

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Measurement of the Top Quark Pair Production Cross Section in pp Collisions
at a Center-of-Mass Energy of 7 TeV with the CMS Experiment at the LHC

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Geng-Yuan Jeng

June 2011

Dissertation Committee:

Dr. Gail Hanson, Chairperson
Dr. Stephen Wimpenny
Dr. Ernest Ma

Copyright by
Geng-Yuan Jeng
2011

The Dissertation of Geng-Yuan Jeng is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

After years of immersion in experimental high energy physics, I finally reach the point to thank all the people who have played special roles during my PhD study. First of all, I would like to express my gratitude to my thesis advisor, Professor Gail Hanson, I would not have gone all the way without her guidance and generous support during the years. Her very own experience and broad knowledge of experimental high energy physics have always been the sources of inspirations. She has also been very considerate of my personal status and open-minded in adjusting the plan for my work. I would also like to acknowledge her assistance in reading and correcting this thesis.

My very special thanks go to Dr. Kevin Burkett for being a good friend and my advisor at the LHC Physics Center at Fermilab. I have a lot of respect for his wisdom of dealing with difficult situations, e.g., when we were under a lot of pressure to improve beamspot monitoring. His encouragement had boosted me for extra miles.

I want to thank Dr. Francisco Yumiceva in particular. He has been resourceful and always ready to provide ideas. Through numerous constructive discussions with him, I was able to crank up to full speed in physics analysis. I also learned from him many useful tricks and techniques in computer programming; this helped ease my life significantly when I finally got the data to analyze.

I am very grateful to Professor Joe Incandela and Dr. Dan Green for their support during my stays at CERN and at Fermilab, respectively. Many thanks to Dr. Leonard Spiegel and Dr. Weimin Wu for their helps during my stay at Fermilab.

Before moving to Fermilab I also spent two and a half years at CERN, where many people made my stay enjoyable. Dr. Gabriella Pasztor and Dr. Asish Satpathy shared their knowledge of physics with me and provided useful advices. Thanks to Dr. Duccio Abbaneo, Dr. Slawek Tkaczyk, Professor Steve Nahm, and Dr. Kristian Hahn for their helps and support.

As a young researcher in experimental high energy physics, I am honored to join the CMS experiment and participate in building such a fantastic experiment. I am greatly indebted to my collaborators for building and operating the CMS detector and to all the staff of CERN for their excellent works in running the Large Hadron Collider.

Finally, I would like to take this opportunity to express my thanks to Professor Kenneth Ming-Der Lan, Professor Lee Lin, and Professor Ta-Kuan Lin for their guidance and support when I started my research career in National Chung Hsing University.

To my parents and my wife for their love and support.

ABSTRACT OF THE DISSERTATION

Measurement of the Top Quark Pair Production Cross Section in pp Collisions
at a Center-of-Mass Energy of 7 TeV with the CMS Experiment at the LHC

by

Geng-Yuan Jeng

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, June 2011
Dr. Gail Hanson, Chairperson

We perform a measurement of the top quark pair production cross section in pp collisions at a center-of-mass energy of 7 TeV at the Large Hadron Collider. The analysis is based on a 36 pb^{-1} dataset collected with the CMS experiment in 2010. We study the kinematic characteristics of events with exactly one isolated muon in association with additional hard jets and exhibiting substantial missing transverse energy to distinguish the $t\bar{t}$ signal from the electroweak and QCD multijet backgrounds. The $t\bar{t}$ signal is extracted by simultaneously fitting the tri-jet invariant mass distributions in events with exactly three jets and at least four jets. The measured cross section is $157 \pm 12(\text{stat.})^{+28}_{-35}(\text{syst.}) \pm 6(\text{lumi.}) \text{ pb}$, which is consistent with the theoretical predictions.

Contents

List of Figures	xi
List of Tables	xvi
Introduction	1
1 The Standard Model of Particle Physics	3
1.1 The Standard Model	3
1.1.1 Gauge Interactions	5
1.1.2 Spontaneous Symmetry Breaking	9
1.2 Top Quark Physics	11
1.2.1 Production Mechanism	13
1.2.2 Production Cross Section	14
2 The Large Hadron Collider	16
3 The CMS Detector	22
3.1 The CMS Coordinates	23
3.2 The CMS Silicon Tracker	24
3.2.1 The Pixel Detector	25
3.2.2 The Silicon Strip Tracker	26
3.3 The CMS Calorimeter System	28
3.3.1 The Electromagnetic Calorimeter	28
3.3.2 The Hadronic Calorimeter	30
3.4 The CMS Muon Spectrometer	32
3.4.1 The Drift Tube Chambers	33
3.4.2 The Cathode Strip Chambers	34
3.4.3 The Resistive Plate Chambers	36
3.5 The Trigger System	36
3.6 The Data Acquisition System	39

4 Data Handling, Event Reconstruction and Monte Carlo Simulation	41
4.1 Data Sample	41
4.2 Analysis Object Reconstruction	42
4.2.1 Beamspot, Track and Vertex Reconstruction	44
4.2.2 Energy Reconstruction in the Calorimeters	47
4.2.3 Muon Reconstruction	47
4.2.4 Electron Reconstruction	48
4.2.5 Jet Reconstruction	49
4.2.6 Missing Transverse Energy	50
4.3 Monte Carlo Simulation	51
4.3.1 Event Generation	52
4.3.2 Detector Simulation	54
4.3.3 Signal Modeling	54
4.3.4 Background Modeling	55
4.3.5 Monte Carlo Samples	57
5 Event Selection and Data Analysis	59
5.1 Event Selection	59
5.2 Muon Efficiencies	67
5.3 QCD Background Estimation	69
5.3.1 The Matrix Method	70
5.3.2 The Results	76
5.4 Comparison between Data and Monte Carlo Predictions	78
5.4.1 Data-driven QCD multijet template	78
5.4.2 Data/Monte Carlo Comparisons	79
6 Measurement of the $t\bar{t}$ Cross Section	89
6.1 The Method	90
6.1.1 Study of the Fit Stability	91
6.1.2 The Results	92
6.2 Estimation of the Systematic Uncertainties	95
6.2.1 Jet Energy Scale	97
6.2.2 Jet Energy Resolution	98
6.2.3 Factorization Scale	98
6.2.4 Matching Threshold	99
6.2.5 Initial and Final State Radiation	100
6.2.6 Underlying Event Modeling	100
6.2.7 Parton Distribution Function	100
6.2.8 Summary of the Systematic Uncertainties	101
7 Conclusions	116
Bibliography	117
Glossary	125

A The L1 and HLT Triggers	130
B CMSSW	132
C The CMS Tracker Integration at CERN	133
C.1 The Tracker Outer Barrel Integration	133
C.2 The Rod-In-a-Box Project	135
C.2.1 Setup of the RIB	135
C.2.2 Local Test	137
C.2.3 Global Test	138
C.3 The CMS Tracker Online Software	138
D CMS Beamspot Data Quality Monitoring	139

List of Figures

1.1	Feynman diagram for the $pp \rightarrow t\bar{t}$ process.	14
1.2	Feynman diagrams of the LO processes for $t\bar{t}$ production via strong interactions: (a) quark-antiquark annihilation; (b) gluon-gluon fusion.	15
2.1	Schematic layout of the LHC (Beam 1 - clockwise, Beam 2 - counterclockwise)	17
2.2	Schematic of the CERN accelerator complex.	18
2.3	Instantaneous and integrated luminosity of the LHC: Left: The instantaneous luminosity as a function of time. The peak values is $2.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$. Right: The integrated luminosity as a function of time.	21
3.1	The Compact Muon Solenoid (CMS) layout.	23
3.2	Schematic cross sectional view of the CMS tracker. Each line represents a detector module. Double lines indicate double sided silicon strip modules mounted back-to-back with a stereo angle of 100 mrad to deliver stereo hits.	25
3.3	The layout of the pixel detector. It consists of three barrel layers (BPix) with two endcap disks (FPix).	26
3.4	Transverse view in the (r, z) plane of a quarter of the ECAL layout.	30
3.5	Transverse $r - z$ view of a quarter of the CMS muon spectrometer. The barrel region is integrated with Drift Tube (DT) chambers and Resistive Plate Chambers (RPCs), while the endcap region is equipped with Cathode Strip Chambers (CSCs) and RPCs. In total four barrel muon stations, MB 1 to MB 4, and four endcap muon stations, ME 1 to ME 4, are installed.	33
3.6	The $r - \phi$ view of a Drift Tube (DT) chamber inside a muon barrel station. One can see the 2 SLs, SL Φ_1 and SL Φ_2 , with wires along the z direction (pointing out of the paper) and the other (SL Θ) perpendicular to it. In between is a honeycomb plate with supports attached to the iron yoke.	34
3.7	Schematic view of a CSC chamber.	35
3.8	Schematic layout of one of the 5 barrel wheels. Each wheel is divided into 12 sectors that are numbered as shown. There are two RPCs for each muon station, except for the special cases for RB 4 in sector 4, which consists of four RPCs, and in sectors 9 and 11, each contains only one RPC.	37
3.9	Architecture of the Level-1 (L1) Trigger.	39
3.10	General architecture of the CMS Data Acquisition (DAQ) system.	40

4.1	The workflow of the tracking algorithm.	46
5.1	Muon ID&ISO efficiencies as functions of kinematic variables: p_T (upper) and η (bottom) of the muon.	69
5.2	Muon ID&ISO efficiencies as functions of kinematic variables: ϕ (upper) and $\Delta R(\mu, \text{jet})$ (bottom) to the closest jet.	70
5.3	Muon ISO efficiencies (upper) and scale factor (bottom) as a function of $\Delta R(\mu, \text{jet})$ to the closest jet in a sample with at least two jets.	71
5.4	ε_{QCD} is measured in the QCD region ($\cancel{E}_T < 10$ GeV), while the estimation of the QCD multijet background is done in the signal region ($\cancel{E}_T > 20$ GeV). The tight sample is a subset of the loose sample consisting of those events with $I_{\text{rel}}^\mu < 0.1$	72
5.5	ε_{QCD} for the QCD Monte Carlo sample as a function of \cancel{E}_T . The dashed lines represent the total uncertainty used in data.	74
5.6	The two-dimensional scatter plot of events with at least two jets, selected as a function of the Relative Isolation and \cancel{E}_T for data. The sizes of the rectangles represent the numbers of events.	74
5.7	The two-dimensional scatter plots of events with at least two jets, selected as a function of the Relative Isolation and \cancel{E}_T for the $t\bar{t}$, $W+\text{jets}$ and QCD Monte Carlo samples. The sizes of the rectangles represent the numbers of events.	75
5.8	Transverse mass of the reconstructed W boson for events with at least one jet in the orthogonal data sample (left) and in the orthogonal QCD Monte Carlo sample (right). A small bump can be observed around 40 GeV/ c^2 . Most of the events that contribute to the bump have a muon which is opposite to the direction from the \cancel{E}_T in the transverse plane, i.e., those events have large $\Delta\phi(\mu, \cancel{E}_T)$ values. This effect is enhanced due to the muon p_T and \cancel{E}_T thresholds both at 20 GeV/ c	80
5.9	Transverse mass of the W boson using the QCD Monte Carlo sample with at least one jet. The requirement of $\Delta\phi(\mu, \cancel{E}_T) < 2$ rad has a small effect on the QCD Monte Carlo sample.	80
5.10	Shape comparisons between QCD Monte Carlo and the orthogonal sample for muon p_T (upper plots) and W transverse mass (lower plots) for events with 1 (left) and 2 (right) jets.	81
5.11	Shape comparisons between QCD Monte Carlo and the orthogonal sample for muon p_T (upper plots) and W transverse mass (lower plots) for events with 3 (left) and 4 or more (right) jets.	82
5.12	Data/Monte Carlo comparisons of the jet multiplicity. The QCD multijet contribution is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z/\gamma^*)+\text{jets}$ Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.	84

- 5.13 Data/Monte Carlo comparisons of the muon p_T for the 1-jet (left) and 2-jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown. 85
- 5.14 Data/Monte Carlo comparisons of the muon p_T for the 3-jet (left) and inclusive four jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown. 85
- 5.15 Data/Monte Carlo comparisons of the missing transverse energy for the 1-jet (left) and 2-jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown. 86
- 5.16 Data/Monte Carlo comparisons of the missing transverse energy for the 3-jet (left) and inclusive four jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown. 86
- 5.17 Data/Monte Carlo comparisons of the $W M_T$ for the 1-jet (left) and 2-jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown. 87
- 5.18 Data/Monte Carlo comparisons of the $W M_T$ for the 3-jet (left) and inclusive four jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown. 87

5.19 Data/Monte Carlo comparisons of M3 for the 3-jet (left) and inclusive four jet (right) samples. M3 is the invariant mass distribution of the combination of three jets with the largest magnitude of the vectorial sum of the jet p_T . The shape for the QCD multijet sample is taken from data with a reverse isolation criteria and the sample is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.	88
6.1 The probability density functions, i.e., templates for M3, used in the simultaneous fit. On the left, the shapes of the templates in the sample with three jets are shown. On the right, the template shapes for the sample with at least four jets are shown. The shapes for the QCD multijet events are taken from the orthogonal data sample as described in Section 5.4.1. The shapes of the $t\bar{t}$ events are clearly distinguishable from other processes.	92
6.2 The stability of the fit is studied using 1000 pseudo-experiments. The upper set of plots are the fit value for the total number of $t\bar{t}$ events in the three-jet sample; the lower plots are the fit value for the ratio of the number of four-or-more-jet events to the number of three-jet events ($R_{t\bar{t}}^{4/3}$). The plots in the middle column are the distributions of the fit errors on the number of fitted $t\bar{t}$ events. The plots on the right are the pull distributions fitted separately with a Gaussian distribution to determine the mean and variance. The mean is close to zero, which means the fit itself gives an unbiased central value. However, the variance being lower than one indicates that the errors are overestimated.	93
6.3 The results of the M3 simultaneous fit to the three-jet (top-left figure) and four-or-more-jet (top-right figure) samples. The MC templates are normalized to the results of the fit. The bottom plot shows the visualization of the covariance matrix.	94
6.4 Schematic cartoon of $2 \rightarrow 2$ hard-scattering event.	99
6.5 The interpolated templates for the M3 distribution as a function of α for JES systematic error determination: (a) JES scaled up and (b) JES scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	102
6.6 Selected MC pseudo-experiments and the fit results for the JES systematic error determination: (a) JES scaled up and (b) JES scaled down.	103
6.7 The interpolated templates for the M3 distribution as a function of α for JER systematic error determination: (a) JER scaled up and (b) JER scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	104
6.8 Selected MC pseudo-experiments and the fit results for the JER systematic error determination: (a) JER scaled up (+20%) and (b) JER scaled down (+0%).	105

6.9	The interpolated templates for the M3 distribution as a function of α for the $t\bar{t}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	106
6.10	The interpolated templates for the M3 distribution as a function of α for the $W(Z)+\text{jets}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	107
6.11	Selected MC pseudo-experiments and the fit results for the $t\bar{t}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down.	108
6.12	Selected MC pseudo-experiments and the fit results for the $W(Z)+\text{jets}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down.	109
6.13	The interpolated templates for the M3 distribution as a function of α for the $t\bar{t}$ ME-PS jet matching threshold systematic error determination: (a) 40 GeV/c and (b) 10 GeV/c. The nominal threshold is 20 GeV/c. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	110
6.14	The interpolated templates for the M3 distribution as a function of α for the $W(Z)+\text{jets}$ ME-PS jet matching threshold systematic error determination: (a) 20 GeV/c and (b) 5 GeV/c. The nominal threshold is 10 GeV/c. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	111
6.15	Selected MC pseudo-experiments and the fit results for the $t\bar{t}$ ME-PS jet matching threshold systematic error determination: (a) 40 GeV/c and (b) 10 GeV/c. The nominal threshold is 20 GeV/c.	112
6.16	Selected MC pseudo-experiments and the fit results for the $W(Z)+\text{jets}$ ME-PS jet matching threshold systematic error determination: (a) 20 GeV/c and (b) 5 GeV/c. The nominal threshold is 10 GeV/c.	113
6.17	The interpolated templates for the M3 distribution as a function of α for ISR/FSR systematic error determination: (a) larger ISR/FSR and (b) smaller ISR/FSR. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.	114
6.18	Selected MC pseudo-experiments and the fit results for the ISR/FSR systematic error determination: (a) larger ISR/FSR and (b) smaller ISR/FSR.	115
C.1	Shown in this picture is the RIB in its position on top of a HCAL wedge in the experimental cavern at P5.	136
C.2	Shown in this picture is the RIB in its position in the ECAL bore in the experimental cavern at P5.	136
C.3	Pedestal profiles of module 6 with (read) and without (black) pp1 daughter card. The blue shade displays wings. Module 6 is the module closest to the power supply cable and the ccu.	137
D.1	Workflow diagram of the BeamMonitorDQM.	140

List of Tables

1.1	The Standard Model fermions. There are three fermion generations, each comprising an up-type quark, a down-type quark, a charged lepton, and the corresponding neutrino. Electric charges are given in units of the elementary charge $e = 1.602176487(40) \times 10^{-19}$ C.	4
1.2	Properties of the force mediating bosons. Electric charges are given in units of the elementary charge $e = 1.602176487(40) \times 10^{-19}$ C.	5
2.1	Comparison of the LHC machine parameters between the nominal design luminosity at design center-of-mass energy of 14 TeV and the 2010 peak operation mode for the proton-proton collisions at the center-of-mass energy of 7 TeV.	20
4.1	Data samples used in this study. The combined integrated luminosity is 36 pb ⁻¹ . The uncertainty due to the measurement of the total integrated luminosity is assigned to be about 4% [45] [46].	42
4.2	L1 and HLT triggers used to produce the “Mu” primary dataset for the “2E32” menu.	43
4.3	High Level Trigger paths applied to the primary datasets used in this study. Note in the period of 2010B two HLT paths are used due to the significant increase of the instantaneous luminosity at the LHC. Both HLT paths use the same L1 trigger: “L1_SingleMu7.”	43
4.4	Monte Carlo dataset names from DBS. This analysis was carried out on the AODSIM Monte Carlo samples; the “/AODSIM” is implied and therefore has been stripped from the dataset names.	58
4.5	The cross sections for the relevant Monte Carlo processes. $\ell = e, \mu, \tau$	58
5.1	The event yields for the QCD Monte Carlo sample in the region $\cancel{E}_T > 20$ GeV (“signal” region) and in the region $\cancel{E}_T < 10$ GeV (“background” or “QCD” region).	63
5.2	Monte Carlo cutflow normalized using theoretical cross sections to 36 pb ⁻¹ . Particle-flow jets and \cancel{E}_T are used.	66
5.3	Data cutflow for 36 pb ⁻¹	67

5.4	Muon efficiencies in data and MC, and the scale factors (statistical errors only).	69
5.5	Results of the selection in the QCD region ($\cancel{E}_T < 10$ GeV): the number of events passing the preselection and $\cancel{E}_T < 10$ GeV is denoted by $N_T^{\cancel{E}_T < 10}$. The results for data and the Monte Carlo samples of QCD and $t\bar{t}$ are shown.	73
5.6	ε_{QCD} values measured before subtracting $W(Z)$ contributions, denoted as default, and after subtracting the $W(Z)$ contributions, denoted as $W(Z)$ subtracted, from the $\cancel{E}_T < 10$ GeV region. The values are shown with statistical uncertainties only.	76
5.7	The results of the Matrix Method. The numbers of W -like and QCD multijet events in the candidate sample are determined with the W and Z boson contamination subtracted in the $\cancel{E}_T < 10$ GeV region. The ε_{QCD} value measured for the events with at least two jets is used for the events with two, three and four or more jets. Both statistical and systematic uncertainties are shown for ε_{QCD}	77
5.8	Scale factors for the estimated number of QCD multijet events, denoted SF_{QCD} , and W -like events, denoted SF_{signal} , calculated with respect to Monte Carlo predictions.	77
6.1	Results of simultaneously fitting the M3 distributions.	93
6.2	Summary of the systematic uncertainties in percentages of the top pair production cross section.	101
A.1	Single muon based L1 trigger definition. H_T is the scalar sum of transverse energies of the leptons and the jets, and the missing transverse energy (\cancel{E}_T).	130
A.2	Single muon based HLT trigger definition. Unless specified otherwise, the eta range is $ \eta < 2.5$ for muons and $ \eta < 3.0$ for electrons.	131

Introduction

The [Standard Model \(SM\)](#) of Particle Physics describes how elementary particles interact with each other. All the predictions of this theory have been verified with an astonishing precision except for the existence of the Higgs boson. The spontaneous symmetry breaking provides plausible mechanisms for the gauge bosons (W^\pm and Z) and fermions to acquire mass and results in a scalar field called the Higgs boson. However, the mass of the Higgs boson cannot be predicted. Direct searches at the [Large Electron Positron \(LEP\)](#) collider set a lower limit on the Higgs mass at $114.4 \text{ GeV}/c^2$ at the 95% confidence level (C.L.) [1], and the [Tevatron](#) experiments at [CDF](#) and [DØ](#) exclude the mass range of 158 to $175 \text{ GeV}/c^2$ at the 95% C.L. [2]. Although the Standard Model is by far the most self-consistent theory and has been stringently scrutinized, there are indications that it is not the fundamental theory at the Planck Scale, which is at a much higher energy scale than currently is accessible in the laboratory, but is an effective theory at the electroweak scale. The need of extreme fine-tuning in the theory to account for the large difference between the grand unification scale and the electroweak scale leads to a naturalness problem, the so-called hierarchy problem. In order to solve the hierarchy problem along with other shortcomings, the Standard Model must expand. Since the top quark is the heaviest fermion observed and fermions interact with the Higgs boson through the Yukawa coupling that is

proportional to the fermion mass, the very large mass of the top quark makes it a unique probe into physics at the electroweak scale. With about one top quark pair produced per second at the design luminosity and unprecedented high center-of-mass energy of 14 TeV, the [Large Hadron Collider \(LHC\)](#) will provide a top quark factory that allows detailed studies of top quark properties and searches for new physics in top decays, and eventually may lead to a deeper understanding of electroweak symmetry breaking and the origin of mass. Meanwhile, the precise measurement of the top-antitop ($t\bar{t}$) pair production cross section serves as a crucial test of the overall framework of perturbative [Quantum Chromodynamics \(QCD\)](#) calculations for processes with large momentum transfers in hadronic collisions, including the use of resummation techniques [3] [4] and the parton distribution functions ([PDFs](#)) describing the partonic content of the protons in hard scattering [5].

