigma$ (%)	-0.09	-0.01	0.03	-0.02	-0.08
Jet energy resolution -1σ (%)	-0.08	-0.02	0.02	0.02	0.04
Jet energy scale $+1\sigma$ (%)	**1.41**	**0.72**	**-0.26**	**-0.98**	**-1.43**
Jet energy scale -1σ (%)	**-1.68**	**-0.79**	**0.30**	**1.15**	**1.78**
Pile-up -1σ (%)	0.13	0.07	-0.03	-0.09	-0.08
Pile-up $+1\sigma$ (%)	-0.05	-0.01	0.02	-0.01	-0.09
QCD shape uncertainty (%)	-0.25	-0.18	0.02	0.31	0.50
hadronisation uncertainty (%)	**7.07**	**2.98**	**-2.00**	**-1.44**	**-9.13**
$p_{\mathrm{T}}(t,\bar{t})$ reweighting (%)	0.66	0.29	-0.14	-0.39	-0.48
$t\bar{t}$ (matching up) (%)	0.04	-0.50	0.02	0.62	-1.27
$t\bar{t}$ (matching down) (%)	-0.34	0.09	0.20	-0.20	-2.24
$t\bar{t}$ (Q^2 up) (%)	**-1.01**	**-1.14**	**-0.03**	**1.78**	**3.91**
$t\bar{t}$ (Q^2 down) (%)	**1.88**	**0.83**	**0.11**	**-2.07**	**-7.81**
V+jets (matching up) (%)	0.17	-0.05	-0.05	0.07	0.08
V+jets (matching down) (%)	**-2.01**	**-1.14**	**0.38**	**1.52**	**1.97**
V+jets (Q^2 up) (%)	0.20	0.43	0.27	-1.25	-2.90
V+jets (Q^2 down) (%)	**1.23**	**0.65**	**-0.30**	**-0.71**	**-0.73**
Electron energy $+1\sigma$ (%)	-1.47	-0.87	0.16	1.43	2.50
Electron energy -1σ (%)	1.33	0.79	-0.17	-1.26	-2.07
Muon energy $+1\sigma$ (%)	-0.48	-0.27	0.08	0.40	0.58
Muon energy -1σ (%)	0.54	0.30	-0.08	-0.46	-0.74
Tau energy $+1\sigma$ (%)	0.02	0.00	-0.01	0.00	0.01
Tau energy -1σ (%)	-0.06	-0.02	0.01	0.05	0.10
Unclustered energy $+1\sigma$ (%)	1.25	0.61	-0.28	-0.73	-0.78
Unclustered energy -1σ (%)	-1.55	-0.79	0.32	1.02	1.24
Total (%)	8.42	4.03	2.33	3.83	13.18

Table D.4: Systematic uncertainties for the normalised $\mathrm{t\bar{t}}$ cross section measurement with respect to $p_{\mathrm{T}}^{\mathrm{W}}$ variable (combination of electron and muon channels). Dominating uncertainties are emphasised in bold.