In this dissertation, a description is given of a measurement of the $t\bar{t}$ pair production cross section at the center-of-mass energy of 7 TeV at the [LHC](#) at the [European Organization for Nuclear Research \(CERN\)](#). The analysis has been completed with a dataset that corresponds to an integrated luminosity of 36 pb^{-1} collected by the [Compact Muon Solenoid \(CMS\)](#) experiment in 2010. This dissertation is structured as follows. It begins with a brief overview of the fascinating Standard Model of particle physics in Chapter 1 followed by a description of the [LHC](#) in Chapter 2. The design of the [CMS](#) experiment is outlined in Chapter 3. Chapter 4 will focus on the data handling and event reconstruction. Chapter 5 gives a detailed description of the event selection and the data analysis. In Chapter 6 the analysis method will be presented in detail, along with the results obtained from this study. Finally the conclusions will be summarized in Chapter 7.

Chapter 1

The Standard Model of Particle Physics

A concise overview of the Standard Model of particle physics is given in Section 1.1, with emphasis on the gauge theory and interactions in Section 1.1.1 and on spontaneous symmetry breaking in Section 1.1.2. It is then followed by an overview of top quark physics in Section 1.2, in which the production mechanism and cross section calculation for $t\bar{t}$ process are discussed in Sections 1.2.1 and 1.2.2, respectively.

1.1 The Standard Model

The Standard Model (SM) [6–11] of particle physics is a relativistic quantum field theory that describes the interactions of the elementary particles. It says that all matter is made up of spin- $\frac{1}{2}$ elementary particles called fermions, which obey Fermi-Dirac statistics and are subject to the Pauli exclusion principle. Note that throughout this chapter the

so-called natural units are employed, in which $\hbar = c = 1$. Fermions are categorized into three generations or families of leptons and quarks. Each generation contains an up-type quark, a down-type quark, a charged lepton, and the corresponding neutrino. There are three known lepton flavors: the electron (e), the muon (μ), the tau (τ), and the three corresponding neutrinos: ν_e , ν_μ , and ν_τ . The electron, muon and tau lepton have similar properties except for the mass. There are six flavors of quarks: the up quark (u), the down quark (d), the charm quark (c), the strange quark (s), the bottom quark (b), and the top quark (t). Each type of particle also has a corresponding antiparticle with identical mass and spin but with opposite quantum numbers, such as charge. The **SM** fermions are summarized in Table 1.1.

Table 1.1: The Standard Model fermions. There are three fermion generations, each comprising an up-type quark, a down-type quark, a charged lepton, and the corresponding neutrino. Electric charges are given in units of the elementary charge $e = 1.602176487(40) \times 10^{-19}$ C.

	Generation	Flavor	Symbol	Charge [e]
Lepton	1	electron neutrino	ν_e	0
		electron	e^-	-1
	2	muon neutrino	ν_μ	0
		muon	μ^-	-1
	3	tau neutrino	ν_τ	0
		tau	τ^-	-1
Quark	1	up	u	+2/3
		down	d	-1/3
	2	charm	c	+2/3
		strange	s	-1/3
	3	top (truth)	t	+2/3
		bottom (beauty)	b	-1/3

Fermions interact with each other through the exchange of spin-1 particles called bosons, which are understood as arising from the so-called gauge symmetry invariance (see Section 1.1.1). Three types of fundamental interactions between the fermions have been experimentally observed. The electromagnetic interaction is mediated by photons (γ), the weak interaction by the massive W^\pm and Z bosons, and the strong interaction is carried by gluons. The forces and the corresponding carriers are summarized in Table 1.2.

Table 1.2: Properties of the force mediating bosons. Electric charges are given in units of the elementary charge $e = 1.602176487(40) \times 10^{-19}$ C.

Type	Symbol	Force	Electric Charge [e]	Mass [GeV/c^2]
gluon	g	strong	0	0
photon	γ	electromagnetic	0	0
W boson	W^\pm	weak	± 1	80.420 ± 0.031
Z boson	Z^0	weak	0	91.188 ± 0.002

1.1.1 Gauge Interactions

In the Standard Model, a free fermion with mass m can be described by the Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \quad (1.1)$$

from which the Dirac equation $(i\cancel{\partial} - m)\psi = 0$ can be derived, where $\cancel{\partial} \equiv \gamma^\mu \partial_\mu$. Consider the local phase transformation:

$$\psi \rightarrow \psi' = U\psi = e^{-i\vec{\alpha}(x) \cdot \vec{\tau}/2}\psi, \quad (1.2)$$

with rotation parameters $\vec{\alpha}(x)$ in an internal space represented by the generators $\vec{\tau}$. The Lagrangian (Eq. 1.1) is no longer invariant. To retain the symmetry in the theory, one can

introduce an interacting vector field, \overrightarrow{A}_μ , such that the space-time derivative, ∂_μ , becomes a covariant derivative (D_μ) by

$$D_\mu = \partial_\mu + ig\frac{\vec{\tau}}{2} \cdot \vec{A}_\mu, \quad (1.3)$$

where g is an arbitrary parameter and \vec{A}_μ transforms as

$$\frac{\vec{\tau}}{2} \cdot \vec{A}'_\mu = -\frac{i}{g}(\partial_\mu U)U^{-1} + U\frac{\vec{\tau}}{2} \cdot \vec{A}_\mu U^{-1}. \quad (1.4)$$

Replacing the space-time derivative in Equation 1.1 by the covariant derivative in Equation 1.3 gives:

$$\begin{aligned} \mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi, \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - g\bar{\psi}\gamma^\mu\frac{\vec{\tau}}{2} \cdot \vec{A}_\mu\psi, \end{aligned} \quad (1.5)$$

which is invariant under the local gauge transformation in Equation 1.2. It is found that the requirement of a theory to be invariant under certain gauge transformations entails the introduction of associated vector fields, called gauge fields.

The **SM** is a non-Abelian Yang-Mills gauge theory based on the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$ where C , L , and Y stand for color quantum number, left-handed chirality (weak isospin) and hypercharge, respectively. The **SM** unifies the electromagnetic and weak interactions, then known as the electroweak interaction, which is described by the $SU(2)_L \times U(1)_Y$ gauge symmetry. The Standard Model Lagrangian can thus be written as

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \quad (1.6)$$

The first two terms, \mathcal{L}_{EW} and \mathcal{L}_{QCD} , describe free fermions, free gauge bosons associated with the $SU(2)_L \times U(1)_Y$ and $SU(3)_C$ gauge symmetries, the interaction between fermions and gauge bosons, and the interactions among gauge bosons themselves. The latter terms \mathcal{L}_{Higgs} and \mathcal{L}_{Yukawa} introduce the Higgs particle and non-zero gauge boson and fermion masses. The covariant derivative, which ensures all three gauge invariances of the theory, takes the form

$$D_\mu = \partial_\mu + ig' \frac{Y}{2} B_\mu + ig \frac{\tau_a}{2} W_\mu^a + ig_s \frac{\lambda_b}{2} G_\mu^b, \quad (1.7)$$

where the scalar Y and the matrices τ_a and λ_b denote the generators for respectively the $U(1)_Y$ hypercharge, the $SU_L(2)$ weak isospin and the $SU(3)_C$ color space.

The Dirac fermion can be represented in terms of left-handed (L) and right-handed (R) chiral states, also known as Weyl fermion fields, as

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}, \quad (1.8)$$

where $\psi_L = \frac{1}{2}(1 - \gamma^5)\psi$ and $\psi_R = \frac{1}{2}(1 + \gamma^5)\psi$. The left-handed fermions, ψ_L , transform as isospin doublets under $SU(2)_L$ gauge transformation, while the right-handed fermions, ψ_R , transform as isospin singlets. The left-handed fermions form isospin doublets Ψ_L :

$$\Psi_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \quad \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \quad \begin{pmatrix} e_L \\ \nu_{e,L} \end{pmatrix}, \quad \begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}, \quad \begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}. \quad (1.9)$$

The interaction of electroweak gauge bosons with the fermions can be described as the coupling of gauge fields to weak isospin and weak hypercharge currents $ig \frac{\tau_a}{2} W_\mu^a$ and $ig' \frac{Y}{2} B_\mu$.

We can then write the electroweak Lagrangian as

$$\begin{aligned}\mathcal{L}_{EW} &= i\bar{\Psi}_L \gamma^\mu D_\mu \Psi_L + i\bar{\Psi}_R \gamma^\mu D_\mu \Psi_R - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^k W^{k\ \mu\nu}, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ W_{\mu\nu}^k &= \partial_\mu W_\nu^k - \partial_\nu W_\mu^k - g\epsilon_{klm} W_\mu^l W_\nu^m \quad [\frac{\tau_i}{2}, \frac{\tau_j}{2}] = i\epsilon_{ijk} \frac{\tau_k}{2},\end{aligned}\tag{1.10}$$

where Ψ_L denotes left-handed fermion doublets and Ψ_R the corresponding right-handed singlets. Taking Pauli matrices as the matrix representation of the $SU(2)_L$ group, we get the interaction terms relating to the weak charged current

$$-\frac{g}{2} \{ \bar{\nu}_{e,L} \gamma^\mu (W_\mu^1 - iW_\mu^2) e_L + \bar{e}_L \gamma^\mu (W_\mu^1 + iW_\mu^2) \nu_{e,L} \} \tag{1.11}$$

and similar terms for muons and taus. The physically observable charged fields are therefore defined by

$$W_\mu^\pm = \frac{1}{2} (W_\mu^1 \mp iW_\mu^2) \tag{1.12}$$

To incorporate the electromagnetic theory (QED), the combination of the remaining gauge fields, W^3 and B , must give rise to the physically observable fields A_μ and Z_μ . Due to the kinetic energy terms for the gauge fields, only orthogonal combination of W^3 and B have independent kinetic energy. Thus, the A_μ and Z_μ can be written as

$$\begin{aligned}A_\mu &= B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W, \\ Z_\mu &= -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W,\end{aligned}$$

where θ_W is the mixing angle called the weak mixing (Weinberg) angle. $\tan \theta_W = g'/g$.

The strong interaction between quarks and gluons is fully described by Quantum Chromodynamics ([QCD](#)), based on the $SU(3)_C$ gauge symmetry, with the associated vector fields denoted by G^μ . The Lagrangian of [QCD](#) is given by [12]

$$\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,a} (i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C G_\mu^C - m_q \delta_{ab}) \psi_{q,b} - \frac{1}{4} F_{\mu\nu}^k F^{k\mu\nu}, \quad (1.13)$$

where repeated indices are summed over. The $\psi_{q,a}$ are the quark-field spinors for a quark of flavor q and mass m_q , with a color index a that runs from 1 to 3. The field tensor is given by

$$F_{\mu\nu}^k = \partial_\mu G_\nu^k - \partial_\nu G_\mu^k - g_s f_{klm} G_\mu^l G_\nu^m \quad [t^A, t^B] = i f_{ABC} t^C, \quad (1.14)$$

where the f_{klm} are the structure constants of the $SU(3)_C$ group, and $t^k = \frac{\lambda_k}{2}$. Since $SU(3)_C$ is a non-Abelian group, there are gluon self interaction terms.

1.1.2 Spontaneous Symmetry Breaking

The requirement of gauge invariance under local $SU(2)_L \times U(1)_Y$ transformation implies that the gauge bosons introduced are massless. The Higgs mechanism, using the idea of spontaneous symmetry breaking, was proposed to solve the shortcomings of the electroweak theory. The mechanism requires the introduction of a complex scalar field ϕ called Higgs field, which is a $SU(2)$ doublet of scalar fields ϕ_i :

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + \phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.15)$$

The gauge invariant Lagrangian for the scalar field is

$$\begin{aligned}\mathcal{L}_{Higgs} &= (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \\ V(\phi) &= \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2,\end{aligned}\tag{1.16}$$

where D_μ is the covariant derivative given in Equation 1.3, excluding the gauge term of $SU(3)_C$, μ^2 is the mass parameter, and $\lambda (> 0)$ is the strength of the Higgs field's self interaction. For $\mu^2 > 0$ the potential $V(\phi)$ has a global minimum for $\phi = 0$. This does not lead to any interesting application. On the other hand when $\mu^2 < 0$, the potential exhibits many non-zero degenerate minima for $\phi^\dagger \phi = (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}$, where v is called the vacuum expectation value (VEV). Choosing arbitrarily a vacuum state at $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$ and making an expansion is made around the minimum:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}.\tag{1.17}$$

Substituting this ϕ in the \mathcal{L}_{Higgs} , one obtains mass terms for the W^\pm and Z bosons, while the photon remains massless:

$$\begin{aligned}(\tfrac{1}{2}vg)^2 W_\mu^+ W^{-\mu}, \quad &\left(\tfrac{1}{2}v\sqrt{g^2 + (g')^2}\right)^2 Z_\mu Z^\mu, \\ M_{W^\pm} = \tfrac{1}{2}vg; \quad M_Z &= M_{W^\pm}/\cos\theta_W.\end{aligned}\tag{1.18}$$

Apart from the mass terms, several couplings arise between the gauge bosons and the field $h(x)$, which is associated to a new scalar particle, the Higgs boson. This new boson has mass $\sqrt{2\lambda}v$, carries no electric charge, has as weak isospin component $-\frac{1}{3}$, and hence as weak

hypercharge $Y = 1$. Since λ is a free parameter, the Standard Model cannot predict the mass of the Higgs boson. The masses of the fermions are obtained through the introduction of additional gauge invariant (Yukawa) terms into the Lagrangian. Taking the electron as an example:

$$\Delta\mathcal{L}_{Yukawa} = -G_e \left[\begin{pmatrix} \bar{\nu}_{e,L} & \bar{e}_L \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \bar{e}_R \begin{pmatrix} \phi^- & \bar{\phi}^0 \end{pmatrix} \begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix} \right]. \quad (1.19)$$

Choose G_e such that $m_e = G_e v / \sqrt{2}$, and use ϕ in Equation 1.17, then $\Delta\mathcal{L}_{Yukawa} = -m_e \bar{e}e - \frac{m_e}{v} \bar{e}eh$, where the second term describe its coupling with the Higgs. The direct searches at the LEP collider set a lower limit on the Higgs mass at 114.4 GeV/c² at the 95% C.L. [1], and the Tevatron experiments at CDF and DØ exclude the mass range of 158 to 175 GeV/c² at the 95% C.L. [2].

1.2 Top Quark Physics

Ever since the inference of the bottom quark from the discovery of the Upsilon family of resonances at Fermilab in 1977, the search for the top quark, which was suggested by the SM as the weak isospin partner of the b quark, was pursued by experimentalists immediately. However, scientists were not equipped with an accelerator powerful enough to create such particles until the Fermilab Tevatron Collider came into play. The SM claimed yet another triumph when the top quark was discovered at the Tevatron in 1995 [13] [14], with a measured mass about 40 times heavier than the b quark. The current world average of the measured top quark mass is 173.3 ± 1.1 GeV/c² [15]. According to the SM, the

top quark decays almost exclusively into a b quark and a W boson, which subsequently decays into leptons ($\ell^-\bar{\nu}_l : \ell = e, \mu, \tau$), referred to as the leptonic decay, or quarks ($q\bar{q}' : q = u, c; q' = d, s$), referred to as the hadronic decay. The top-quark decay rate, including first order QCD corrections, is given by

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} |V_{tb}|^2 \left(1 - \frac{M_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_t^2}\right) \left[1 - \frac{2\alpha_s}{3\pi} \cdot f(y)\right], \quad (1.20)$$

with $y = (M_W/m_t)^2$ and $f(y) = 2\pi^2/3 - 5/2 - 3y + 9/2y^2 - 3y^2 \ln y$ [16–18]. The width for $m_t = 171$ GeV/ c^2 is 1.29 GeV (using $\alpha_s = 0.118$) and increases with mass [12]. Its corresponding lifetime ($\approx 5 \times 10^{-25}$ s) is shorter than the characteristic formation time of hadrons, so that the top quark decays so quickly that it cannot form bound states (hadrons) with other quark(s). Therefore, the observation of the top quark relies on precise reconstruction of the final state of its decay products.

At hadron colliders, top quark production is dominated by the production of top-antitop ($t\bar{t}$) pairs [19]. Therefore, the final state of $t\bar{t}$ events is usually categorized in terms of the decay mode of the W bosons. The final state is referred to as the “dilepton” channel if both W bosons decay leptonically, and as the “all-jets” channel if both W bosons decay into jets, where jets are the physical observables of the hadronization of quarks or gluons [20]. If one W boson decays leptonically and the other hadronically, it is identified as the “semileptonic” or “lepton+jets” channel; depending on the lepton flavor of the leptonic W boson decay, the categorization can be further categorized as “electron+jets” ($e+$ jets), “muon+jets” ($\mu+$ jets), and tau+jets ($\tau+$ jets). The dilepton channel is usually classified into sub-categories such as ee +jets, $e\mu$ +jets, and $\mu\mu$ +jets. The branching fractions are

roughly 1/9 for the dilepton channel, 4/9 for the lepton+jets channel, and 4/9 for the all-jets channel. The presence of two leptons provides an effective criterion to distinguish signal from backgrounds; however, there are also two neutrinos from the W boson decays that cannot be detected. Since one can reconstruct only a missing transverse energy (\cancel{E}_T) for such events, this makes the reconstruction of W bosons less trivial as well as the top quark reconstruction. On the contrary, the all-jets channel events exhibit no \cancel{E}_T , but the disadvantage is the large number of jets in the final state, which results in huge combinatorial backgrounds that complicate the extraction of the signal. While for the lepton+jets channel, events contain one energetic lepton and one \cancel{E}_T yielding cleaner W boson reconstruction and lower jet multiplicity (compared to all-jets channel) greatly reducing the combinatorics. Meanwhile, the muon reconstruction has higher efficiency and purity than the electron and tau reconstructions. We therefore focus on the μ +jets channel in this study.

1.2.1 Production Mechanism

At the [LHC](#), top quark pair production is dominated by gluon-gluon fusion ($\approx 90\%$) and quark-antiquark annihilation ($\approx 10\%$) [21]. The Leading Order (LO) differential cross section for $t\bar{t}$ production via $q\bar{q}$ annihilation is given by [12] [22]:

$$\frac{d\hat{\sigma}}{d\hat{t}}(q\bar{q} \rightarrow t\bar{t}) = \frac{4\pi\alpha_s^2}{9\hat{s}^4} \cdot \left[(m_t^2 - \hat{t})^2 + (m_t^2 - \hat{u}^2)^2 + 2m_t^2\hat{s} \right], \quad (1.21)$$

and for gluon-gluon fusion by:

$$\begin{aligned} \frac{d\hat{\sigma}}{d\hat{t}}(gg \rightarrow t\bar{t}) &= \frac{\pi\alpha_s^2}{8\hat{s}^2} \cdot \left[\frac{6(m_t^2 - \hat{t})(m_t^2 - \hat{u})}{\hat{s}^2} - \frac{m_t^2(\hat{s} - 4m_t^2)}{3(m_t^2 - \hat{t})(m_t^2 - \hat{u})} \right. \\ &\quad + \frac{4}{3} \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) - 2m_t^2(m_t^2 + \hat{t})}{(m_t^2 - \hat{t})^2} + \frac{4}{3} \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) - 2m_t^2(m_t^2 + \hat{u})}{(m_t^2 - \hat{u})^2} \\ &\quad \left. - 3 \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) + m_t^2(\hat{u} - \hat{t})}{\hat{s}(m_t^2 - \hat{t})} - 3 \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) + m_t^2(\hat{t} - \hat{u})}{\hat{s}(m_t^2 - \hat{u})} \right], \end{aligned} \quad (1.22)$$

where m_t denotes the top-quark mass and \hat{s} , \hat{t} , and \hat{u} are the Lorentz-invariant Mandelstam variables of the partonic process. For quark-antiquark annihilation (Figure 1.2(a)), they are defined by $\hat{s} = (p_q + p_{\bar{q}})^2$, $\hat{t} = (p_q - p_t)^2$, and $\hat{u} = (p_q - p_{\bar{t}})^2$. For gluon-gluon fusion (Figure 1.2(b)), they are defined by $\hat{s} = (p_{g_1} + p_{g_2})^2$, $\hat{t} = (p_{g_1} - p_t)^2$, and $\hat{u} = (p_{g_1} - p_{\bar{t}})^2$.

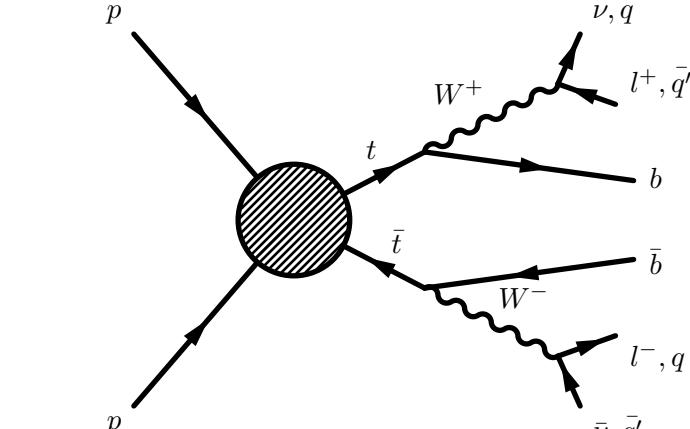
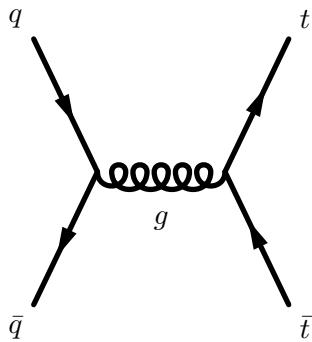


Figure 1.1: Feynman diagram for the $pp \rightarrow t\bar{t}$ process.

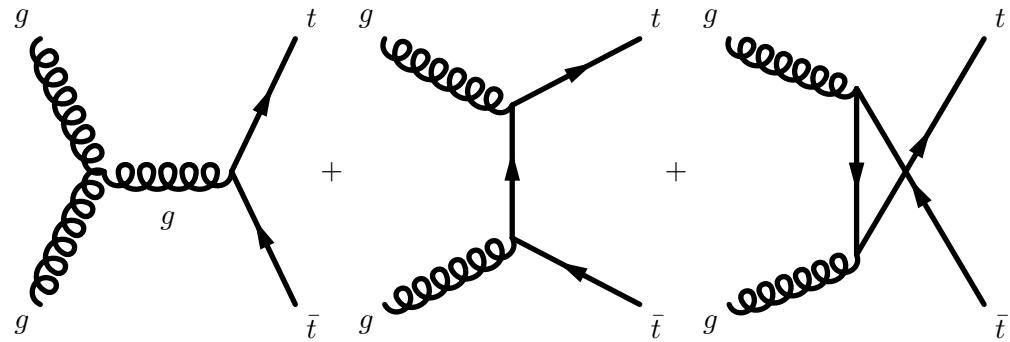
1.2.2 Production Cross Section

According to perturbative QCD (pQCD), inelastic proton-proton (pp) scattering can be described by the direct interactions between constituent partons (i.e., quarks and gluons). The two partons entering the hard scattering carry fractions x_1 and x_2 of the proton momenta. The PDF, $f_{i/p}(x, \mu^2)$, describes the probability density for a parton of type i to carry a fractional momentum x and squared four momentum (μ^2 , or virtuality). The inclusive cross section for $t\bar{t}$ production in pp collisions can be written as [12]:

$$\begin{aligned} \sigma(pp \rightarrow t\bar{t} + X) &= \sum_{n=0}^{\infty} \alpha_s^n(\mu_R^2) \sum_{i,j} \int dx_1 dx_2 f_{i/p}(x_1, \mu_F^2) f_{j/p}(x_2, \mu_F^2) \\ &\quad \times \hat{\sigma}_{ij \rightarrow t\bar{t} + X}^{(n)}(x_1 x_2 s, \mu_R^2, \mu_F^2), \end{aligned} \quad (1.23)$$



(a)



(b)

Figure 1.2: Feynman diagrams of the LO processes for $t\bar{t}$ production via strong interactions: (a) quark-antiquark annihilation; (b) gluon-gluon fusion.

where s is the squared center-of-mass energy of the collision, μ_F is the factorization scale, and μ_R the normalization scale.

Chapter 2

The Large Hadron Collider

As discussed in the previous chapter, it is necessary to explore the TeV energy scale regime to reveal secrets to most fundamental questions of particle physics. The design of the Large Hadron Collider ([LHC](#)) [23] [24] has been driven by the goal to explore physics at the TeV scale.

The [LHC](#) is a two-ring superconducting hadron accelerator and collider located at [CERN](#) near Geneva, Switzerland. It was designed to collide protons at a center-of-mass energy of 14 TeV with a instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Also, the [LHC](#) is capable of colliding heavy ions in dedicated runs; with nominal dipole field the [LHC](#) will be able to accelerate ionized lead nuclei (Pb^{82+}) to an energy of 2.76 TeV per nucleon and a peak luminosity of $10^{27} \text{ cm}^{-2}\text{s}^{-1}$. The schematic layout of the [LHC](#) is shown in Fig. 2.1. The accelerator rings, each constituted by eight arcs interconnected by eight straight sections where detectors, Radio Frequency (RF) cavities, and other facilities are accommodated, are installed in the existing 26.7 km tunnel that was constructed and used for the [LEP](#) machine. There are four main experiments established in different straight sections of the [LHC](#). The

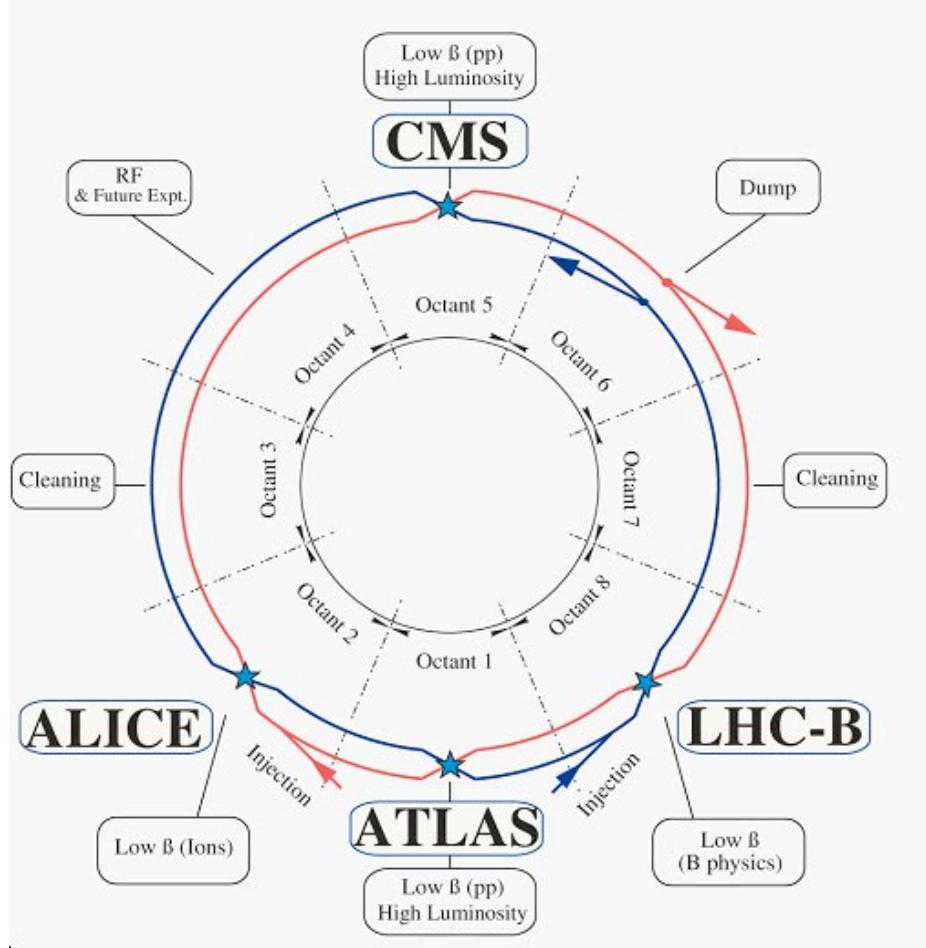


Figure 2.1: Schematic layout of the LHC (Beam 1 - clockwise, Beam 2 - counterclockwise)

two general purpose detectors ATLAS (A Toroidal LHC Apparatus) [25] and CMS [26] are installed at the two high luminosity interaction points Point 1 (P1) and Point 5 (P5), respectively. The LHC also hosts two special-purpose detectors located at the experimental sections in Octant 2 and Octant 8: Alice (A Large Ion Collider Experiment), located at Point 2 (P2), is designed to study heavy ions collisions exclusively; LHCb (LHC beauty), located at Point 8 (P8), is an experiment dedicated to the study of CP violation and rare decays in the range of b -quark physics arising from pp collisions.

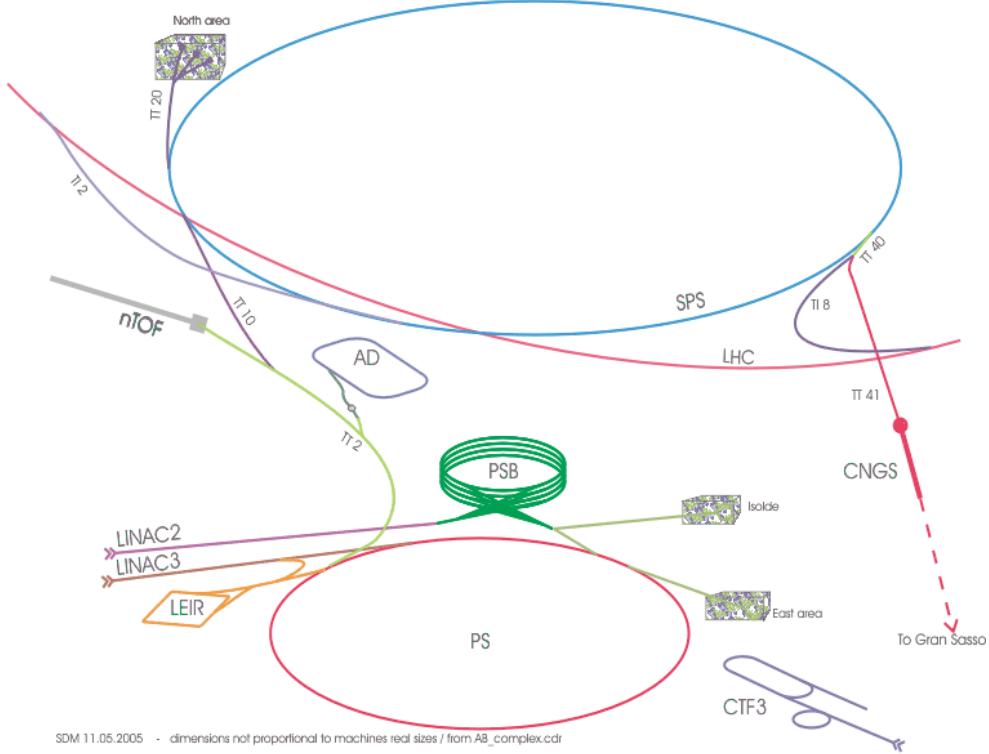


Figure 2.2: Schematic of the CERN accelerator complex.

Before being accelerated to the colliding energy by the [LHC](#), the protons are delivered by the Super Proton Synchrotron (SPS) at an injection energy of 450 GeV through the two injection tunnels (approximately 2.5 km in length) linking the [LHC](#) to the [CERN](#) accelerator complex (Figure 2.2). Before entering the SPS, the protons are accelerated to 26 GeV by the Proton Synchrotron (PS), which is fed with the 1.4 GeV protons accelerated by the Proton Synchrotron Booster (PSB). The beam injection into the different acceleration facilities is carried out by special fast-ramping magnets called kickers. The various rise times of these injection kickers introduce gaps in the beams. This results in a total of 2,808 bunches of 1.15×10^{11} protons filling the 3,564 available RF buckets per LHC beam at the nominal designed operation mode.

Searches for rare phenomena and measurements with increasing precision require ever increasing interaction rates. The acceleration in colliders is performed in bunches with a high particle density. The interaction rate is proportional to the luminosity \mathcal{L} of the accelerator, which is given for head-on collisions by

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F, \quad (2.1)$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta function at the collision point, and F the geometric luminosity reduction factor due to the crossing angle at the interaction point:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right)^{-1/2}, \quad (2.2)$$

where θ_c is the full crossing angle at the IP, σ_z the RMS bunch length, and σ^* the transverse RMS beam size at the IP. The above expression assumes round beams and with equal beam parameters for both beams. The exploration of rare events in the **LHC** collisions therefore requires both large beam energies and large beam intensities. The **LHC** has been devised to operate at the center-of-mass energy half of the design value for *pp* collisions [27] [28]. The machine parameters are summarized in Table 2.1. Figure 2.3 shows the performance of the machine and about 92% efficiency for the **CMS** data taking [29].

Table 2.1: Comparison of the LHC machine parameters between the nominal design luminosity at design center-of-mass energy of 14 TeV and the 2010 peak operation mode for the proton-proton collisions at the center-of-mass energy of 7 TeV.

Quantity	Design	2010	Operation
Luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	10^{34}	2×10^{32}	
Number of bunches per proton beam n_b	2,808		368
Number of protons per bunch N_b	1.15×10^{11}	1.15×10^{11}	
Normalised transverse emittance $\epsilon_n [\mu\text{m}]$	3.75		1.6
Amplitude function $\beta^*[m]$	0.55		3.5
Crossing angle at interaction point $\theta_c [\mu\text{rad}]$	285		100
Nominal bunch separation [ns]	25		75
Nominal energy of protons [GeV]	7,000		3,500
Peak magnetic dipole field [T]	8.33		4.17

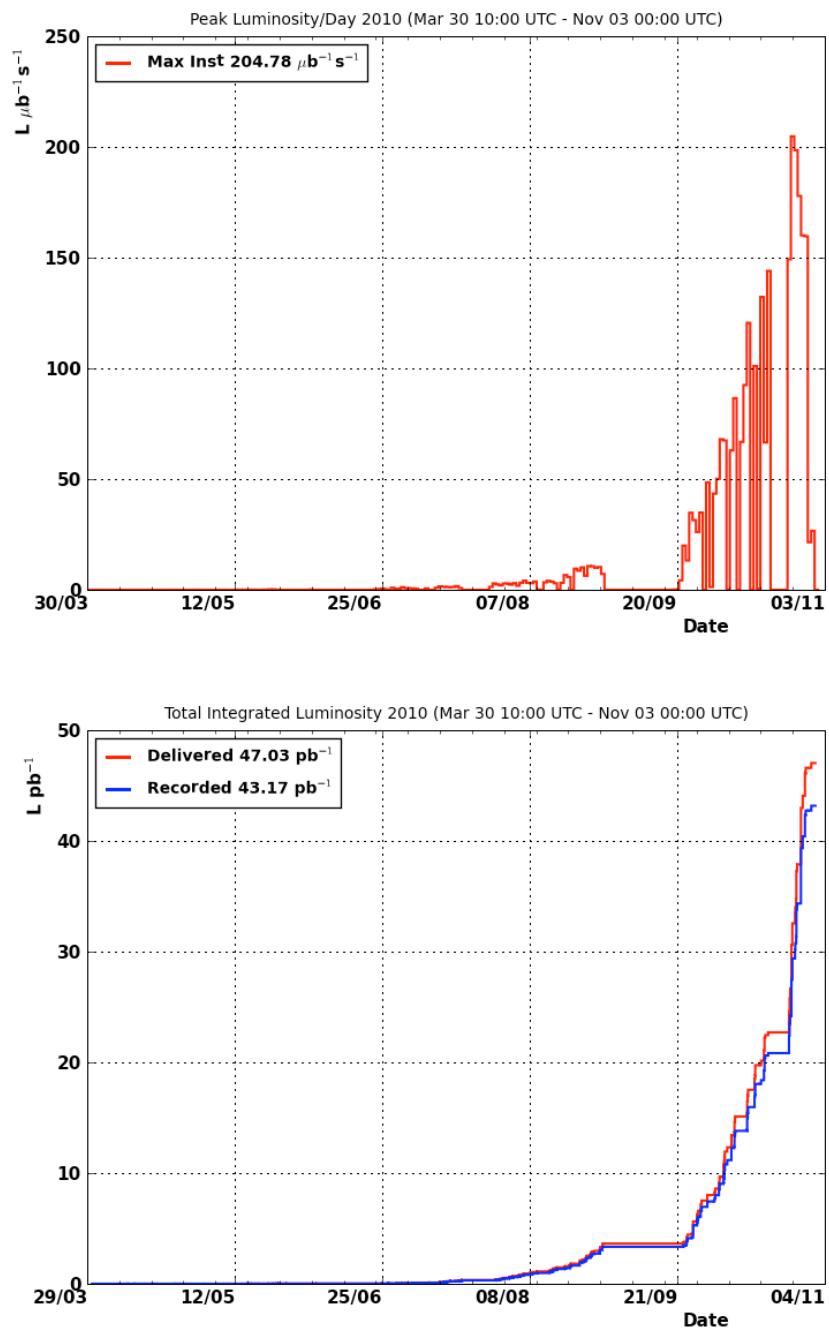


Figure 2.3: Instantaneous and integrated luminosity of the LHC: Left: The instantaneous luminosity as a function of time. The peak values is $2.0 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$. Right: The integrated luminosity as a function of time.

Chapter 3

The CMS Detector

The general purpose Compact Muon Solenoid ([CMS](#)) detector has been designed to perform various particle measurements to reconstruct physics events from the pp collisions in as much detail as possible. The objective is to capture the ephemeral phenomena of the rich and diverse physics program at the [LHC](#). The overall layout of [CMS](#) is shown in Fig. 3.1. The [CMS](#) detector is about 21.6 m long and 14.6 m in diameter, weighs roughly 12,500 tons, and is composed of several state-of-the-art subdetectors that are immersed in a strong magnetic field provided by the superconducting solenoid. The magnet provides up to 3.8 T of uniform magnetic field inside its bore, which accommodates the inner tracking system and the calorimetry. Outside the solenoid is the muon spectrometer. The CMS coordinate system is defined in Section 3.1, and the designs of the separate subdetectors are sketched out in the following sections (the tracking system in Section 3.2, the calorimetry in Section 3.3, the muon system in Section 3.4, and the trigger system in Section 3.5). The corresponding low level object reconstructions, which provide the building blocks for the high level physics object reconstructions are described in Chapter 4. A more detailed

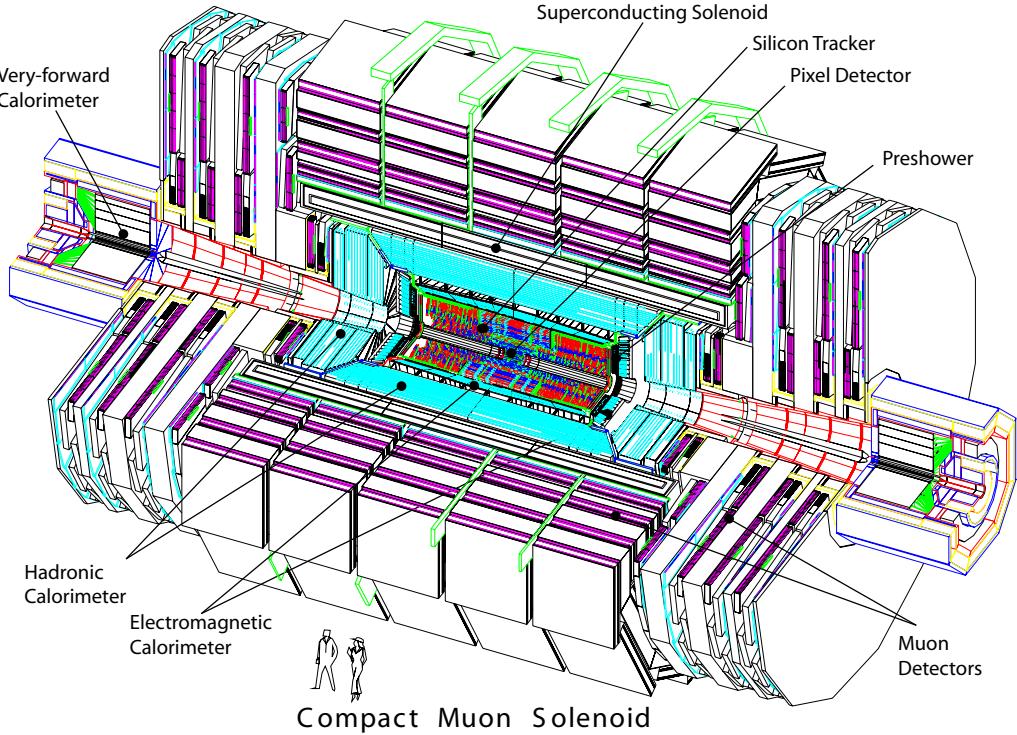


Figure 3.1: The Compact Muon Solenoid (CMS) layout.

description of the technical design has been published [26].