Systematic	0–40 GeV	40–70 GeV	70–100 GeV	100–130 GeV	130–170 GeV	$\geq$ 170 GeV
Luminosity $+1\sigma$ (%)	-0.02	-0.01	0.01	0.02	0.01	-0.01
Luminosity -1σ (%)	0.02	0.01	-0.01	-0.02	-0.01	0.01
Single top cross section $+1\sigma$ (%)	0.01	0.01	0.01	-0.00	-0.02	-0.05
Single top cross section -1σ (%)	-0.01	-0.01	-0.01	0.00	0.02	0.05
$t\bar{t}$ cross section $+1\sigma$ (%)	-0.01	-0.01	-0.01	0.00	0.03	0.06
$t\bar{t}$ cross section -1σ (%)	0.01	0.01	0.01	-0.01	-0.03	-0.06
b-tagging efficiency $+1\sigma$ (%)	0.02	-0.00	-0.01	0.01	-0.00	-0.04
b-tagging efficiency -1σ (%)	0.05	0.02	-0.03	-0.06	-0.04	0.01
b-tagging mis-tag rate $+1\sigma$ (%)	0.01	-0.00	-0.00	0.01	-0.00	-0.04
b-tagging mis-tag rate -1σ (%)	0.06	0.02	-0.03	-0.06	-0.03	0.02
Jet energy resolution $+1\sigma$ (%)	0.06	0.02	-0.03	-0.04	-0.02	-0.00
Jet energy resolution -1σ (%)	-0.00	-0.03	-0.03	0.00	0.07	0.14
Jet energy scale $+1\sigma$ (%)	**-1.23**	**-0.74**	**0.09**	**0.97**	**1.66**	**2.15**
Jet energy scale -1σ (%)	**1.13**	**0.75**	**0.04**	**-0.86**	**-1.74**	**-2.49**
Pile-up $+1\sigma$ (%)	0.11	0.04	-0.04	-0.07	-0.08	-0.10
Pile-up -1σ (%)	-0.02	-0.01	0.00	0.00	0.02	0.07
QCD shape uncertainty (%)	-0.42	-0.24	0.09	0.38	0.48	0.46
hadronisation uncertainty (%)	**-1.28**	**1.46**	**-0.53**	**-1.99**	**1.06**	**2.41**
$p_{\mathrm{T}}(t,\bar{t})$ reweighting (%)	0.24	0.19	0.05	-0.17	-0.45	-0.72
$t\bar{t}$ (matching up) (%)	0.49	-0.14	-0.09	-0.21	0.15	-0.32
$t\bar{t}$ (matching down) (%)	0.15	0.50	0.27	-0.82	-0.82	-0.63
$t\bar{t}$ (Q^2 up) (%)	0.43	0.19	0.04	-0.61	-0.74	0.12
$t\bar{t}$ (Q^2 down) (%)	0.93	0.34	-0.34	-0.95	-0.14	-0.80
V+jets (matching up) (%)	0.24	0.12	0.01	-0.08	-0.28	-0.72
V+jets (matching down) (%)	**1.06**	**0.36**	**-0.35**	**-0.73**	**-0.84**	**-0.75**
V+jets (Q^2 up) (%)	-0.04	0.25	0.24	-0.23	-0.53	-0.55
V+jets (Q^2 down) (%)	-0.62	-0.16	0.32	0.46	0.24	0.07
Electron energy $+1\sigma$ (%)	0.27	0.16	0.01	-0.16	-0.39	-0.65
Electron energy -1σ (%)	-0.36	-0.16	0.06	0.23	0.39	0.53
Muon energy $+1\sigma$ (%)	0.10	0.07	0.01	-0.07	-0.16	-0.26
Muon energy -1σ (%)	-0.09	-0.05	-0.00	0.04	0.12	0.23
Tau energy $+1\sigma$ (%)	0.00	-0.01	-0.02	-0.00	0.03	0.06
Tau energy -1σ (%)	0.01	0.00	0.01	0.00	-0.02	-0.03
Unclustered energy $+1\sigma$ (%)	-0.92	-0.55	0.04	0.67	1.26	1.79
Unclustered energy -1σ (%)	0.82	0.51	-0.00	-0.59	-1.18	-1.78
Total (%)	2.75	2.10	1.11	2.92	3.02	4.48

References

[1] Sheldon L. Glashow. ‘Partial-symmetries of weak interactions’. *Nuclear Physics* 22.4 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2 (cited on pages 5, 9).

[2] Steven Weinberg. ‘A Model of Leptons’. *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264 (cited on pages 5, 9).

[3] Abdus Salam and J.C. Ward. ‘Weak and electromagnetic interactions’. *Il Nuovo Cimento Series 10* 11.4 (1959), pp. 568–577. DOI: 10.1007/BF02726525 (cited on pages 5, 9).

[4] M. Gell-Mann. ‘A schematic model of baryons and mesons’. *Physics Letters* 8.3 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3 (cited on page 5).

[5] David J. Gross and Frank Wilczek. ‘Asymptotically Free Gauge Theories. I’. *Phys. Rev. D* 8 (10 1973), pp. 3633–3652. DOI: 10.1103/PhysRevD.8.3633 (cited on pages 5, 11).

[6] H. David Politzer. ‘Reliable Perturbative Results for Strong Interactions?’ *Phys. Rev. Lett.* 30 (26 1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346 (cited on pages 5, 11).

[7] The Super-Kamiokande Collaboration. ‘Evidence for Oscillation of Atmospheric Neutrinos’. *Phys. Rev. Lett.* 81 (8 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562 (cited on pages 5, 16).

[8] J. Beringer et al. ‘Review of Particle Physics (and 2013 partial update for the 2014 edition)’. *Phys. Rev. D* 86 (1 2012). DOI: 10.1103/PhysRevD.86.010001 (cited on pages 6, 18, 20).