3.1 The CMS Coordinates

The CMS reference frame is centered at the [Interaction Point of CMS \(IP5\)](#) of the [LHC](#). With respect to this point a right-handed Cartesian coordinate system is defined, where the positive z direction is chosen along the direction of the proton beam 2 of the [LHC](#) (see Fig. 2.1), the x axis points radially inward toward the center of the [LHC](#) ring, and the y axis points upward perpendicular to the ground.

As described in Section 1.2.1, the proton-proton collision is inelastic scattering between the constituent partons that in general carry different energies. Hence, the rest

frame of the hard scattering is boosted along the beamline with respect to the lab frame. Because it is impossible to reconstruct the boost due to the loss of key information, such as the remnants of the colliding protons carried away along the beam pipe, it is more convenient to introduce the coordinate system (r, η, ϕ) , where r is the radial distance, ϕ is the azimuthal angle in the transverse xy-plane measured with respect to the +x axis, and the pseudorapidity η is defined as

$$\eta = -\ln \tan \frac{\theta}{2}, \quad (3.1)$$

where θ is the polar angle measured with respect to the +z axis. In the high energy limit $m/E \rightarrow 0$, η approaches the rapidity (y) of the particle.

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \approx \eta. \quad (3.2)$$

For a longitudinal Lorentz boost along the z axis to a frame with velocity β , $y \rightarrow y + \tanh^{-1} \beta$. Therefore, not only transverse quantities but differences in η (y) are invariant under the boost. As a consequence, a solid angle ΔR in (η, ϕ) space is also invariant.

$$\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}. \quad (3.3)$$

3.2 The CMS Silicon Tracker

The CMS inner tracking system consists of two major subsystems, the inner [pixel detector \(PIXEL\)](#) and the [silicon strip tracker \(SST\)](#). Together they provide robust and

precise measurements of vertices and tracks of the charged particles produced in the pp collisions. The all silicon tracker design is a realization of balancing many competing factors, such as scientific objectives (i.e., detector performance and robustness) and budgetary constraints [26] [30] [31]. A schematic drawing of the CMS tracker is shown in Fig. 3.2. The CMS tracker covers about 200 m^2 of active silicon sensors that are populated in the cylindrical volume with a radius of 116 cm and a length of 540 cm; it provides tracking coverage up to $|\eta| < 2.5$.

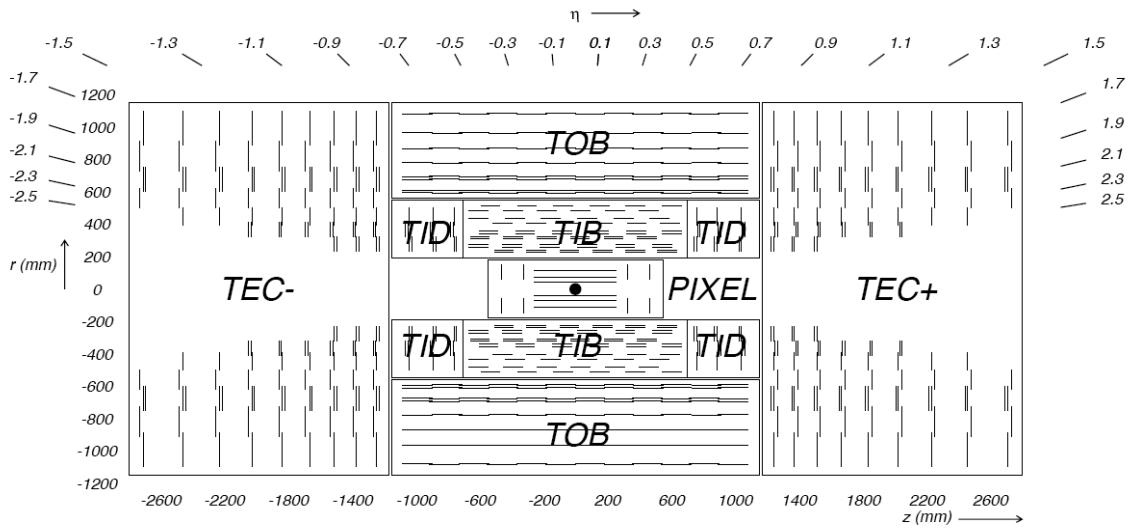


Figure 3.2: Schematic cross sectional view of the CMS tracker. Each line represents a detector module. Double lines indicate double sided silicon strip modules mounted back-to-back with a stereo angle of 100 mrad to deliver stereo hits.

3.2.1 The Pixel Detector

As shown in Fig. 3.2 the pixel detector (PIXEL) is the subdetector system closest to the interaction region. It is designed to deliver precise tracking points (hits) on charged particle trajectories in $r - \phi$ and z , with a spatial resolution in the range of 15-20 μm .

The pixel detector system has a total of 1440 pixel detector modules with about 66 million pixels (each cell is $100 \times 150 \mu\text{m}^2$) and covers an area of about 1 m^2 . It contains two main partitions, the [barrel pixel detector \(BPix\)](#) and the [forward pixel detectors \(FPix\)](#); the layout is shown in Fig. 3.3. The [BPix](#) contains three 53 cm long cylindrical layers of hybrid pixel detector modules at radii of 4.4, 7.3 and 10.2 cm that surround the beamline; it is complemented by two disks ([FPix](#)), which extend from ≈ 6 to 15 cm in radius, of pixel modules on each side at $|z| = 34.5$ and $|z| = 46.5$ cm.

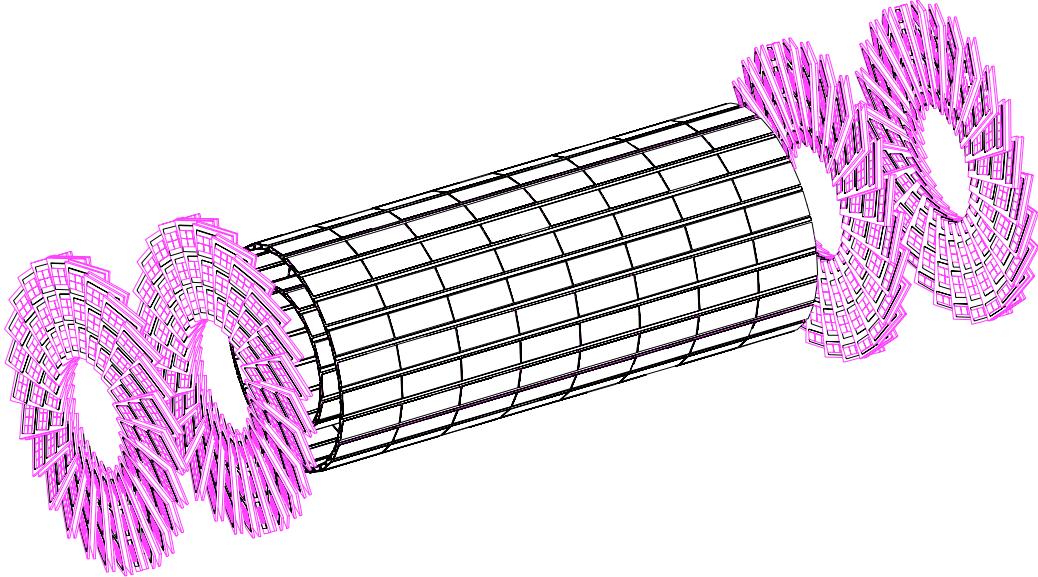


Figure 3.3: The layout of the pixel detector. It consists of three barrel layers (BPix) with two endcap disks (FPix).

3.2.2 The Silicon Strip Tracker

The CMS silicon strip tracker has 15,148 silicon strip detector modules with a total of about 9.3 million strips (channels) and 198 m^2 of active silicon area. It is composed of three different subsystems that extend in the radial region between 20 cm and 116 cm (see Fig. 3.2). The inner part consists of the [Tracker Inner Barrel \(TIB\)](#), which has four layers

of $320\ \mu\text{m}$ thick silicon micro-strip sensors (referred to as “thin” sensors) with their strips parallel to the beam axis, and the **Tracker Inner Disks (TID)**, which complete each end of the **TIB** by three disks of thin sensors with their strips fanned out in the radial direction. The modules in the first two ¹ layers of the **TIB** carry a second module which is mounted back-to-back with a stereo angle of 100 mrad; we call these “double-sided” modules. Also, the strip pitch is $80\ \mu\text{m}$ in layers 1 and 2 and $120\ \mu\text{m}$ in layers 3 and 4, resulting in a single point resolution of $23\ \mu\text{m}$ and $35\ \mu\text{m}$ in $r - \phi$, respectively. The double-sided modules in layers 1 and 2 deliver a single point resolution of $230\ \mu\text{m}$ in z as well. Each **TID** disk contains 3 rings of strip sensors with mean pitch varying between $100\ \mu\text{m}$ and $141\ \mu\text{m}$, with the first two rings populated with double-sided modules.

The **TIB/TID** is surrounded by the **Tracker Outer Barrel (TOB)**, which covers the radial region between 55 cm and 116 cm and extends in z between ± 118 cm. To achieve better signal to noise performance while limiting the number of readout channels, the **TOB** is populated with 6 longer layers (strip length of 183.2 mm compared to 116.9mm in **TIB/TID**) of $500\ \mu\text{m}$ thick silicon strip sensors (referred to as “thick” sensors). The strip pitch is $183\ \mu\text{m}$ for layers 1 to 4 and $122\ \mu\text{m}$ for layers 5 and 6, leading to a single point resolution of $53\ \mu\text{m}$ and $35\ \mu\text{m}$ in $r - \phi$, respectively. In addition, the first two layers of the **TOB** are populated with double-sided modules and therefore provide a resolution of $530\ \mu\text{m}$ in z as well.

The **Tracker EndCaps (TEC)** consist of two partitions that cover the region $124\ \text{cm} < |z| < 282\ \text{cm}$ and $22.5\ \text{cm} < r < 113.5\ \text{cm}$; the partition residing in the positive (negative) z region is referred to as TEC+ (TEC-). Each partition consists of 9 disks carrying up to

¹Unless specified otherwise, we will always count outward from the interaction region.

7 rings of silicon strip sensors (thin sensors on the inner 4 rings, thick sensors on rings 5-7) with radial strips of 97 μm to 184 μm average pitch. As shown in Fig. 3.2, disks 1-3 each contain 7 rings, disks 4-6 each contain ring 2 to 7, disks 7 and 8 each contain rings 3 to 7, and disk 9 contains only rings 4 to 7. Meanwhile, the rings 1, 2, and 5 are populated with double-sided modules that provide measurements of r as well.

3.3 The CMS Calorimeter System

The CMS detector is equipped with two calorimeters to perform energy measurements of particles leaving the tracking volume. The [Electromagnetic Calorimeter \(ECAL\)](#) [32], which surrounds the inner tracking system, serves the purpose of measuring the energy of electrons and photons, and the [Hadronic Calorimeter \(HCAL\)](#) [33], which is placed outside the [ECAL](#), is designed to measure charged- and neutral-hadron energies. Both of the calorimeter systems are located inside the superconducting solenoid to reduce energy loss due to magnet materials.

3.3.1 The Electromagnetic Calorimeter

The [ECAL](#) absorbs the energies of the electrons and photons traversing it and produces signals proportional to those energies. It is finely subdivided so that it can measure the directional dependence of the electromagnetic energy. There are three main parts in the [ECAL](#) system: [ECAL Barrel \(EB\)](#), [ECAL Endcaps \(EE\)](#), and [Preshower Detector \(ES\)](#). Lead tungstate (PbWO_4) scintillating crystals have been chosen as the detection material for the [ECAL](#) because of the high density (8.28 g/cm³), short radiation length ($X_0 = 0.89$ cm), short scintillation decay time (about 80% light is emitted in 25 ns), and

radiation hardness [12]. The tapered-shape crystals in the **EB** have a cross section of approximately 0.0174×0.0174 in the $(\eta - \phi)$ plane or $22 \times 22 \text{ mm}^2$ at the front face, and a $26 \times 26 \text{ mm}^2$ cross section at the rear face. The crystal length is 23 cm corresponding to 25.8 X_0 . A total of 61,200 lead tungstate crystals mounted in the central barrel part (**EB**) cover up to $|\eta| < 1.479$. The barrel crystal volume is 8.14 m^3 and weighs 67.4 t. The **ECAL** is closed by the two endcaps (**EE**) each of which is mounted with 7,324 crystals and covers the range $1.479 < |\eta| < 3.0$. The **EE** crystals have a front face cross section of $28.6 \times 28.6 \text{ mm}^2$, a rear face cross section of $30 \times 30 \text{ mm}^2$, and a length of 22 cm corresponding to 24.7 X_0 . The crystals in the **EE** are grouped in mechanical units of 5×5 supercrystals (or SCs). The total **EE** crystal volume is 2.90 m^3 and weights 24.0 t. To avoid cracks aligned with the particle trajectory, the **EB** and **EE** crystals are mounted in such a way that their axes make a 3° and 2° to 8° angle, respectively, with respect to the vector from the nominal interaction vertex, in both the η and ϕ projections. An ECAL preshower (**ES**) is placed on the inner side of each **EE** with the aim to identify neutral pions in the endcaps within a fiducial region $1.653 < |\eta| < 2.6$, which helps the identification of electrons against minimum ionizing particles (MIP), and improves the position determination of electrons and photons with high granularity. There are around 4,300 micromodules, each containing a silicon sensor with a $61 \times 61 \text{ mm}^2$ active area (32 strips) and a thickness of $320 \mu\text{m}$ [34].

The transverse view in the (r, z) plane of the **ECAL** is shown in Figure 3.4.

When electrons, positrons or photons hit the **ECAL** crystals they create electromagnetic showers. The number of produced particles in the electromagnetic shower caused by the incident particle, such as an electron or photon, is proportional to the energy of the incoming particle. Energy deposits within the scintillating crystals create a flash of

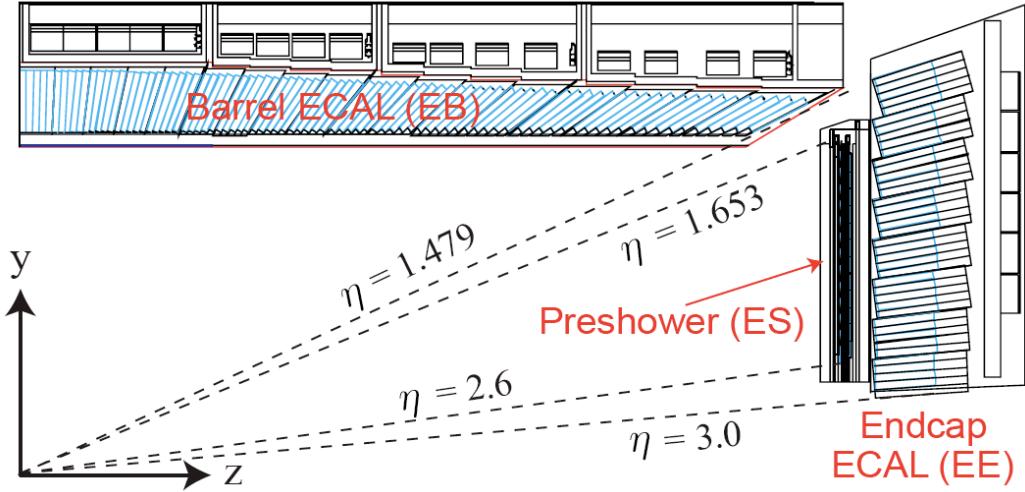


Figure 3.4: Transverse view in the (r, z) plane of a quarter of the ECAL layout.

blue-green light (420-430 nm [35] [36]), which is measured by silicon avalanche photo-diodes (APDs) in the barrel part and vacuum phototriodes (VPTs) in the endcaps. At 18°C about 4.5 photoelectrons per MeV are collected in both APDs and VPTs. The energy resolution for the calorimeter can be parameterized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \quad (3.4)$$

where E is in GeV, S is the stochastic term, N the noise term, and C the constant term. They are measured (with electron beams having momenta between 20 and 250 GeV/c) to be $0.028 \text{ GeV}^{\frac{1}{2}}$, 0.12 GeV , and 0.3% , respectively [37].

3.3.2 The Hadronic Calorimeter

The **HCAL**, which surrounds the **ECAL**, plays an important role in measuring the energies of hadrons, including protons, neutrons, pions and kaons (electrons and photons

are stopped by the **ECAL** materials); it also serves a crucial purpose to determine the missing transverse energy in events containing neutrinos or exotic particles. The CMS **HCAL** includes the **HCAL Barrel (HB)**, the **HCAL Endcaps (HE)**, the **Outer Hadronic Calorimeter (HO)** and the **Forward Hadronic Calorimeter (HF)**. The central barrel part (**HB**) covering the pseudorapidity range $|\eta| < 1.3$ is closed by the two endcaps (**HE**), which cover the range $1.3 < |\eta| < 3.0$. Since they are located inside the superconducting magnet coil, a dense, non-magnetic, short interaction length ($\lambda_I = 16.42$ cm), and radiation hard absorber material, known as C26000 cartridge brass, is chosen to build the calorimetry. The **HB** is composed of two half-barrels (**HB+** and **HB-**), each consisting of 18 identical wedges constructed of flat brass absorber plates aligned parallel to the beam axis. Each wedge is segmented into four azimuthal sectors. The plastic scintillator system is divided into 16 η sectors, resulting in a segmentation of 0.087×0.087 in the (η, ϕ) plane. The scintillator trays of the **HE** provide energy measurement with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ for $1.6 \leq |\eta| < 3.0$.

In the central pseudorapidity region, the combined stopping power of **EB** and **HB** does not provide sufficient containment for deeply penetrating or late initiated hadron showers. To ensure adequate sampling depth for $|\eta| < 1.3$, the **HCAL** is extended outside the superconducting solenoid (serving as additional absorbing material) with the outer hadronic calorimeter (**HO**), which covers the region $|\eta| < 1.26$. The **HO** serves to detect the tails of the hadronic shower and aids in the missing transverse energy measurement by sampling the energy leakage of deeply penetrating hadron showers. The **HO** extends the minimum effective absorber thickness to $11.8 \lambda_I$. The **HF** is designed as a cylindrical steel absorber structure with an inner radius of 12.5 cm and an outer radius of 130.0 cm;

it provides coverage of the pseudorapidity range $3.0 < |\eta| < 5.2$. The front face of the forward calorimeters is located at ± 11.2 m from the nominal interaction point. Each **HF** is azimuthally subdivided into 18 wedges; each wedge is further subdivided into two halves. The active elements of **HF** (quartz fibers) run parallel to the beam axis and are bundled into 13 eta to form towers of $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$

The hadronic energy resolution of the combined barrel **HCAL** and **ECAL** is parameterized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + C^2, \quad (3.5)$$

where E is in GeV, S corresponds to a stochastic term, and C to a constant term, one measures $S = 0.847 \pm 0.016$ GeV $^{\frac{1}{2}}$ and $C = 0.074 \pm 0.008$. The energy resolution in the endcaps is similar to that in the barrel and the corresponding values are measured to be $S = 1.98$ GeV $^{\frac{1}{2}}$ and $C = 0.09$ [38].