[9] The ATLAS Collaboration. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.' *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020 (cited on pages 6, 15).

[10] The CMS Collaboration. 'Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.' *Physics Letters B* 716.1 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021 (cited on pages 6, 15, 36).

[11] ATLAS, CDF, CMS and DØ Collaborations. *First combination of Tevatron and LHC measurements of the top-quark mass.* Tech. rep. arXiv:1403.4427, ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014. Geneva: CERN, 2014 (cited on pages 6, 26, 27, 108).

[12] C. N. Yang and R. L. Mills. 'Conservation of Isotopic Spin and Isotopic Gauge Invariance'. *Phys. Rev.* 96 (1 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191 (cited on page 8).

[13] C. S. Wu et al. 'Experimental Test of Parity Conservation in Beta Decay'. *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413 (cited on page 9).

[14] D. Griffiths. *Introduction to Elementary Particles.* Physics textbook. Wiley, 2008 (cited on page 12).

[15] The LHCb Collaboration. *Observation of the resonant character of the $Z(4430)^-$ state.* Tech. rep. arXiv:1404.1903. LHCB-PAPER-2014-014. CERN-PH-EP-2014-061. CERN, 2014 (cited on page 12).

[16] Peter W. Higgs. 'Broken Symmetries and the Masses of Gauge Bosons'. *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cited on page 13).

[17] F. Englert and R. Brout. 'Broken Symmetry and the Mass of Gauge Vector Mesons'. *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cited on page 13).

[18] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble. 'Global Conservation Laws and Massless Particles'. *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cited on page 13).

[19] A. Quadt. 'Top quark physics at hadron colliders'. *The European Physical Journal C - Particles and Fields* 48.3 (2006), pp. 835–1000. DOI: 10.1140/epjc/s2006-02631-6. URL: 10.1140/epjc/s2006-02631-6 (cited on page 14).

[20] The CMS Collaboration. 'Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV'. *Journal of High Energy Physics* 2013.6 (2013), pp. 1–127. DOI: 10.1007/JHEP06(2013)081 (cited on page 15).

[21] The ATLAS Collaboration. 'Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC'. *Physics Letters B* 726.1-3 (2013), pp. 88–119. DOI: 10.1016/j.physletb.2013.08.010 (cited on page 15).

[22] Takehiko Asaka, Steve Blanchet and Mikhail Shaposhnikov. 'The νMSM, dark matter and neutrino masses'. *Physics Letters B* 631.4 (2005), pp. 151–156. DOI: 10.1016/j.physletb.2005.09.070 (cited on page 16).

[23] Michael E. Peskin. 'The Matter with antimatter'. *Nature* 419 (2002), pp. 24–27. DOI: 10.1038/419024a (cited on page 16).

[24] Michael Dine and Alexander Kusenko. 'Origin of the matter-antimatter asymmetry'. *Rev. Mod. Phys.* 76 (1 2003), pp. 1–30. DOI: 10.1103/RevModPhys.76.1 (cited on page 16).

[25] Planck Collaboration. 'Planck 2013 results. I. Overview of products and scientific results'. *Submitted to Astronomy & Astrophysics* (2013). arXiv: 1303.5062 (cited on page 16).

[26] Planck Collaboration. 'Planck 2013 results. XVI. Cosmological parameters'. *Submitted to Astronomy & Astrophysics* (2013). arXiv: 1303.5076 (cited on page 16).

[27] Stephen P. Martin. 'A Supersymmetry primer' (1997). arXiv: hep-ph/9709356 [hep-ph] (cited on page 17).

[28] Nima Arkani-Hamed, Savas Dimopoulos and Gia Dvali. 'The hierarchy problem and new dimensions at a millimeter'. *Physics Letters B* 429.3-4 (1998), pp. 263–272. DOI: 10.1016/S0370-2693(98)00466-3 (cited on page 17).

[29] Lisa Randall and Raman Sundrum. 'Large Mass Hierarchy from a Small Extra Dimension'. *Phys. Rev. Lett.* 83 (17 1999), pp. 3370–3373. DOI: 10.1103/PhysRevLett.83.3370 (cited on page 17).

[30] Christopher T. Hill. 'Topcolor: top quark condensation in a gauge extension of the standard model'. *Physics Letters B* 266.3-4 (1991), pp. 419–424. DOI: 10.1016/0370-2693(91)91061-Y (cited on page 17).

[31] Kaustubh Agashe, Roberto Contino and Alex Pomarol. 'The minimal composite Higgs model'. *Nuclear Physics B* 719.1-2 (2005), pp. 165–187. DOI: 10.1016/j.nuclphysb.2005.04.035 (cited on page 17).

[32] CDF Collaboration. 'Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab'. *Phys. Rev. Lett.* 74 (14 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626 (cited on page 18).

[33] DØ Collaboration. 'Observation of the Top Quark'. *Phys. Rev. Lett.* 74 (14 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632 (cited on page 18).

[34] The CMS Collaboration. *Measurement of the t-channel single-top-quark production cross section and of the $|V_{tb}|$ CKM matrix element in pp collisions at $\sqrt{s}$ = 8 TeV*. Tech. rep. arXiv:1403.7366. CMS-TOP-12-038-003. CERN-PH-EP-2014-032. Submitted to JHEP. Geneva: CERN, 2014 (cited on page 20).

[35] F.A. Berends et al. ‘Multijet production in W, Z events at pp colliders’. *Physics Letters B* 224.1-2 (1989), pp. 237–242. DOI: 10.1016/0370-2693(89)91081-2 (cited on page 22).

[36] R. Tarrach. ‘The pole mass in perturbative QCD’. *Nuclear Physics B* 183.3 (1981), pp. 384–396. DOI: http://dx.doi.org/10.1016/0550-3213(81)90140-1 (cited on page 24).

[37] Martin C. Smith and Scott S. Willenbrock. ‘Top-quark Pole Mass’. *Phys. Rev. Lett.* 79 (20 1997), pp. 3825–3828. DOI: 10.1103/PhysRevLett.79.3825 (cited on pages 24, 25).

[38] Andreas S. Kronfeld. ‘Perturbative pole mass in QCD’. *Phys. Rev. D* 58 (5 1998). DOI: 10.1103/PhysRevD.58.051501 (cited on page 24).

[39] M. Beneke and V.M. Braun. ‘Heavy quark effective theory beyond perturbation theory: renormalons, the pole mass and the residual mass term’. *Nuclear Physics B* 426.2 (1994), pp. 301–343. DOI: 10.1016/0550-3213(94)90314-X (cited on page 24).

[40] M. Beneke. ‘Renormalons’. *Physics Reports* 317.1-2 (1999), pp. 1–142. DOI: 10.1016/S0370-1573(98)00130-6 (cited on page 24).

[41] Fred Jegerlehner, Mikhail Yu. Kalmykov and Bernd A. Kniehl. ‘On the difference between the pole and the $\overline{\mathrm{MS}}$ masses of the top quark at the electroweak scale’. *Physics Letters B* 722.1-3 (2013), pp. 123–129. DOI: 10.1016/j.physletb.2013.04.012 (cited on pages 24, 25).

[42] André H. Hoang and Iain W. Stewart. ‘Top Mass Measurements from Jets and the Tevatron Top-Quark Mass’. *Nuclear Physics B - Proceedings Supplements* 185.0 (2008). Proceedings of the Second Workshop on Theory, Phenomenology and Experiments in Heavy Flavour Physics, pp. 220–226. DOI: 10.1016/j.nuclphysbps.2008.10.028 (cited on page 25).

[43] Giuseppe Degrassi et al. 'Higgs mass and vacuum stability in the Standard Model at NNLO'. *Journal of High Energy Physics* 2012.8 (2012). DOI: 10.1007/JHEP08(2012)098 (cited on pages 25, 26).

[44] S. Alekhin, A. Djouadi and S. Moch. 'The top quark and Higgs boson masses and the stability of the electroweak vacuum'. *Physics Letters B* 716.1 (2012), pp. 214–219. DOI: 10.1016/j.physletb.2012.08.024 (cited on page 25).

[45] Tevatron Electroweak Working Group, CDF and DØ Collaborations. *Combination of CDF and DO results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron*. Tech. rep. 2013. arXiv: 1305.3929 (cited on page 26).

[46] ATLAS and CMS Collaborations. *Combination of ATLAS and CMS results on the mass of the top-quark using up to 4.9 fb^{-1} of $\sqrt{s} = 7$ TeV LHC data*. Tech. rep. ATLAS-CONF-2013-102. Geneva: CERN, 2013 (cited on page 26).

[47] Howard Baer et al. 'The International Linear Collider Technical Design Report - Volume 2: Physics' (2013). arXiv: 1306.6352 [hep-ph] (cited on page 26).