3.4 The CMS Muon Spectrometer

The CMS Muon Spectrometer [39] is designed to precisely reconstruct the momentum and charge of muons over the entire kinematic range of the **LHC** and provides a full geometric coverage for muon identification up to $|\eta| = 2.4$. The muon system is capable of providing quick response to the passage of muons and plays an important role in triggering. The muon stations are integrated within the flux return iron yoke, which also serves as a hadron absorber improving the muon identification; high magnetic field (about 2T) in the return yoke enables the muon system to have good momentum resolution and triggering

capabilities. The CMS detector utilizes three types of gaseous particle detectors, which combine about $25,000 \text{ m}^2$ of detection planes, for muon identification: **Drift Tube (DT)**, **Cathode Strip Chamber (CSC)**, and **Resistivity Plate Chamber (RPC)**. The muon system comprises a barrel part and two planar endcaps. The layout of the muon system is shown in Figure 3.5.

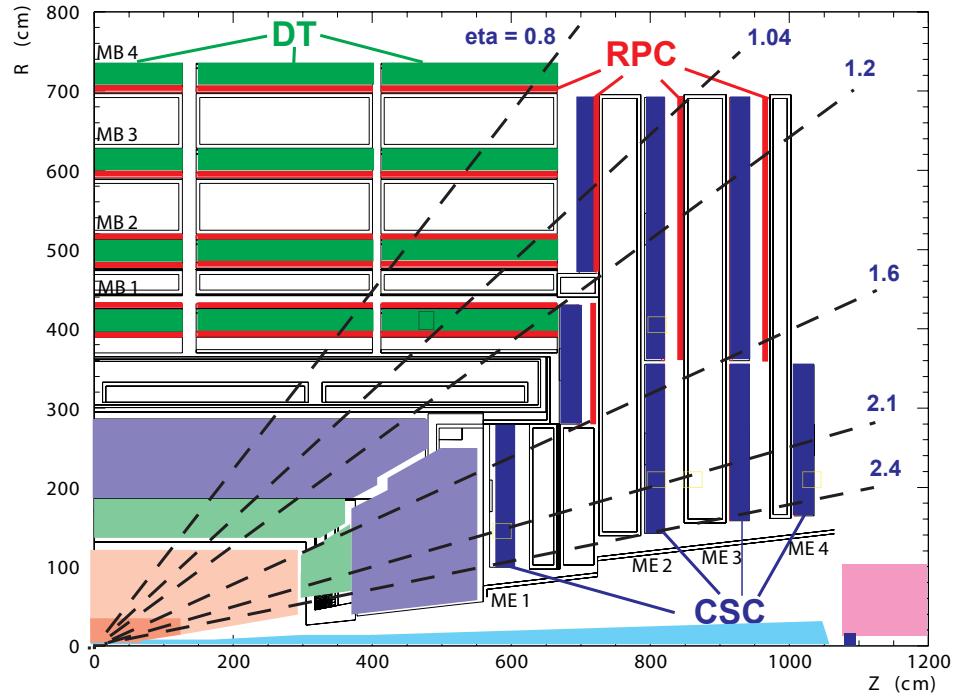


Figure 3.5: Transverse $r - z$ view of a quarter of the CMS muon spectrometer. The barrel region is integrated with Drift Tube (DT) chambers and Resistive Plate Chambers (RPCs), while the endcap region is equipped with Cathode Strip Chambers (CSCs) and RPCs. In total four barrel muon stations, MB 1 to MB 4, and four endcap muon stations, ME 1 to ME 4, are installed.

3.4.1 The Drift Tube Chambers

There are 250 DT chambers in the barrel region, arranged concentrically around the beam line, and covering the pseudorapidity range $|\eta| < 1.2$. The barrel is divided in 5

wheels, each has four layers of **DT** stations, MB 1 to MB 4, interspersed among the layers of the iron yoke plates. The first three **DT** stations each consist of 12 layers of rectangular **DT** cells, which are further grouped into 3 superlayers (SLs) each containing 4 layers of rectangular **DT** cells staggered by half a cell. As shown in Figure 3.6, the two outer SLs are used for the measurement of muon trajectories in the $r - \phi$ bending plane, and one SL is used for the measurement in the z direction along the beamline. The fourth station, MB 4, does not contain z-measuring SLs.

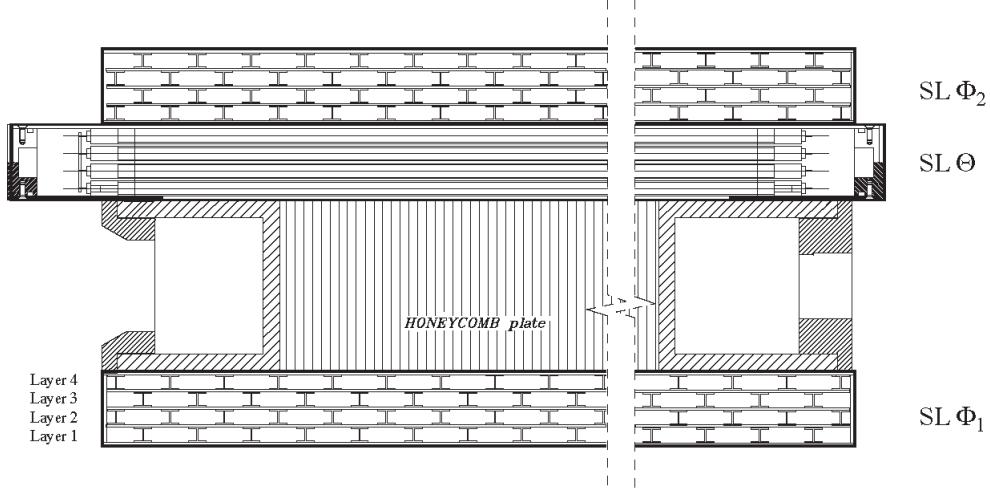


Figure 3.6: The $r - \phi$ view of a Drift Tube (DT) chamber inside a muon barrel station. One can see the 2 SLs, SL Φ_1 and SL Φ_2 , with wires along the z direction (pointing out of the paper) and the other (SL Θ) perpendicular to it. In between is a honeycomb plate with supports attached to the iron yoke.

3.4.2 The Cathode Strip Chambers

There are 468 Cathode Strip Chambers (CSCs) in the two muon endcaps (four stations in each endcap) covering the pseudorapidity range $0.9 < |\eta| < 2.4$. As shown in Figure 3.5, the CSCs are positioned perpendicular to the beamline and interspersed between

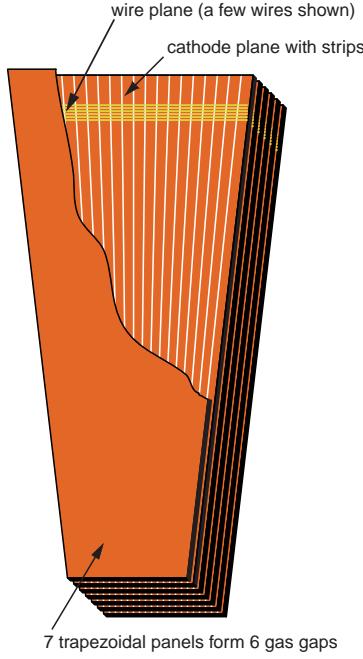


Figure 3.7: Schematic view of a CSC chamber.

iron yoke plates in three rings in ME 1, two rings in ME 2 and ME 3, and one ring in ME 4. All chambers, except for the outermost ring in ME 1, overlap to provide contiguous azimuthal coverage. A [CSC](#) chamber is a multi-wire proportional chamber constituted by six trapezoidal anode wire planes interleaved among seven trapezoidal cathode strip panels covering either 10° or 20° in ϕ , as shown in Figure 3.7. The wires run in the azimuthal direction, and together with the lengthwise (radial) cathode strips, provide precise coordinates in $r - \phi$ of particle passage. The 144 largest CSCs are 3.4 m long along the strip direction and up to 1.5 m wide along the wire direction. The sensitive planes of all chambers cover a total of about 5000 m^2 . The CSCs are designed to provide a $75\text{ }\mu\text{m}$ spatial resolution (per chamber) in $r - \phi$ for the inner two rings of ME 1 and about $150\text{ }\mu\text{m}$ for all others.

3.4.3 The Resistive Plate Chambers

In addition to the **CSC** and **DT** chambers, a complementary system of Resistive Plate Chambers (RPCs) is installed in both the barrel and endcap regions, covering a large portion of the muon system in the range $|\eta| < 1.6$ (see Figure 3.5). An RPC chamber is a gaseous parallel-plate detector that combines adequate spatial resolution with a time resolution comparable to that of scintillators [40] [41]. There are 480 rectangular RPCs (each RPC chamber is composed of two or three double-gap modules operating in avalanche mode [42]) populated in four barrel layers (RB 1 to RB 4) as follows: In the first and second muon stations each **DT** chamber is attached to two RPC chambers via support plates glued to the inner and outer faces. This allows four layers of trigger information even for low- p_T muons that do not reach the outer barrel layers. There are again two RPCs for each **DT** chamber in the third and fourth muon stations (except for some special cases in RB 4 as shown in Figure 3.8), but they both are attached to the inner side of the **DT** chambers.

There are 216 RPCs in each endcap region, populated in the outer two concentric rings located in the inner face of the three iron endcap disks. All trapezoidal shaped RPCs in the endcaps overlap to provide contiguous azimuthal coverage.

3.5 The Trigger System

At the full design luminosity the LHC provides proton-proton collisions with a bunch crossing rate of 40 MHz, approximately 10^9 interactions per second. The CMS experiment uses a two-level online trigger system with an aim to keep events of interest while reducing the data rate to about 100 Hz due to limited storage and computing capacity.

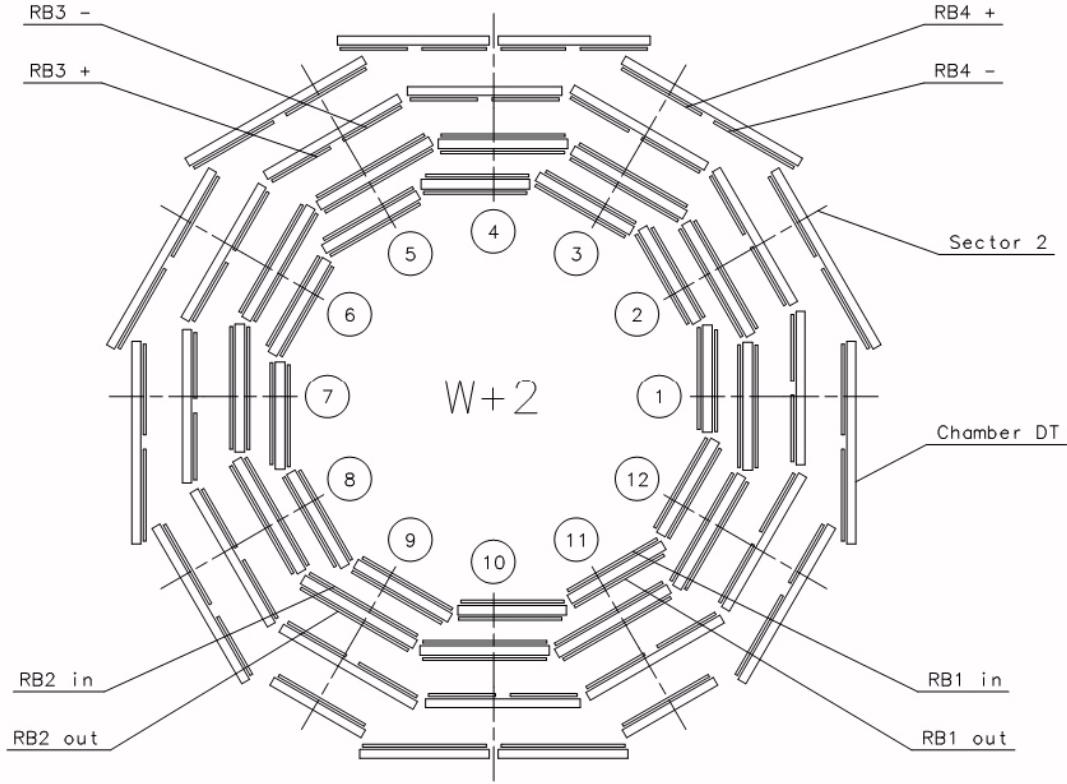


Figure 3.8: Schematic layout of one of the 5 barrel wheels. Each wheel is divided into 12 sectors that are numbered as shown. There are two RPCs for each muon station, except for the special cases for RB 4 in sector 4, which consists of four RPCs, and in sectors 9 and 11, each contains only one RPC.

The first level is hardware-based and is referred to as the “Level-1 (L1) Trigger” [43] while the second level is software-based, referred to as “High-Level Trigger” ([HLT](#)) [44].

The design output rate of the L1 Trigger is limited to 100 kHz, while holding the entire event information in pipelined memories in the front-end electronics. The depth of these pipeline buffers is limited to 128 LHC bunch crossings, leading to a total L1 trigger latency of $3.2 \mu\text{s}$ at the nominal luminosity. The architecture of the L1 trigger, depicted in Figure 3.9, is composed of three subsystems: the Calorimeter Trigger, the Muon Trigger and the [Global Trigger \(GT\)](#). In the first step, the calorimeter and muon triggers each

involve algorithms for the local or regional reconstruction and ranking of calorimeter and muon trigger primitive objects such as electrons, photons, muons and jets above certain E_T or p_T thresholds, using coarsely segmented information from the calorimeters and the muon system. Next, the [Global Calorimeter Trigger \(GCT\)](#) and [Global Muon Trigger \(GMT\)](#) determine the highest-rank calorimeter and muon trigger objects throughout the entire experiment and transfer the four best candidates to the [GT](#), which also receives the total and missing transverse energy measurements and the jet multiplicity above certain programmable thresholds from the [GCT](#). Finally, the [GT](#) checks if the trigger objects fulfill certain programmable trigger algorithms and takes the decision to reject an event or to accept it for further evaluation by the [HLT](#).

Since the [HLT](#) runs on a large farm of commercial multi-core computers and has access to the complete readout data, it can perform more sophisticated calculations similar to those in the offline physics analysis, providing the capability for a full reconstruction of the event topologies and filtering according to several desired physics criteria. The number and type of trigger levels can be optimized according to the running condition of the LHC. For the online data taking, the [HLT](#) performs partial reconstruction of physics objects relying only on the information from subdetectors crossed by the probable passage of candidate objects. The [HLT](#) processes the L1 output and reduces the event rate by a factor of 10^3 to about 100 Hz.

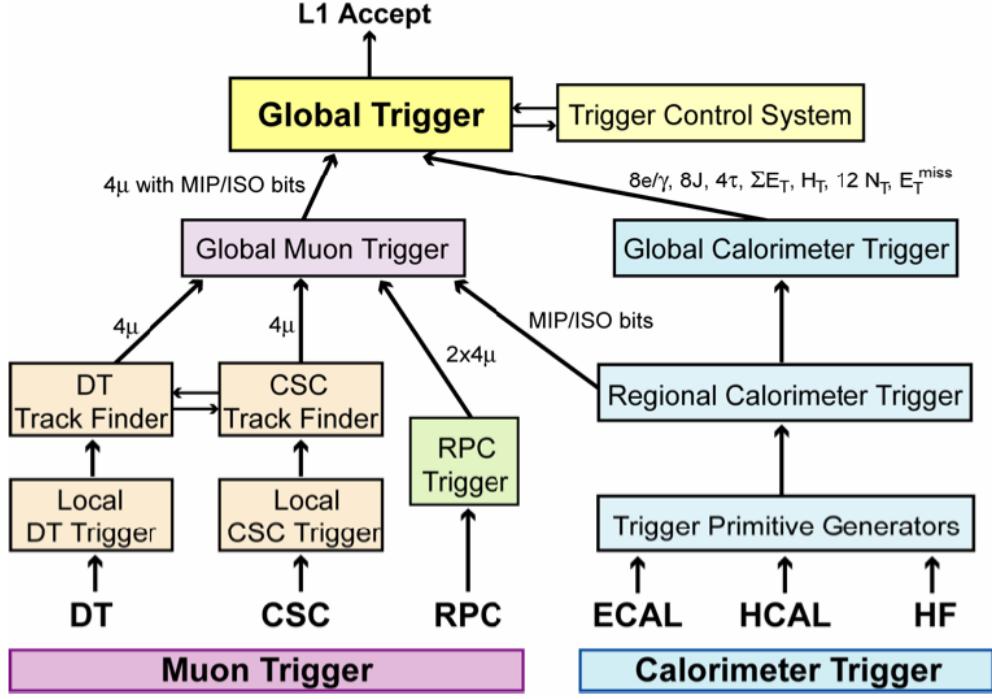


Figure 3.9: Architecture of the Level-1 (L1) Trigger.

3.6 The Data Acquisition System

A schematic of the CMS [Data Acquisition \(DAQ\)](#) system is shown in Figure 3.10.

It is comprised of several interconnected systems: an L1 trigger system; a high-throughput builder network capable of transferring data at 100 GB/s, interfaced to subdetector Front-End Drivers (FEDs), which store data continuously in 40-MHz pipelined buffers, and to the computing resources. The L1 trigger system decides which data in the pipelined buffers should be conveyed through the builder network to a computer farm (HLT Event Filter) that performs simple physics selections, using dedicated (simplified and faster) algorithms similar to the offline physics reconstruction, to achieve the required output rate (a few 10^2 Hz) of data for offline processing and physics analysis.

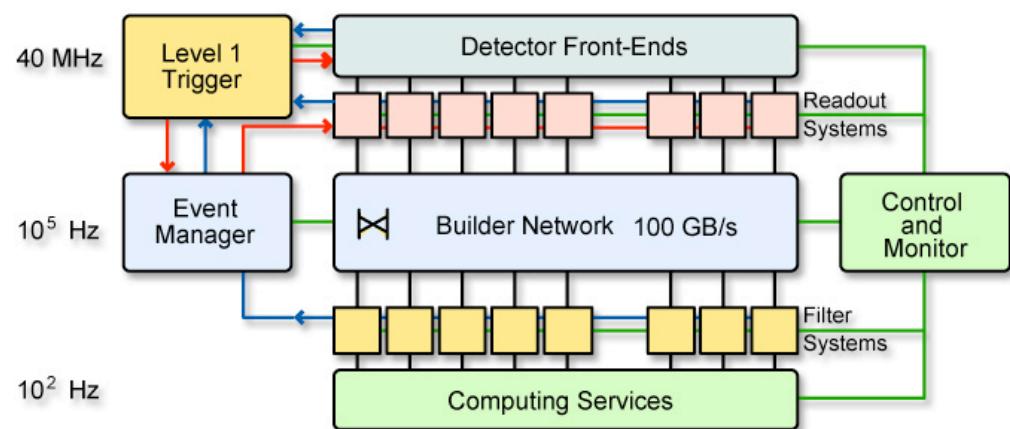


Figure 3.10: General architecture of the CMS Data Acquisition (DAQ) system.

Chapter 4

Data Handling, Event Reconstruction and Monte Carlo Simulation

The datasets and corresponding triggers used in this analysis are described in Section 4.1. Section 4.2 outlines the reconstruction of relevant analysis objects such as muon, jet, and \cancel{E}_T . And finally a detailed account of the signal and background modeling by Monte Carlo simulation is given in Section 4.3.

4.1 Data Sample

The data sample used in this analysis was collected with the CMS detector between June and November 2010. During this time the LHC operated at the center-of-mass energy of 7 TeV and delivered beam luminosity up to $2.07 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$. To take advantage

of the highly parallelized data processing model and make the data access easier for the Physics Analysis Groups ([PAGs](#)), the data sample is, based on the trigger decision, split into primary datasets ([PDs](#)), where events are grouped according to similar use cases, such as experimental signatures, alignments and calibrations, or monitoring. The experimental signature of the $t\bar{t} \rightarrow \mu + \text{jets}$ channel consists of one energetic muon, hard hadronic jets, and large missing transverse energy. Therefore, we are interested in the primary dataset that contains events with at least one muon, i.e., the “Mu” primary dataset.

[Table 4.1](#) shows the list of datasets with corresponding amount of data used in this analysis. The combined integrated luminosity is 36 pb^{-1} . The uncertainty due to the measurement of the total integrated luminosity is assigned to be about 4% [45] [46]. [Table 4.2](#) shows the list of L1 seeds and High Level Trigger ([HLT](#)) paths used to produce the “Mu” primary dataset for the “2E32” menu. Detailed definitions of the L1 and [HLT](#) triggers can be found in [Appendix A](#). [Table 4.3](#) shows the list of runs and the corresponding [HLT](#) paths applied to skim the datasets.

[Table 4.1](#): Data samples used in this study. The combined integrated luminosity is 36 pb^{-1} . The uncertainty due to the measurement of the total integrated luminosity is assigned to be about 4% [45] [46].

Period	Runs	Dataset	$\int \mathcal{L} dt$
2010A	136035 - 144114	/Mu/Run2010A-Nov4ReReco_v1/AOD	3.18 pb^{-1}
2010B	146428 - 149294	/Mu/Run2010B-Nov4ReReco_v1/AOD	32.97 pb^{-1}

4.2 Analysis Object Reconstruction

After the data sample of interest is identified, the next step is to make sure it contains the information needed for the analysis. In this section, the relevant analysis

Table 4.2: L1 and HLT triggers used to produce the “Mu” primary dataset for the “2E32” menu.