[48] Jun Gao et al. 'CT10 next-to-next-to-leading order global analysis of QCD'. *Phys. Rev. D* 89 (3 2014). DOI: 10.1103/PhysRevD.89.033009 (cited on pages 28, 29).

[49] John C. Collins, Davison E. Soper and George Sterman. 'Heavy particle production in high-energy hadron collisions'. *Nuclear Physics B* 263 (1986), pp. 37–60. DOI: 10.1016/0550-3213(86)90026-X (cited on page 28).

[50] J M Campbell, J W Huston and W J Stirling. 'Hard interactions of quarks and gluons: a primer for LHC physics'. *Reports on Progress in Physics* 70.1 (2007), p. 89. DOI: 10.1088/0034-4885/70/1/R02 (cited on pages 28, 35).

[51] Michał Czakon, Paul Fiedler and Alexander Mitov. 'The total top quark pair production cross-section at hadron colliders through O(α_S^4)'. *Phys. Rev. Lett.*

110 (25 2013). DOI: 10.1103/PhysRevLett.110.252004 (cited on pages 29, 30, 126).

[52] The ATLAS Collaboration. 'Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV'. *The European Physical Journal C* 73.1 (2013), pp. 1–28. DOI: 10.1140/epjc/s10052-012-2261-1 (cited on pages 30, 31).

[53] The CMS Collaboration. 'Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV'. *The European Physical Journal C* 73.3 (2013), pp. 1–29. DOI: 10.1140/epjc/s10052-013-2339-4 (cited on pages 30, 31, 145).

[54] L. Evans and P. Bryant. 'LHC Machine'. *Journal of Instrumentation* 3.08 (2008). DOI: 10.1088/1748-0221/3/08/S08001 (cited on page 33).

[55] J.P. Blewett. '200 GeV intersecting storage accelerators'. *Proceedings of The 8th International Conference on High-Energy Accelerators*. CERN. Geneva, Switzerland, 1971 (cited on page 33).

[56] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1 (cited on page 35).

[57] Mike Lamont. 'The LHC's first long run'. *CERN Courier* 53.7 (2013) (cited on page 35).

[58] The CMS Collaboration. 'The CMS experiment at the CERN LHC'. *Journal of Instrumentation* 3.08 (2008). DOI: 10.1088/1748-0221/3/08/S08004 (cited on pages 36, 38, 41–43, 46).

[59] The CMS Collaboration. *CMS Tracker Detector Performance Results*. Online, accessed 9 April 2014. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK (cited on pages 39, 40).

[60] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. Geneva, Switzerland: CERN, 2006 (cited on pages 41, 42, 44, 121, 122).

[61] The CMS Collaboration. *CMS ECAL Detector Performance Results*. Online, accessed 9 April 2014. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResults (cited on page 42).

[62] The CMS Collaboration. 'Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays'. *Journal of Instrumentation* 5.03 (2010). DOI: 10.1088/1748-0221/5/03/T03021 (cited on page 44).

[63] 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 (1997). New Computing Techniques in Physics Research V, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X (cited on page 48).

[64] *CMSSW software framework*. URL: http://cms-sw.github.io/cmssw (cited on pages 49, 54).

[65] *Bristol Analysis Tools*. URL: https://github.com/BristolTopGroup/AnalysisSoftware (cited on page 49).

[66] *Daily Python Scripts*. URL: https://github.com/BristolTopGroup/DailyPythonScripts (cited on page 49).

[67] *rootpy: Pythonic ROOT package*. URL: http://rootpy.org (cited on page 50).

[68] *matplotlib plotting library*. URL: http://matplotlib.org (cited on page 50).

[69] GEANT4 Collaboration. 'GEANT4 - a simulation toolkit'. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cited on pages 50, 54).

[70] Stefan Höche. *Introduction to Monte Carlo Event Generators.* Slides from 2012 CTEQ-Fermilab School on QCD and Electroweak Phenomenology. URL: http://www.physics.smu.edu/olness/ftp/misc/cteq/2012/Hoeche1.pdf (cited on page 51).

[71] Yuri L. Dokshitzer. 'Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics'. *Sov. Phys. JETP* 46 (1977), pp. 641–653 (cited on page 52).

[72] V.N. Gribov and L.N. Lipatov. 'Deep inelastic *ep* scattering in perturbation theory'. *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450 (cited on page 52).

[73] G. Altarelli and G. Parisi. 'Asymptotic freedom in parton language'. *Nuclear Physics B* 126.2 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4 (cited on page 52).

[74] B. Andersson et al. 'Parton fragmentation and string dynamics'. *Physics Reports* 97.2-3 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7 (cited on page 52).

[75] Gennaro Corcella et al. 'HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)'. *Journal of High Energy Physics* 2001.01 (2001). DOI: 0.1088/1126-6708/2001/01/010 (cited on page 52).

[76] Rick D. Field. *The underlying event in hard scattering processes.* Tech. rep. SNOWMASS-2001-P501, UFIFT-HEP-01-19, FERMILAB-CONF-02-009-E. 2001. arXiv: hep-ph/0201192 (cited on page 52).

[77] Torbjörn Sjöstrand. 'Colour reconnection and its effects on precise measurements at the LHC' (2013). arXiv: 1310.8073 (cited on page 52).

[78] Johan Alwall et al. 'MadGraph 5: going beyond'. *Journal of High Energy Physics* 2011.6 (2011), pp. 1–40. DOI: 10.1007/JHEP06(2011)128 (cited on page 53).

[79] Torbjörn Sjöstrand et al. 'High-energy-physics event generation with Pythia 6.1'. *Computer Physics Communications* 135.2 (2001), pp. 238–259. DOI: 10.1016/S0010-4655(00)00236-8 (cited on page 53).

[80] Torbjörn Sjöstrand, Stephen Mrenna and Peter Z. Skands. 'PYTHIA 6.4 Physics and Manual'. *Journal of High Energy Physics* 05 (2006). DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 (cited on page 53).

[81] Stefano Frixione and Bryan R. Webber. 'Matching NLO QCD computations and parton shower simulations'. *Journal of High Energy Physics* 06 (2002). DOI: 10.1088/1126-6708/2002/06/029 (cited on page 53).

[82] Stefano Frixione, Paolo Nason and Carlo Oleari. 'Matching NLO QCD computations with parton shower simulations: the POWHEG method'. *Journal of High Energy Physics* 11 (2007). DOI: 10.1088/1126-6708/2007/11/070 (cited on pages 53, 54).

[83] Stefan Höche et al. 'Matching parton showers and matrix elements' (2006). Proceedings of the "HERA and the LHC" workshop, CERN/DESY 2004/2005". arXiv: hep-ph/0602031 (cited on page 53).

[84] The CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. CMS-PAS-PFT-09-001. Geneva, Switzerland: CERN, 2009 (cited on pages 55, 60).

[85] R. Frühwirth. 'Application of Kalman filtering to track and vertex fitting'. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2-3 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4 (cited on pages 56, 62).

[86] W Adam et al. '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 (2005). DOI: 10.1088/0954-3899/31/9/N01 (cited on page 56).

[87] S. Baffioni et al. 'Electron reconstruction in CMS'. *The European Physical Journal C* 49.4 (2007), pp. 1099–1116. DOI: 10.1140/epjc/s10052-006-0175-5 (cited on page 58).

[88] The CMS Collaboration. *Cut-based Electron Identification algorithm.* Online, accessed 04 September 2013. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCutBasedElectronID (cited on page 60).

[89] The CMS Collaboration. *Cuts in Categories (CiC) Electron Identification.* Online, accessed 04 September 2013. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCategoryBasedElectronID (cited on pages 60, 89).

[90] The CMS Collaboration. *Electron reconstruction and identification at $sqrt(s) = 7$ TeV.* Tech. rep. CMS-PAS-EGM-10-004. Geneva: CERN, 2010 (cited on pages 61, 62).

[91] The CMS collaboration. 'Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV'. *Journal of Instrumentation* 7.10 (2012). DOI: 10.1088/1748-0221/7/10/P10002 (cited on page 62).

[92] Matteo Cacciari, Gavin P. Salam and Gregory Soyez. 'The anti-k_t jet clustering algorithm'. *Journal of High Energy Physics* 2008.04 (2008). DOI: 10.1088/1126-6708/2008/04/063 (cited on page 64).

[93] Ricardo Eusebi 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 (2012). DOI: 10.1088/1742-6596/404/1/012014 (cited on page 64).

[94] The CMS collaboration. 'Identification of b-quark jets with the CMS experiment'. *Journal of Instrumentation* 8.04 (2013). DOI: 10.1088/1748-0221/8/04/P04013 (cited on pages 66, 67, 106).