L1	HLT
L1_SingleMu7	HLT_Mu9
L1_SingleMu7	HLT_Mu11
L1_SingleMu7	HLT_Mu13_v1
L1_SingleMu7	HLT_Mu15_v1
L1_SingleMu7	HLT_IsoMu9
L1_SingleMu7	HLT_IsoMu11_v1
L1_SingleMu7	HLT_IsoMu9_PFTau15_v1
L1_SingleMu7	HLT_Mu11_PFTau15_v1
L1_DoubleMuOpen	HLT_DoubleMu3_v2
L1_DoubleMuOpen	HLT_DoubleMu5_v1
L1_DoubleMuOpen	HLT_L2DoubleMu20_NoVertex_v1
(L1_SingleMu3) AND (L1_ETM20)	HLT_Mu5_MET45_v1
L1_Mu3_Jet10	HLT_Mu5_Jet50U_v1
L1_Mu3_Jet10	HLT_Mu5_HT70U_v1
L1_HTT50	HLT_Mu3_Ele8_HT70U_v1

Table 4.3: High Level Trigger paths applied to the primary datasets used in this study. Note in the period of 2010B two HLT paths are used due to the significant increase of the instantaneous luminosity at the LHC. Both HLT paths use the same L1 trigger: “L1_SingleMu7.”

Period	Runs	HLT path	$\int \mathcal{L} dt$
2010A	136035 - 144114	HLT_Mu9	3.18 pb^{-1}
2010B	146428 - 147116	HLT_Mu9	5.06 pb^{-1}
	147196 - 149294	HLT_Mu15	27.91 pb^{-1}

objects that are available in the sample are described and how they are formulated or reconstructed from the raw data in which is recorded the low level information of detector responses/electronic signals. The reconstruction starts with the basic objects such as track, vertex and energy cluster, and then based on these, higher level physics analysis objects are built, such as muon, electron, jet, and missing transverse energy. For more technical details, one can review the published literature [47].

4.2.1 Beamspot, Track and Vertex Reconstruction

- **Beamspot**

The proton-proton interaction region, referred to as the beamspot, is used as a precise estimate of the primary interaction point prior to primary vertex reconstruction and as the sole primary interaction point if no primary vertex is found. When the beamspot center is displaced from the expected position there is a correlation between the transverse impact parameter (d_{xy}) and the angle (ϕ_0) of the track at the point of closest approach. The beamspot fitter [48] uses an iterative χ^2 fitter to exploit this correlation between d_{xy} and ϕ_0 , looping over a set of reconstructed tracks (reconstructed with respect to the current beamspot parameters in the database) to determine the new beamspot parameters, i.e., positions and widths. Since the proton-proton interaction region is stable for the beam life time and the precision of the beamspot fit increases with more input tracks, the beamspot is determined by averaging over many events. The input track collection is reset frequently so that the beamspot fit maintains sensitivity for the change of the beam condition. The fitted beamspot can then be updated to the database and used to re-reconstruct tracks.

- **Tracking**

As shown in Figure 4.1 the track reconstruction is composed of several logical parts: local reconstruction, seed finding, pattern recognition, and final track fit. Track reconstruction uses primarily the experimental raw data of electronic signals, referred to as “digis,” left by charged particles traversing the tracking detectors. The digis are taken as input to the local reconstruction, which consists of clustering adjacent pixel or strip digis and “RecHit” conversion, to give the reconstructed “hit” positions with uncertainties. Starting from the location of the interaction region, i.e., the beamspot, an initial round of track and vertex reconstruction is performed using only pixel hits. The pixel vertices found at this stage are used in the standard iterative tracking. The standard iterative track reconstruction at CMS is performed by the combinatorial track finder (CTF) [49], which is based on the Kalman filter method [50]. The starting points of tracks (“seeds”) are formed from either triplets of hits in the tracker or pairs of hits with an additional constraint from the beamspot or a pixel vertex, yielding an initial estimate of the trajectory, including its uncertainty. The seed is then propagated outward in a search for compatible hits (“inside-out”). As compatible hits are found, they are added to the trajectory and the track parameters and uncertainties are updated. This search continues until either it has reached the boundary of the tracker or no more compatible hits can be found. An additional search for hits is performed starting from the outermost hits and propagating inward (“outside-in”). In the case that two tracks share a significant number of hits, the one with fewer hits is discarded to prevent double counting. In the final step, the collection of hits is fit to obtain the best estimate of the track parameters [51].

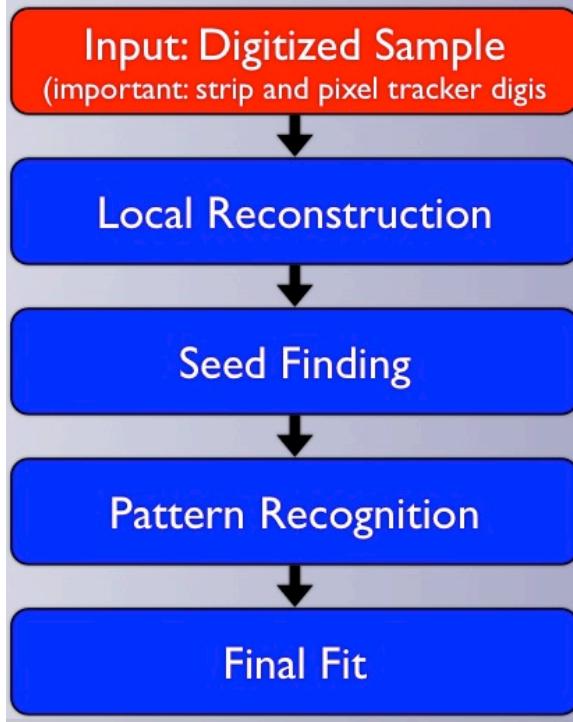


Figure 4.1: The workflow of the tracking algorithm.

- **Primary Vertex Reconstruction**

The reconstruction of the primary interaction vertex in the event starts from the track collection. Tracks are first filtered using cuts on the number of hits and fit quantities including the impact parameter with respect to the beamspot, and the normalized χ^2 . The filtered tracks are then clustered into several primary vertex candidates according to their z-coordinate at the point of closest approach. The tracks in the cluster are then fit with an Adaptive Vertex Fitting [52], where tracks in the vertex are assigned a weight based on their compatibility with the common vertex. The primary vertex candidates compatible with the beamspot are retained in the primary vertex collection [51].

4.2.2 Energy Reconstruction in the Calorimeters

The energy reconstruction in the calorimeter system consists of two steps: First, a local reconstruction identifies the position, time and energy of the energy deposits (clusters) in the Electromagnetic Calorimeter ([ECAL](#)) and Hadronic Calorimeter ([HCAL](#)) cells. Second, a global reconstruction combines the [ECAL](#) and [HCAL](#) clusters into calorimeter towers. A calorimeter tower has a definite position in the (η, ϕ) plane and consists of one or more [HCAL](#) cells and the geometrically corresponding [ECAL](#) crystals. The towers can then be used as the basis for the calorimeter based jet reconstruction algorithm as described in Section [4.2.5](#).

4.2.3 Muon Reconstruction

The muon is the key physics object for this analysis. There are three types of muons in CMSSW, described below [53]:

- **Standalone Muon:**

Standalone muons are built from muon chamber information only. The initial trajectory estimates, i.e., seeds, are formed by matching up pairs of [DT](#) and [CSC](#) segments. The momentum is estimated from the bending of the muon. The Kalman Filter [50] based fitter picks up and fits [CSC](#) and [RPC](#) hits and [DT](#) segments along the trajectory.

- **Global Muon:**

Global muons are a combination (using the “outside-in” approach) of the standalone muons with silicon tracker tracks. The approach starts with each standalone muon and search for matched tracker tracks. It then performs a Kalman Filter fit using track

candidate hits and muon hits/segments. The one with the highest fit probability is chosen as a “Global Muon.”

- **Tracker Muon:**

Tracker muons are built from the “inside-out” approach. It is a complementary approach for low p_T muons that fail standalone muon reconstruction due to insufficient information in the muon system. The method treats each silicon track as a seed and picks up muon signatures along its extrapolated path through subsequent subdetectors, such as compatibility with a muon hypothesis based on energy depositions in calorimeter cells crossed by the track extrapolation. Finally, only those candidates satisfying the following criteria are identified as “Tracker Muon”:

- Track $p_T > 0.5$ GeV/ c and $p > 2.5$ GeV/ c .
- At least one muon segment with a measurement matches to the extrapolation of candidate track.

The three different muon types are merged into one common muon collection, which is used for this analysis.

4.2.4 Electron Reconstruction

When a high-energy electron travels through matter, it initiates an electromagnetic cascade (shower) as bremsstrahlung and pair production generate more electrons and photons with lower energy. Because of the strong magnetic field in the CMS detector an electron shower deposits its energy in several [ECAL](#) crystals spread in ϕ . Therefore, electron reconstruction is based on the association of an [ECAL](#) “supercluster,” which is a collection

of the ECAL clusters extended in ϕ , with an inner tracker track, which is built from seeds (pairs or triplets of hits) in the pixel detector found via a “ECAL driven” algorithm or a complimentary “tracker-driven” seeding algorithm [54] [55], followed by a reconstruction of the trajectory in the silicon strip tracker with a “Gaussian Sum Filter” (GSF) [56] using a dedicated energy loss modeling [57]. The electron energy is deduced from a weighted combination of the corrected supercluster energy and tracker momentum measurements.

4.2.5 Jet Reconstruction

The quarks and gluons produced in the pp collisions cannot be observed singularly due to color confinement in Quantum Chromodynamics; instead, they fragment into final state particles through spontaneous creation of quark-antiquark pairs, which is referred to as hadronization or fragmentation, resulting in bunches of hadrons called jets that are collimated in the direction of the initial partons. Therefore, most jet reconstruction algorithms make use of a clustering technique, in which energy deposits, e.g., calorimeter towers, close in (η, ϕ) to a high E_T tower, the “seed” tower, are summed together to form a jet, subject to some constraints such as the cone size in the (η, ϕ) plane.

There are three major types of jet reconstruction algorithms in CMSSW using the Anti- k_T [58] clustering algorithm with the cone size parameter $R = 0.5$: (1) calo jets, (2) jet-plus-track (JPT) jets, and (3) particle-flow (PF) jets. Calo jets are reconstructed based only on calorimeter information (calorimeter towers) [59]. JPT jets improve the p_T response and resolution of calo jets by taking into account the momenta of the associated tracks of the charged particles [60]. PF jets are reconstructed using the list of particles, namely muons, electrons, photons, charged hadrons, and neutral hadrons, from the particle-flow

method [61], which reconstructs each particle in the event based on information from all CMS subdetectors.

4.2.6 Missing Transverse Energy

Although the CMS detector covers nearly 4π solid angle, it is however not completely hermetic, as it is necessary to have an opening in the very forward direction to allow the beam pipe to go through. The design of the calorimeter system covering up to $|\eta| \approx 5$ has significantly limited the escaping forward particles to carry away large transverse momentum, since

$$p_T \stackrel{\frac{m}{E} \ll 1}{=} \frac{E}{\cosh \eta} < \frac{E}{\cosh(5)} \approx 0.013 \times E. \quad (4.1)$$

For most physics processes of interest the typical momentum of forward particles in hard collisions is significantly smaller than the full beam energy, so the transverse momentum carried away by the particles beyond the detector acceptance is negligible, which allows the detector to examine rather precisely the momentum conservation in the transverse plane perpendicular to the beam direction. As a result, any apparent imbalance of the measured transverse momentum is indicative of the presence of one or more weakly interacting particles (e.g., neutrinos) that escape from the detector without leaving a trace. This imbalance is referred to as missing transverse momentum, denoted by \vec{E}_T , and its magnitude as missing transverse energy, denoted by E_T . The CMS collaboration has implemented four major types of E_T [62]:

1. E_T^{calo} , derived from energy deposits in the reconstructed calorimeter towers [63].

2. \cancel{E}_T^{tc} (or track-corrected \cancel{E}_T), which is the same as \cancel{E}_T^{calo} except that those calorimeter tower energies that are matched to muons and charged hadrons are replaced by the corresponding charged track momenta [64].
3. \cancel{E}_T^{PF} , calculated using a complete particle-flow technique [61].
4. \cancel{H}_T , calculated using reconstructed jets.

The $\overrightarrow{\cancel{E}_T}$ is calculated by

$$\overrightarrow{\cancel{E}_T} = - \sum_n (E_n \sin \theta_n \cos \phi_n \hat{x} + E_n \sin \theta_n \sin \phi_n \hat{y}), \quad (4.2)$$

where the index n runs over all input objects of the algorithm used (e.g., energy deposits in calorimeter towers for \cancel{E}_T^{calo} , reconstructed particles for \cancel{E}_T^{PF}). Note that since we use \cancel{E}_T^{PF} in this study, unless stated otherwise, the \cancel{E}_T stands for \cancel{E}_T^{PF} in the following.

4.3 Monte Carlo Simulation

Monte Carlo methods are a class of well established mathematical techniques that can be applied to simulate the indeterministic behavior of High Energy Physics phenomena. The simulation for the CMS physics studies involves two main steps. First, the “event generation” simulates the physics process of interest using a dedicated Monte Carlo approach to generate randomly the inputs, e.g., the flavor and kinematic information about the initial partons involved in the hard scattering, using a random number generator. It then calculates the probability of certain physics outcome, e.g., a $t\bar{t}$ event, using a specific theory model over the domain of possible inputs and provides kinematic information for the final state

particles. The generated events are weighted according to various probability distribution functions involved in the method. Second, the “detector simulation,” using similar Monte Carlo methods, models how these final state particles travel through the detector and leave distinguishable signatures. These are described in more detail below.

4.3.1 Event Generation

High energy physics event generation can typically be split into the following steps: [Matrix Element \(ME\)](#) calculation, [Parton Shower \(PS\)](#), Underlying Event modeling and Hadronization. The matrix element calculation involves the leading order or tree level Feynman diagrams of parton-parton hard scattering; the momenta, spins and color connections of the final state partons (quarks, gluons, leptons and photons) are assigned based on the expected distributions from the theory. Since the experimental observables are mesons, baryons, leptons and photons, the connection between final state partons and experimental observables is normally done by making use of the parton shower formulation. Following the hard scattering, the final state partons are evolved through repeated parton branching (e.g., $q \rightarrow qg$, $g \rightarrow gg$, $g \rightarrow q\bar{q}$, $q \rightarrow q\gamma$, and $\ell \rightarrow \ell\gamma$) according to the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi ([DGLAP](#)) perturbative evolution [65] (taking PYTHIA as an example). The initial and final state radiation corrections (due to soft and collinear emissions of gluons and quarks) are included by the resummation of leading logarithmic contributions. To prevent double counting due to the showering, a matrix element-parton shower matching prescription is carried out before creating the parton configuration of the generated event. The underlying events, which consist of “beam remnants” and particles arising from soft or semi-soft multiple parton interactions, are modeled by the generator as well.

The process by which partons evolve to produce hadrons is referred to as hadronization or fragmentation, which cannot be calculated using perturbative QCD. Instead, several hadronization models are available as alternatives to describe such process. Therefore, once the partonic configuration is available, it is fed to a specific hadronization model (depending on the generator used) to produce final state hadrons. Once generated, unstable hadrons or leptons are allowed to decay into the final state particles according to their known branching ratios. There are many software packages available in the high energy physics community for the Monte Carlo simulation. Those relevant to this analysis are briefly described below:

- **PYTHIA** [66] is a multi-purpose Leading Order (LO) generator that can calculate hard scattering process using the tree level Feynman diagrams. It is fully equipped with functionalities such as parton shower, underlying event modeling, and hadronization. In this analysis PYTHIA version 6.4 [66] is used and is referred to as PYTHIA6. The main hadronization option in PYTHIA is the “Lund” string scheme [67] [68], which involves stretching a color “string” across quarks and gluons, and breaking it up into final state hadrons. PYTHIA6 can also interface with other Matrix Element generators, such as ALPGEN [69] and MADGRAPH [70].
- **MADGRAPH** [70] is a Matrix Element based generator, which takes into account leading order Feynman diagrams with additional hard (wide angle) emissions before going through the parton shower step (produces soft/collinear emissions). Given a physics process of interest, MADGRAPH automatically generates the amplitudes for all the relevant subprocesses and produces the mappings for the integration over the phase space. This process-dependent information is passed to built-in MADEVENT to pro-

duce a stand-alone code, which allows the user to calculate cross sections and to obtain unweighted events in a standard output format, i.e., the [Les Houches Events \(LHE\)](#) file format [71], which provides a common convention for the Monte Carlo event generators to define Matrix Element level event listing. Since MADGRAPH and MADEVENT are normally deployed in sequence, they are collectively referred to as MADGRAPH in this dissertation. The output of MADGRAPH is consequently interfaced to a parton shower Monte Carlo program, such as PYTHIA, to simulate showering, underlying events, hadronization, and so on.

4.3.2 Detector Simulation

The CMS detector simulation is based on GEANT4 [72] [73] and is fully integrated within the CMSSW framework. It describes the detector geometry and materials, uses information about the magnetic field, and simulates the detector response, the effects of energy loss, multiple scattering and showering in the detector materials when the generated particles traverse the CMS detector.

4.3.3 Signal Modeling

The simulation of signal $t\bar{t}$ events was done using MADGRAPH [70] version 4.4.12 with CTEQ6L1 [74] parton distribution functions and assuming a top quark mass of 172.5 GeV/c^2 . A detailed description of all the parameters used can be found in the Monte Carlo Database ([MCDB](#)) [75] article 846. The relevant subprocesses of top-antitop pairs accompanied by up to three additional jets are included in the matrix element calculation. The W bosons from the top quark decays are allowed to decay inclusively. For those

processes containing tau (τ) leptons, the decay of the τ is taken care of by a dedicated Monte Carlo program called TAUOLA [76]. Because of their substantial life time, τ decays can be well separated from their production process. The TAUOLA interface supplements events generated by MADGRAPH with a more precise description of τ decays. The resulting parton configurations generated by MADGRAPH (+TAUOLA) are then processed by the parton shower Monte Carlo program PYTHIA; in this step the MLM prescription [77] is adopted for matrix element-parton shower matching to remove double counting of events.

4.3.4 Background Modeling

The main backgrounds to top quark pair production in the μ +jets channel are modeled using the MADGRAPH and PYTHIA generators. The W +jets, Drell-Yan ($Z/\gamma^* \rightarrow \ell\ell$) and single top backgrounds are generated using MADGRAPH, while the diboson (W^+W^-) and QCD multijet events are generated using PYTHIA. The event signatures of the separate processes are described below:

- **W +jets** process, in which the W boson decays leptonically, exhibits very similar signatures as the signal events, especially when with higher jet multiplicity. Although the differential cross section decreases in terms of jet multiplicity, since the total cross section of the W +jets process is two orders of magnitude larger than the signal process, there are still significant amounts of W +jets events that will pass simple cut-based event selection and mimic the signal. It is in fact the largest background to this analysis. A detailed description of the sample is available in the MCDB article 886.

- **Single Top** contributions are categorized according to the production mechanism as s-channel, t-channel and tW-channel. The s-channel cross section is small and therefore ignored. Since the single top process contains a real top quark, there is a substantial amount of single top contributions to the backgrounds. Detailed descriptions of the tW-channel and t-channel are available in the [MCDB](#) article 689 and 690, respectively.
- **Drell-Yan Process** ($Z/\gamma^* \rightarrow \ell\ell$) [78] can imitate signal events when one of the leptons escapes the detector acceptance, resulting in missing energy, or is misidentified as jet. Due to the large cross section, the Drell-Yan process is the second largest backgrounds for this analysis. A detailed description of the sample is available in the [MCDB](#) article 851.
- **Diboson Processes** include W^+W^- , ZZ and $W^\pm Z$. The contributions from ZZ and $W^\pm Z$ are negligible due to their small cross sections. W^+W^- with extra jets can mimic the signal events very well due to the existence of two W bosons, of which one decays leptonically and one hadronically. It is therefore included in this study.
- **QCD** acts like a signal event when one of its jets is misidentified as a lepton. The predicted [QCD](#) production cross section is so large that generating such processes with [Monte Carlo](#) is neither practical nor effective. Therefore, the [QCD](#) multijet events were generated with generator-level filters, depending on the processes wanted. For this analysis the QCD samples are generated with $\hat{p}_T > 20$ GeV/ c and at least one muon of $p_T > 15$ GeV/ c .