[95] The CMS Collaboration. *CMS TriDAS project: Technical Design Report, Vol. 1: The Trigger Systems*. CERN-LHCC-2000-038; CMS-TDR-6-1. CERN, 2000 (cited on pages 70, 122).

[96] J. Brooke. 'Performance of the CMS Level-1 Trigger'. *Proceedings of Science,* ICHEP2012 (2013). arXiv: 1302.2469 (cited on page 70).

[97] The CMS collaboration. 'The CMS High Level Trigger System'. *2007 15th IEEE-NPSS Real-Time Conference*. Institute of Electrical and Electronics Engineers, 2007. DOI: 10.1109/RTC.2007.4382773 (cited on page 71).

[98] The CMS collaboration. 'Commissioning of the CMS High Level Trigger'. *Journal of Instrumentation* 4.10 (2009). DOI: 10.1088/1748-0221/4/10/P10005 (cited on page 72).

[99] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. CMS-NOTE-2006-041. Geneva, Swtizerland: CERN, 2006 (cited on page 73).

[100] Diego Casadei. 'Estimating the selection efficiency'. *Journal of Instrumentation* 7.08 (2012). DOI: 10.1088/1748-0221/7/08/P08021 (cited on page 75).

[101] M. Agelou et al. *Top Trigger Efficiency Measurements and the top trigger package*. Tech. rep. DØ Note 4512. 2004 (cited on page 76).

[102] The CMS Collaboration. 'Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s} = 7$ TeV'. *Journal of High Energy Physics* 2012.12 (2012), pp. 1–37. DOI: 10.1007/JHEP12(2012)105 (cited on pages 85, 91, 93, 97, 108).

[103] The CMS Collaboration. 'Measurement of the mass difference between top and antitop quarks'. *Journal of High Energy Physics* 2012.6 (2012), pp. 1–36. DOI: 10.1007/JHEP06(2012)109 (cited on page 85).

[104] The CMS Collaboration. *Measurement of the top quark mass in the l+jets channel*. Tech. rep. CMS-PAS-TOP-10-009. Geneva, Switzerland: CERN, 2011 (cited on page 85).

[105] Daniel R. Stump. 'A new generation of CTEQ parton distribution functions with uncertainty analysis'. *Proceedings to ICHEP 2002* 117 (2003), pp. 265–267. DOI: 10.1016/S0920-5632(03)90541-X (cited on pages 86, 107).

[106] The CMS Collaboration. *Tracking and Primary Vertex Results in First 7 TeV Collisions*. Tech. rep. CMS-PAS-TRK-10-005. Geneva, Switzerland: CERN, 2010 (cited on page 89).

[107] Suharyo Sumowidagdo. *Programming documentation of HitFit: A Kinematic Fitter for top quark-antiquark pair lepton+jets event*. CMS AN-2011/171. 2011 (cited on page 93).

[108] Orin I. Dahl, T.B. Day and Frank T. Solmitz. *SQUAW Kinematic Fitting Program*. Tech. rep. P-126. Berkeley, California: Lawrence Radiation Laboratory, 1968 (cited on page 93).

[109] DØ Collaboration. 'Direct measurement of the top quark mass by the DØ Collaboration'. *Phys. Rev. D* 58 (5 1998). DOI: 10.1103/PhysRevD.58.052001 (cited on page 93).

[110] DØ Collaboration. 'Measurement of the top quark mass in the lepton+jets channel using the ideogram method'. *Phys. Rev. D* 75 (9 2007). DOI: 10.1103/PhysRevD.75.092001 (cited on pages 93, 97).

[111] DØ Collaboration. 'Measurement of the Forward-Backward Charge Asymmetry in Top-Quark Pair Production'. *Phys. Rev. Lett.* 100 (14 2008). DOI: 10.1103/PhysRevLett.100.142002 (cited on page 93).

[112] *CLHEP - A Class Library for High Energy Physics*. URL: http://cern.ch/CLHEP (cited on page 93).

[113] Scott Stuart Snyder. 'Measurement of the Top Quark Mass at DØ'. FERMILAB-THESIS-1995-27. PhD thesis. Stony Brook University, 1995 (cited on page 94).

[114] DELPHI Collaboration. 'Measurement of the W-pair cross-section and of the W mass in e^+e^- interactions at 172 GeV'. *The European Physical Journal C - Particles and Fields* 2.4 (1998), pp. 581–595. DOI: 10.1007/s100529800857 (cited on page 97).

[115] DELPHI Collaboration. 'Measurement of the mass and width of the W boson in e^+e^- collisions at $\sqrt{s}$ = 161-209 GeV'. *The European Physical Journal C* 55.1 (2008), pp. 1–38. DOI: 10.1140/epjc/s10052-008-0585-7 (cited on page 97).

[116] CDF Collaboration. 'Measurement of the Top-Quark Mass in All-Hadronic Decays in $p\overline{p}$ Collisions at CDF II'. *Phys. Rev. Lett.* 98 (14 2007). DOI: 10.1103/PhysRevLett.98.142001 (cited on page 97).

[117] DØ Collaboration. 'Improved determination of the width of the top quark'. *Phys. Rev. D* 85 (9 2012). DOI: 10.1103/PhysRevD.85.091104 (cited on page 99).

[118] The ALEPH Collaboration et al. 'Precision electroweak measurements on the Z resonance'. *Physics Reports* 427.5-6 (2006), pp. 257–454. DOI: 10.1016/j.physrep.2005.12.006 (cited on page 99).

[119] M. Oreglia. 'A Study of the Reactions $\psi' \to \gamma\gamma\psi$'. PhD thesis. 1980 (cited on page 100).

[120] Diego Martínez Santos and Frederic Dupertuis. 'Mass distributions marginalized over per-event errors' (2013). arXiv: 1312.5000 [hep-ex] (cited on page 100).

[121] K.S. Kölbig and B. Schorr. 'A program package for the Landau distribution'. *Computer Physics Communications* 31.1 (1984), pp. 97–111. DOI: 10.1016/0010-4655(84)90085-7 (cited on page 100).

[122] The CMS collaboration. 'Determination of jet energy calibration and transverse momentum resolution in CMS'. *Journal of Instrumentation* 6.11 (2011). DOI: 10.1088/1748-0221/6/11/P11002 (cited on page 106).

[123] Peter Z. Skands. 'Tuning Monte Carlo generators: The Perugia tunes'. *Phys. Rev. D* 82 (7 2010). DOI: 10.1103/PhysRevD.82.074018 (cited on page 106).

[124] The CMS Collaboration. *Measurement of missing transverse energy in top pair events at* $\sqrt{s} = 7$ *TeV*. Tech. rep. CMS-PAS-TOP-12-019. Geneva: CERN, 2012 (cited on pages 109, 152).

[125] The CMS Collaboration. *Measurement of missing transverse energy and other event-level observables in top pair events at* $\sqrt{s} = 8$ *TeV*. Tech. rep. CMS-PAS-TOP-12-042. Geneva: CERN, 2013 (cited on page 109).

[126] The CMS Collaboration. *CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update*. Tech. rep. CMS-PAS-LUM-12-001. Geneva: CERN, 2012 (cited on page 112).

[127] The CMS Collaboration. *Inelastic pp cross section at 7 TeV*. Tech. rep. CMS-PAS-FWD-11-001. Geneva: CERN, 2011 (cited on page 112).

[128] The CMS Collaboration. *Results on b-tagging identification in 8 TeV pp collisions*. Tech. rep. CMS-DP-2013-005. 2013 (cited on page 114).

[129] Andrea Rizzi. *"Events weights from b-tagging effciency SF"*. URL: https://twiki.cern.ch/twiki/pub/CMS/BTagWeight/BTagWeight.pdf (cited on page 114).

[130] Pushpalatha Bhat et al. *Simulation Studies of the Radiation Environment in the CMS Detector for pp Collisions at the LHC*. Tech. rep. CERN-CMS-NOTE-2013-001. Geneva: CERN, 2010 (cited on page 121).

[131] The CMS Collaboration. 'Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV'. *The European Physical*

Journal C 73.3 (2013). DOI: 10.1140/epjc/s10052-013-2339-4 (cited on page 136).

[132] The CMS Collaboration. *Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV*. Tech. rep. CMS-PAS-TOP-12-027. Geneva: CERN, 2013 (cited on page 136).

[133] Andreas Höcker and Vakhtang Kartvelishvili. 'SVD approach to data unfolding'. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 372.3 (1996), pp. 469–481. DOI: 10.1016/0168-9002(95)01478-0 (cited on pages 139, 140).

[134] T. Adye. 'Unfolding algorithms and tests using RooUnfold'. *Proceedings of the PHYSTAT 2011 Workshop*. CERN. Geneva, Switzerland, 2011, pp. 313–318 (cited on page 140).