4.3.5 Monte Carlo Samples

This analysis was performed using CMSSW (see Appendix B) version 3.87. The Monte Carlo simulated samples used in this analysis are listed in Table 4.4. The general purpose Monte Carlo program MCFM [79] [80] is used to calculate NLO cross sections for the $t\bar{t}$ process ($\sigma_{t\bar{t}} = 157.5^{+23.2}_{-24.4} \text{ pb}$ [81]), the t-channel single top process ($\sigma_t = 20.93^{+1.10}_{-1.04} \text{ pb}$ [82, 83]), the tW-channel single top process ($\sigma_{tW} = 10.6 \pm 0.8 \text{ pb}$ [84]), and the diboson process (W^+W^- , where $\sigma_{W^+W^-} = 43 \pm 1.5 \text{ pb}$). The NNLO cross section for $W+\text{jets}$, where the W boson decays into leptons, has been determined to be $31.3 \pm 1.6 \text{ nb}$ using FEWZ [85]. The NNLO cross section of the Drell-Yan dilepton process ($m_{\ell\ell} > 50 \text{ GeV}/c^2$) is also calculated with FEWZ to be $3.05 \pm 0.13 \text{ nb}$. The LO cross section value for the filtered QCD sample is calculated with PYTHIA. A complete analytic result for the NLO QCD cross section has been published only recently [86]. There also exist higher order calculations for the $t\bar{t}$ and single top processes [87–89]. The estimated cross sections are comparable to the NLO values. For the samples that are produced inclusively, the quoted values correspond to the inclusive cross sections. For those only the leptonic decay is simulated, so the effective cross sections also include the corresponding branching ratios so that $\sigma_{\text{eff}} = \sigma \cdot \text{BR}$. The QCD sample is generated using filters on the generator level in order to increase the number of simulated events that are likely to pass the event selection cuts. The LO cross section value quoted for the QCD sample is thus the effective cross section including the filter efficiencies (ε_f), $\sigma_{\text{eff}} = \sigma \cdot \varepsilon_f$.

Table 4.4: Monte Carlo dataset names from DBS. This analysis was carried out on the AODSIM Monte Carlo samples; the “/AODSIM” is implied and therefore has been stripped from the dataset names.

Process	Dataset
$t\bar{t}$	TTJets_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v3
single top (t)	TToBLNu_TuneZ2_t-channel_7TeV-madgraph/Fall10-START38_V12-v2
single top (tW)	TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/Fall10-START38_V12-v2
W +jets	WJetsToLNu_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v1
$Z/\gamma^* \rightarrow \ell\ell$	DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola/Fall10-START38_V12-v2
W^+W^-	WWtoAnything_TuneZ2_7TeV-pythia6-tauola/Fall10-START38_V12-v1
QCD	QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1

Table 4.5: The cross sections for the relevant Monte Carlo processes. $\ell = e, \mu, \tau$.

Process	σ_{eff} [pb]	# of events generated
$t\bar{t}$, inclusive	$157.5^{+23.2}_{-24.4}$	1,286,491
single top (t), $t \rightarrow b\ell\nu$	$20.93^{+1.10}_{-1.04}$	484,060
single top (tW), inclusive	10.6 ± 0.8	494,961
W +jets, $W \rightarrow \ell\nu$	$31,314 \pm 1,558$	14,805,546
$Z/\gamma^* +\text{jets}$, $Z/\gamma^* \rightarrow \ell\ell$ ($m_{\ell\ell} > 50 \text{ GeV}/c^2$)	$3,048 \pm 132$	2,543,727
W^+W^- , inclusive	43 ± 1.5	2,061,760
QCD, filtered	84,679.3	29,504,866

Chapter 5

Event Selection and Data Analysis

A detailed description of the event selection is given in Section 5.1. The measurement of the muon selection efficiencies by using the “tag-and-probe” method [53] is described in Section 5.2, followed by the estimation of the QCD multijet background by using the Matrix Method (MM) [90] in Section 5.3. A direct comparison between the preselected data and the Monte Carlo predictions is presented in Section 5.4.

5.1 Event Selection

The final state of $t\bar{t}$ events in the $\mu+\text{jets}$ channel exhibits one prompt muon, four hadronic jets (two b jets from the top quark decays and two jets from the hadronic W boson decay) and large missing transverse energy. Therefore, the data sample used in this analysis is collected by triggering on collision events that contain at least one muon with $|\eta| < 2.5$ and p_T ranging from 9 GeV to 15 GeV. The muon candidates are also required to be compatible with the luminous region of the beams (i.e., beamspots [48]) to reject

contamination from cosmic muons. In addition to the trigger requirement, the events must have at least one reconstructible primary proton-proton interaction vertex (PV) compatible with the nominal interaction point of the beams. Due to the presence of pile-up events, there might exist more than one PV candidate. In this case, the one with larger sum of squared p_T of associated tracks is chosen as the reference vertex of the event, because in general the hard scattering physics processes of interest exhibit larger transverse energy than the pile-up events produced by low-momentum-transfer (low- Q^2) forward scattering.

Because of high throughput of information ($\gtrsim 100$ GB/s) and finite computing resource in the online data acquisition workflow, it is impossible and unnecessary to fully reconstruct every event for the High-Level Trigger (HLT) during data taking. Instead, only those events that fire Level-1 hardware triggers will be processed by the HLT computer farm, and only the information from subdetectors crossed by the probable passage of candidate objects will be extracted (“unpacked”) for online reconstruction, which is subsequently exploited for trigger decisions. To cover larger acceptance without overloading the computer farm and to prevent rejecting valuable events, muon selection criteria at the HLT are much looser than the requirements for physics analysis. It is therefore essential to include more information and tighter cuts in the offline selection to improve the purity of the selected muons in the events.

The main background for this analysis comes from the $W + \text{jets}$, single top, Drell-Yan (Z/γ^* + jets), diboson (VV , where $V = W$ or Z/γ^*) and QCD multijet processes. Since the QCD multijet cross section is many orders of magnitude larger than the $t\bar{t}$ signal cross section (Table 4.5), we need to find a good way to eliminate such contamination without reducing our signal yields. The muon from the W boson in a top quark decay

is expected to be isolated (well separated in $\eta - \phi$ space) from other high p_T particles in the event, whereas the muon from leptonic decay of a heavy flavor jet in QCD multijet events is surrounded by other particles inside the jet and has a larger impact parameter with respect to the interaction point. Also, the muon loses very little energy when crossing detector materials. We can therefore take advantage of these intrinsic characteristics to reduce the QCD multijet background. The combined relative isolation (denoted by I_{rel}^μ) has been widely used to select isolated leptons; it is defined as

$$I_{rel}^\mu \equiv \frac{I_{trk} + I_{ecal} + I_{hcal}}{p_T^\mu}, \quad (5.1)$$

where p_T^μ is the transverse momentum of the selected muon associated to the ℓ -th track in the event. The isolation variables, I_{trk} , I_{ecal} and I_{hcal} , are defined for the track p_T in the silicon tracker and for energy depositions in the **ECAL** and in the **HCAL**, respectively, within the solid angle cone of 0.3 in (η, ϕ) space around the muon:

$$I_{trk} \equiv \sum_{\substack{i \neq \ell \\ \Delta R < 0.3}} p_{T,i}^{track}, \quad (5.2)$$

$$I_{ecal} \equiv \sum_i^{\Delta R < 0.3} E_{T,i}^{ecal}, I_{hcal} \equiv \sum_i^{\Delta R < 0.3} E_{T,i}^{hcal}, \quad (5.3)$$

where $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, in which $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and in azimuthal angle between the selected muon and the i^{th} track or calorimeter cell, respectively. In Equation 5.3 the energy depositions in the “veto cone” ($\Delta R = 0.07$) around the extrapolated path of the muon in the calorimeter are excluded [53]. The calculation of the combined relative isolation for the electron, I_{rel}^e , is similar.

The I_{rel}^μ is expected to be small for isolated muons. To calculate this variable more precisely, the selected muon is required to have an associated high quality tracker track. A high quality reconstructed track usually passes through many layers of tracker detector modules, resulting in many measurements (hits) compatible with the extrapolated path and therefore good χ^2 -fit probability. For this analysis, we select events that contain only one good isolated prompt muon, which is found by both the Global and Tracker Muon reconstruction algorithms (see Section 4.2.3), associated to a good tracker track and satisfying $I_{rel}^\mu < 0.1$. “Prompt” means that its extrapolated track is within 0.02 cm of the luminous region in the transverse plane perpendicular to the beams and within 1 cm along the direction of the beams. Furthermore, we reject events that have an extra muon satisfying $p_T > 10$ GeV/ c , $|\eta| < 2.5$ and $I_{rel}^\mu < 0.2$ to suppress the contribution coming from the Drell-Yan Z/γ^* +jets background. Since we are interested only in the decay into muon and jets, we reject events containing at least one electron with $E_T > 15$ GeV, $|\eta| < 2.5$ and $I_{rel}^e < 0.2$.

The neutrino from the leptonic W boson decay is manifested as missing transverse energy (\cancel{E}_T). Since the W boson in top quark decay is boosted, \cancel{E}_T in $t\bar{t}$ events tends to be larger than \cancel{E}_T in electroweak W +jets events. On the other hand, QCD multijet events usually exhibit no substantial missing transverse energy and thus can be further reduced by applying a \cancel{E}_T cut. Table 5.1 shows the comparison of the selections for the QCD Monte Carlo sample in the region $\cancel{E}_T > 20$ GeV (“signal” region) and in the region $\cancel{E}_T < 10$ GeV (“background” or “QCD” region). In the leading order perturbative QCD (pQCD), jet production in pp collisions occurs primarily when two partons interact via the strong force to produce two final state partons ($2 \rightarrow 2$ process), which undergo parton showering and

hadronization to form jets. Since QCD multijet events with three or more jets in the final state originate from hard gluon radiation and higher order QCD processes, the differential cross section decreases significantly as a function of jet multiplicity. Therefore, only the results for the one- and two-jet bins are presented (the event yields in the three- and four-jet bins are too small to reach any meaningful conclusion).

Table 5.1: The event yields for the QCD Monte Carlo sample in the region $\cancel{E}_T > 20$ GeV (“signal” region) and in the region $\cancel{E}_T < 10$ GeV (“background” or “QCD” region).

N_{jets}	$\cancel{E}_T < 10$ GeV	$\cancel{E}_T > 20$ GeV
= 1	4385	648
= 2	408	186

W +jets events show less hadronic jet activity than the $t\bar{t}$ signal, which exhibits at least four jets in the final state including two bottom jets and two from hadronic W boson decay. Therefore, events exhibiting more well reconstructed high- p_T jets are more likely to be from the $t\bar{t}$ signal. A good jet should have at least some fraction of total jet energy deposited in the [ECAL](#) and not all of the energy appearing in a single [HCAL Hybrid Photodiode \(HPD\)](#) readout. In this analysis, we used the particle-flow (PF) jets and the PF missing transverse energy [61]; however, since the PF muon collection is very similar to the standard RECO muon collection, we therefore used the standard muon collection. The complete list of selection cuts used in this analysis is summarized below:

1. The event is required to pass the single muon trigger requirements (see Section 4.1).
2. The event should contain a good primary vertex satisfying (ref. Section 4.2.1):
 - Primary Vertex (PV) fit converged (reconstructible and not fake).
 - The number of degrees of freedom of the PV fit (n_{dof}) > 4 .

- $|z| < 24$ cm; $\rho < 2.0$ cm, where $\rho = \sqrt{x^2 + y^2}$.

3. The event contains exactly one isolated muon satisfying (ref. Section 4.2.3):

- The muon is a Global Muon and a Tracker Muon.
- $p_T^\mu > 20$ GeV/c, $|\eta| < 2.1$ and the combined relative isolation (I_{rel}^μ) < 0.1 .
- The global track associated with the muon satisfies: χ^2/n_{dof} of the global muon fit < 10 and the number of valid muon-detector hits used in the global fit > 0 .
- The number of valid hits of the inner track associated with the muon > 10 .
- The inner track of the muon must contain at least one pixel hit.
- The 2D impact parameter calculated with respect to the beamspot satisfies: $|d_0| < 0.02$ cm.
- The longitudinal distance between the vertex of the inner track of the muon (z^μ) and the primary vertex (z^{PV}) satisfies: $|z^\mu - z^{PV}| < 1$ cm.
- The muon must be isolated from hadronic activity, such that $\Delta R(\mu, \text{jet}) > 0.3$.
- The segments matched to a Global Muon must be part of at least two muon stations.

4. The event cannot contain an additional loose muon, defined as:

- The muon is a Global Muon.
- $p_T^\mu > 10$ GeV/c, $|\eta^\mu| < 2.5$ and $I_{rel}^\mu < 0.2$.

5. The event cannot contain an electron, defined as:

- $E_T^e > 15$ GeV, $|\eta^e| < 2.5$ and $I_{rel}^e < 0.2$.

6. The event is required to have missing transverse energy $\cancel{E}_T > 20$ GeV.

7. The selected jets must satisfy the following requirements:

- $p_T^{\text{jet}} > 25$ GeV/c and $|\eta^{\text{jet}}| < 2.4$.
- The jets must satisfy the jet ID conditions:
 - The electromagnetic fraction > 0.01 .
 - The number of RecHits containing 90% of the jet energy > 1 .
 - The fraction of energy in the hottest HPD readout < 0.98 .
- Jets must be isolated from the muon with $\Delta R(\mu, \text{jet}) < 0.1$.

In this analysis, we categorize the preselected (passing cut 1 to 6) data and Monte Carlo samples according to the number of selected jets (passing cut 7), e.g., the sample with exactly one jet is referred to as the [1-jet sample](#), the sample with exactly two jets as the [2-jet sample](#), the sample with at least two jets as the [2plus-jet sample](#), and so on. Although $t\bar{t}$ events in the $\mu+\text{jets}$ channel should contain at least four jets in the final state, due to the limitations on the detector acceptance and resolution, one can expect non-negligible contributions of $t\bar{t}$ events with only three reconstructible jets. Also, if only events with more than three jets are selected, the statistics of the orthogonal data sample (as for the data driven QCD model; described in Section [5.4.1](#)) is low; this prevents us from obtaining a reasonable shape for the [QCD](#) background template and results in large uncertainties in the [QCD](#) background modeling. We therefore include in this analysis the samples with exactly three jets, referred to as the the exclusive three jet or [3-jet samples](#), and those with at least four jets, referred to as the [4plus-jet samples](#) or the inclusive four jet samples. Table [5.2](#) shows the cut flow for the [Monte Carlo](#) samples and Table [5.3](#) for the data.

Cut	$t\bar{t}$	$W \rightarrow \mu\nu_\mu$	$Z \rightarrow \mu^+\mu^-$	QCD	single top		W^+W^-	Total
					t channel	tW channel		
Processed	5670.0 \pm 5.0	1127304.0 \pm 293.0	109728.0 \pm 68.8	3048454.8 \pm 561.2	753.5 \pm 1.1	381.6 \pm 0.5	54.0 \pm 0.1	1548.0 \pm 1.1
HLT	1784.8 \pm 2.8	253269.6 \pm 138.9	34376.0 \pm 38.5	2586537.3 \pm 517.0	261.7 \pm 0.6	104.3 \pm 0.3	19.7 \pm 0.0	272.2 \pm 0.5
Good PV	1784.4 \pm 2.8	252315.1 \pm 138.6	34301.6 \pm 38.5	2585475.4 \pm 516.8	261.5 \pm 0.6	104.3 \pm 0.3	19.7 \pm 0.0	271.7 \pm 0.5
One Iso mu	765.1 \pm 1.8	174836.8 \pm 115.4	14210.1 \pm 24.8	19928.9 \pm 45.4	157.3 \pm 0.5	54.0 \pm 0.2	10.7 \pm 0.0	184.3 \pm 0.4
Loose mu veto	741.7 \pm 1.8	174829.4 \pm 115.4	9283.8 \pm 20.0	19906.5 \pm 45.4	156.9 \pm 0.5	52.4 \pm 0.2	10.7 \pm 0.0	179.4 \pm 0.4
Electron veto	643.0 \pm 1.7	174440.2 \pm 115.2	9063.8 \pm 19.8	19735.6 \pm 45.2	153.8 \pm 0.5	44.9 \pm 0.2	10.5 \pm 0.0	158.8 \pm 0.3
$\cancel{E}_T > 20$ GeV	5784.4 \pm 1.6	157490.0 \pm 109.5	5968.4 \pm 16.0	1121.5 \pm 10.8	136.8 \pm 0.5	39.4 \pm 0.2	9.3 \pm 0.0	137.1 \pm 0.3
1 jet	32.6 \pm 0.4	19209.8 \pm 38.2	665.6 \pm 5.4	647.8 \pm 8.2	51.8 \pm 0.3	6.3 \pm 0.1	2.9 \pm 0.0	51.7 \pm 0.2
2 jets	116.2 \pm 0.7	3900.6 \pm 17.2	148.5 \pm 2.5	185.9 \pm 4.4	47.8 \pm 0.3	13.1 \pm 0.1	4.2 \pm 0.0	36.9 \pm 0.2
3 jets	184.6 \pm 0.9	680.8 \pm 7.2	31.8 \pm 1.2	41.5 \pm 2.1	18.6 \pm 0.2	12.2 \pm 0.1	1.3 \pm 0.0	9.9 \pm 0.1
≥ 4 jets	241.9 \pm 1.0	180.2 \pm 3.7	9.8 \pm 0.7	8.3 \pm 0.9	7.2 \pm 0.1	6.8 \pm 0.1	0.3 \pm 0.0	2.4 \pm 0.0

Table 5.2: Monte Carlo cutflow normalized using theoretical cross sections to 36 pb $^{-1}$. Particle-flow jets and \cancel{E}_T are used.

Table 5.3: Data cutflow for 36 pb^{-1} .

Cut	Data
CleanFilters	49034698
HLT	7727624
Good PV	7718212
One Iso mu	214368
Loose mu veto	208313
Electron veto	207536
$\cancel{E}_T > 20 \text{ GeV}$	157654
1 jet	20012
2 jets	4506
3 jets	1111
≥ 4 jets	459

5.2 Muon Efficiencies

In this section we investigate the differences of measuring muon selection efficiencies in data and in [Monte Carlo](#) simulation. We measured the efficiencies by a data driven approach known as the “tag-and-probe” method [53]. The tag-and-probe method utilizes known mass resonances (e.g., J/ψ , Υ and Z boson) to select particles of the desired type, muons in our case, and probes (measures) the efficiency of a particular set of selection criteria on those particles. Since the analysis of this dissertation relies on the selection of good muons, the accurate and reliable estimation of the muon selection efficiency is essential. Its effect on the cross section calculation can be counteracted by a scale factor, which is the ratio between the measured efficiencies in data and in [Monte Carlo](#) simulation.

We factorize the muon efficiency into a combined muon identification and isolation (ID&ISO) efficiency and a trigger efficiency. The former is the probability of a reconstructed muon to pass the muon quality selections including the identification and isolation cuts (ref. Section 5.1); the latter is the probability of a muon that has already passed

ID&ISO cuts to fire a certain [HLT](#) trigger, which is either HLT_Mu9 or HLT_Mu15 in our case, depending on the period of data taking.

An unbiased $Z \rightarrow \mu^+ \mu^-$ control sample can be collected efficiently using the single-muon [HLT](#) trigger. Events are selected with stricter selection requirements on one muon (the ‘tag’ muon), and with more relaxed selection criteria on the other muon (the ‘probe’ muon). In our study, we started with a muon that has been reconstructed as both Global and Tracker muon (see [4.2.3](#)), and look for another Global muon such that the two muon candidates are required to be oppositely charged, and their invariant mass is within 15 GeV/c^2 of the Z boson mass, which ensures a high purity $Z \rightarrow \mu^+ \mu^-$ sample. The Z mass peak is fitted with a Gaussian distribution and the background is fitted using an exponential function and subtracted. The fit is performed before and after cuts, and the ratio of the numbers of survival events from these two fits gives us the efficiency. The same procedures have been repeated in collision data as well as the Z +jets [Monte Carlo](#) sample. Figures [5.1](#) and [5.2](#) show the ID&ISO efficiency as a function of several kinematic variables. Data/MC scale factors can be extracted from these plots. Since the scale factors are found to be almost constant as a function of the kinematic variables, a single correction factor is used in this analysis. The factors with statistical errors are summarized in Table [5.4](#).

We have also estimated the efficiencies and scale factor in a sample with at least two jets. Figure [5.3](#) shows the selection efficiency and data/MC scale factor as a function of $\Delta R(\mu, \text{jet})$. The statistics are reduced in this case but on average we observe that the scale factor is constant as a function of $\Delta R(\mu, \text{jet})$ within the statistical uncertainties. The average muon isolation efficiencies for data and [Monte Carlo](#) simulation and the scale factor for events with at least two jets are shown in the last column of Table [5.4](#).

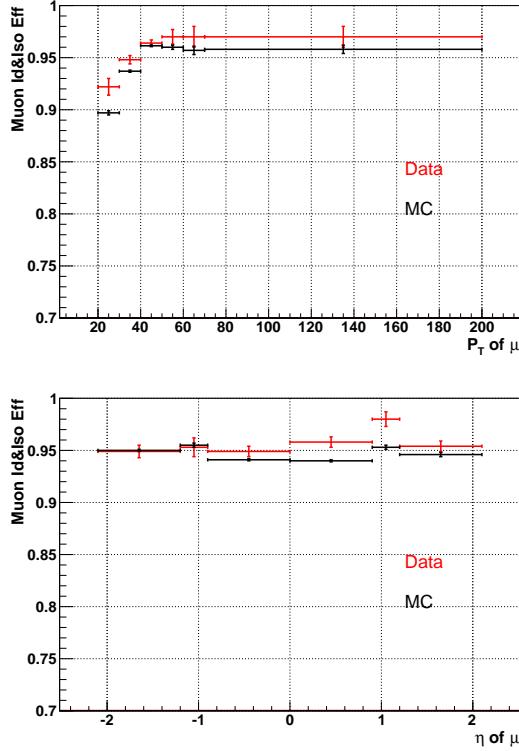


Figure 5.1: Muon ID&ISO efficiencies as functions of kinematic variables: p_T (upper) and η (bottom) of the muon.

Table 5.4: Muon efficiencies in data and MC, and the scale factors (statistical errors only).

Selection	Trigger	ID&ISO	ISO and ≥ 2 jets
Data	0.918 ± 0.003	0.957 ± 0.002	0.97 ± 0.01
MC	0.931 ± 0.001	0.944 ± 0.001	0.957 ± 0.002
Scale Factor	0.987 ± 0.002	1.013 ± 0.001	1.01 ± 0.01

5.3 QCD Background Estimation

After applying the full event selection described in Section 5.1, the preselected data sample is enriched with events containing an isolated high p_T muon, referred to as a “real” muon, originating from the W boson decay. Since there are substantial contributions from the QCD multijet events that contain a muon originating from the semileptonic decay of a heavy quark, referred to as a “fake” muon as opposed to a genuine W boson decay, and

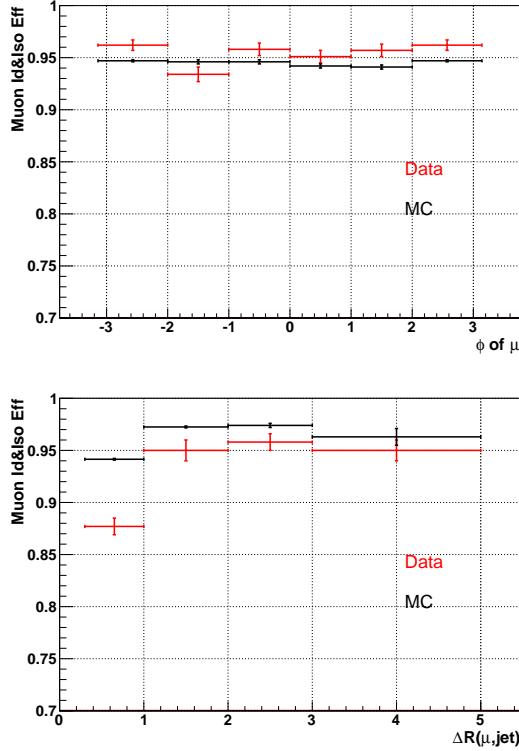


Figure 5.2: Muon ID&ISO efficiencies as functions of kinematic variables: ϕ (upper) and $\Delta R(\mu, \text{jet})$ (bottom) to the closest jet.

the QCD multijet events are not well modeled by the Monte Carlo simulation, the QCD multijet background is better estimated with a data driven method, the Matrix Method (MM) [90], that does not rely heavily on the Monte Carlo simulation.

5.3.1 The Matrix Method

The Matrix Method relies on two data samples: the ‘tight’ sample that corresponds to the preselected sample containing N_T events, and the ‘loose’ sample that consists of N_L events that pass the preselection cuts except for the muon relative isolation requirement. The number of W -like events with leptons originating from the W boson decays is denoted by N^{sig} , and the number of events originating from the QCD multijet production is denoted

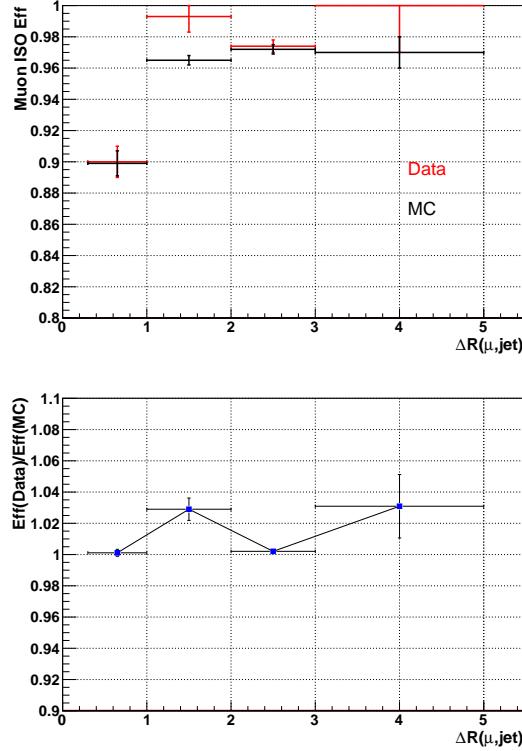


Figure 5.3: Muon ISO efficiencies (upper) and scale factor (bottom) as a function of $\Delta R(\mu, \text{jet})$ to the closest jet in a sample with at least two jets.

by N^{QCD} . N_L and N_T can be written as:

$$\begin{aligned} N_L &= N^{sig} + N^{QCD}, \\ N_T &= \varepsilon_{sig} N^{sig} + \varepsilon_{QCD} N^{QCD}, \end{aligned} \quad (5.4)$$

where ε_{sig} is the efficiency for a loose muon from a W boson decay to pass the tight criteria; it is measured in a W -like sample obtained by combining the Monte Carlo samples of the $W+\text{jets}$, $Z/\gamma^* \rightarrow \ell\ell$, $t\bar{t}$ and single top processes. ε_{sig} measured in this W -like sample is corrected by a data-to-simulation scale factor derived from $Z \rightarrow \mu^+ \mu^-$ events (see Section 5.2). A conservative uncertainty of 2% (compared to the less than 1% uncertainty

obtained from the tag-and-probe method) is assigned to ε_{sig} . ε_{QCD} is the (fake) rate at which a loose muon in the QCD multijet events also passes the tight criteria.

The fact that the signal events can leak into the QCD dominated region ($I_{rel}^\mu > 0.1$) may introduce a bias to the method. In order to minimize the signal contamination, the region corresponding to a low missing transverse energy ($\cancel{E}_T < 10$ GeV) was chosen to measure the ε_{QCD} . The “QCD region” with $\cancel{E}_T < 10$ GeV where we measure ε_{QCD} and the “signal region” with $\cancel{E}_T > 20$ GeV where we estimate the QCD multijet background in the final sample are visualized in Figure 5.4; the area for $I_{rel}^\mu < 0.1$ in the signal region indicates the tight sample as a subset of the loose sample. From Table 5.5 it can be observed that the number of the Monte Carlo $t\bar{t}$ events measured in the region $\cancel{E}_T < 10$ GeV is smaller than the number of the Monte Carlo QCD events in this region for all jet multiplicity bins. The Monte Carlo QCD event yields are apparently underestimated for the [3-jet sample](#) and the [4plus-jet sample](#), since QCD is simulated using the PYTHIA $2 \rightarrow 2$ processes.

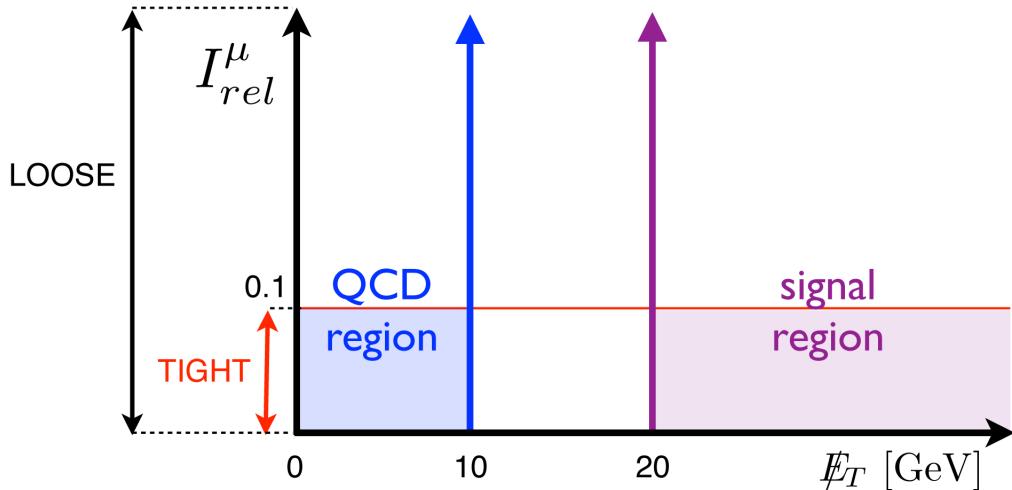


Figure 5.4: ε_{QCD} is measured in the QCD region ($\cancel{E}_T < 10$ GeV), while the estimation of the QCD multijet background is done in the signal region ($\cancel{E}_T > 20$ GeV). The tight sample is a subset of the loose sample consisting of those events with $I_{rel}^\mu < 0.1$.

Table 5.5: Results of the selection in the QCD region ($\cancel{E}_T < 10$ GeV): the number of events passing the preselection and $\cancel{E}_T < 10$ GeV is denoted by $N_T^{\cancel{E}_T < 10}$. The results for data and the Monte Carlo samples of QCD and $t\bar{t}$ are shown.

N_{jets}	$N_T^{\cancel{E}_T < 10}$ data	$N_T^{\cancel{E}_T < 10}$ QCD	$N_T^{\cancel{E}_T < 10}$ $t\bar{t}$
= 1	6426	4385	1
= 2	782	408	3
= 3	113	50	5
≥ 4	30	9	7

With the measurements of ε_{sig} and ε_{QCD} at hand, the linear system in Equation 5.4 can be solved for N^{QCD} and N^{Sig} :

$$\begin{aligned} N^{sig} &= \frac{N_T - \varepsilon_{QCD} \cdot N_L}{\varepsilon_{sig} - \varepsilon_{QCD}}, \\ N^{QCD} &= \frac{\varepsilon_{sig} \cdot N_L - N_T}{\varepsilon_{sig} - \varepsilon_{QCD}}. \end{aligned} \quad (5.5)$$

The number of W -like events in the preselected sample is obtained as $N_T^{Sig} = \varepsilon_{sig} N^{Sig}$, and the number of QCD multijet events as $N_T^{QCD} = \varepsilon_{QCD} N^{QCD}$. The method relies on the fact that the fake rate measured in the QCD sample is independent of the \cancel{E}_T in the event. As shown in Figure 5.5, the fake rate distribution for the QCD Monte Carlo sample for the events with two or more jets is flat with respect to the missing transverse energy; the fluctuations observed at higher \cancel{E}_T are due to the lack of statistics. The uncertainty lines represent the total uncertainty used in data, as described in section 5.3.2. Figures 5.6 and 5.7 show the two-dimensional scatter plot of the events with at least two jets as a function of the relative isolation and \cancel{E}_T for data and the Monte Carlo simulated samples ($t\bar{t}$, $W+jets$ and QCD). Figure 5.7 also indicates that the main contribution in the QCD region ($\cancel{E}_T < 10$ GeV) comes from the QCD multijet events.

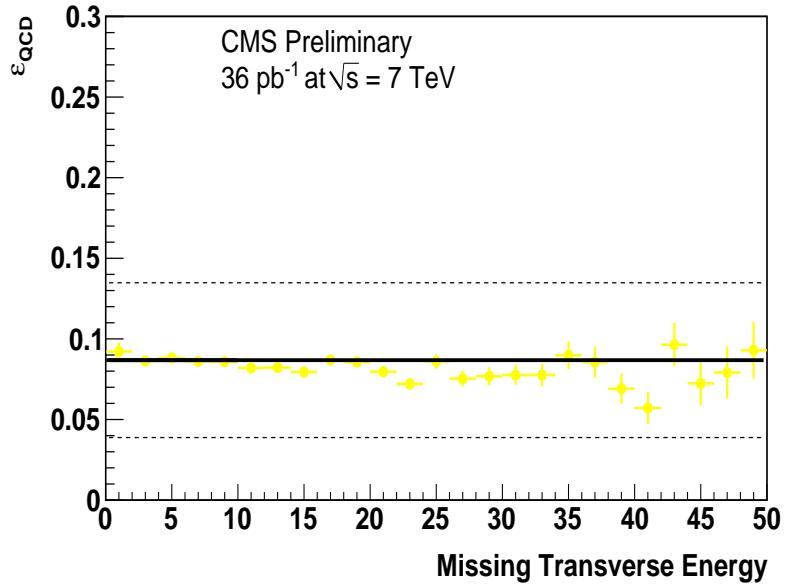


Figure 5.5: ε_{QCD} for the QCD Monte Carlo sample as a function of \cancel{E}_T . The dashed lines represent the total uncertainty used in data.

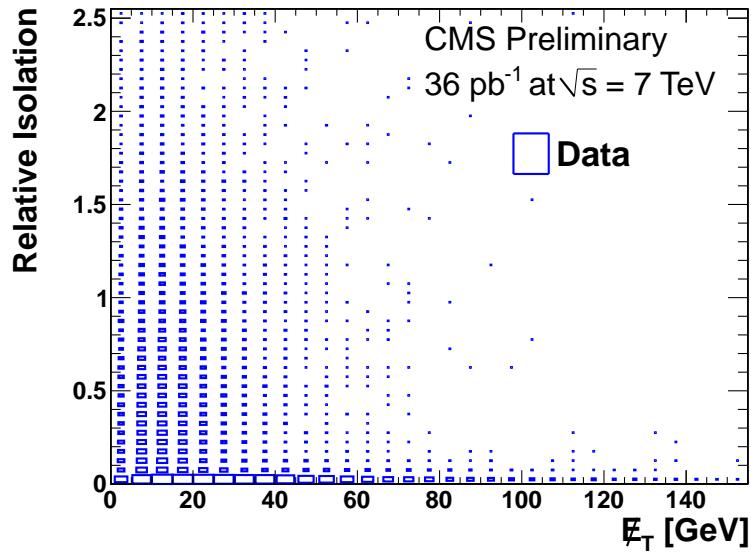


Figure 5.6: The two-dimensional scatter plot of events with at least two jets, selected as a function of the Relative Isolation and \cancel{E}_T for data. The sizes of the rectangles represent the numbers of events.

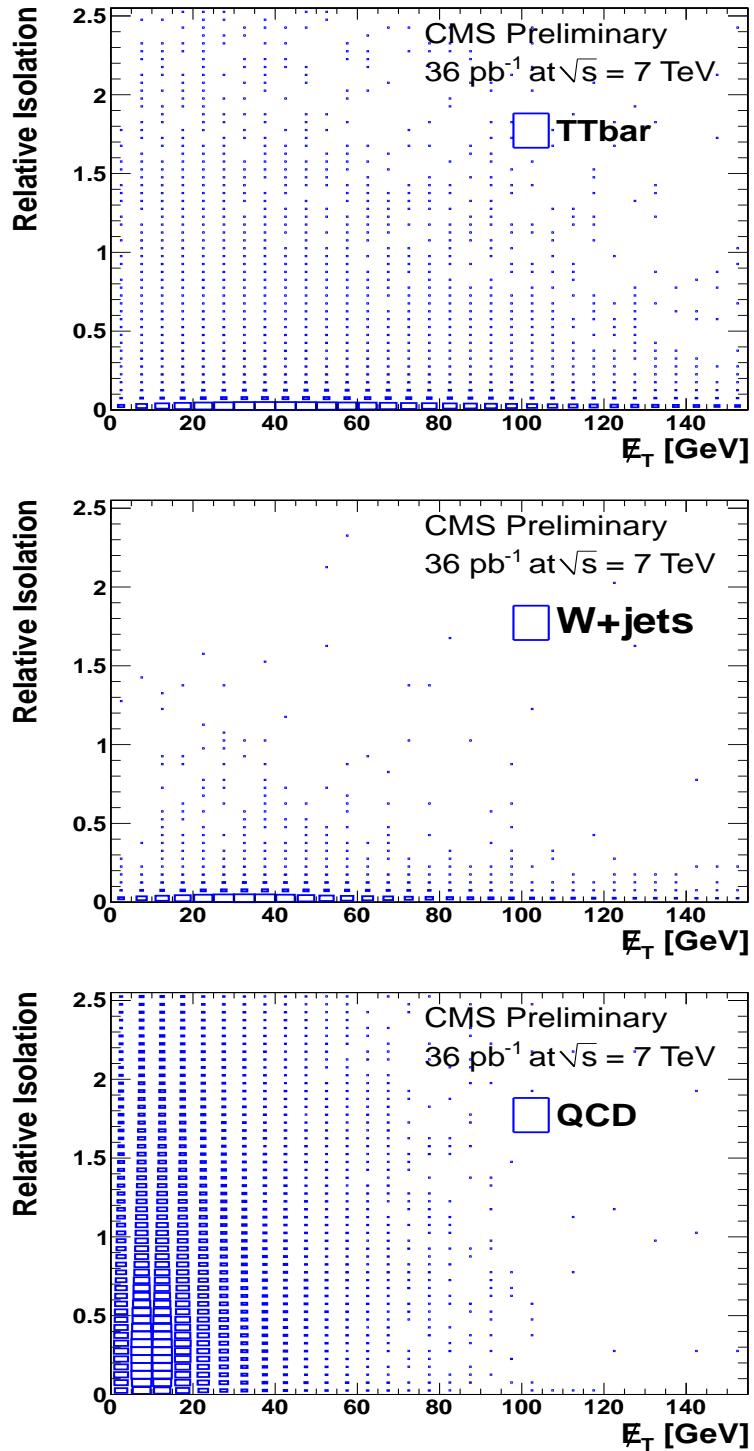


Figure 5.7: The two-dimensional scatter plots of events with at least two jets, selected as a function of the Relative Isolation and E_T for the $t\bar{t}$, $W+\text{jets}$ and QCD Monte Carlo samples. The sizes of the rectangles represent the numbers of events.

ε_{QCD} is sensitive to the threshold used to define the QCD-dominated region because of the “signal contamination,” mainly from W +jets and Drell-Yan processes. The choice of the threshold is limited by the statistics. In order to get a purer QCD sample, one can subtract the signal contamination present in the region $\cancel{E}_T < 10$ GeV by using the Monte Carlo predictions. The values of ε_{QCD} measured for each jet multiplicity bin, before and after subtracting the $W(Z)$ contribution from the $\cancel{E}_T < 10$ GeV region are summarized in Table 5.6. One can conclude from Table 5.6 that the values of ε_{QCD} corresponding to the exclusive two- and three-jet bins and the inclusive four-jet bin are comparable within statistical errors. We therefore use ε_{QCD} measured for the inclusive two-jet bin to estimate the background in the two-, three- and four-or-more jet bins. Because the topology for events with one jet is very different, we use the ε_{QCD} value obtained in the one-jet bin sample to estimate the background for those events.

Table 5.6: ε_{QCD} values measured before subtracting $W(Z)$ contributions, denoted as default, and after subtracting the $W(Z)$ contributions, denoted as $W(Z)$ subtracted, from the $\cancel{E}_T < 10$ GeV region. The values are shown with statistical uncertainties only.

N_{jets}	default	$W(Z)$ subtracted
= 1	0.231 ± 0.003	0.189 ± 0.002
≥ 2	0.209 ± 0.006	0.151 ± 0.006
= 2	0.208 ± 0.007	0.152 ± 0.006
= 3	0.211 ± 0.018	0.140 ± 0.016
≥ 4	0.242 ± 0.039	0.165 ± 0.035

5.3.2 The Results

The results obtained with the Matrix Method are summarized in Table 5.7. ε_{QCD} is estimated with the signal contamination, calculated using the Monte Carlo samples for W +jets and $Z/\gamma^* \rightarrow \ell\ell$, subtracted from the data in the $\cancel{E}_T < 10$ GeV region; it is determined for the 1-jet sample and 2plus-jet sample; the contribution from the $t\bar{t}$ and

single-top processes in these samples in the QCD region is negligible and therefore ignored.

As shown in Table 5.6, the ε_{QCD} values depend on the amounts of W -like contamination in the QCD region. Therefore, the difference between the values of the ε_{QCD} measured with and without subtracting the contributions from the $W(Z)$ events in the QCD region is taken as the systematic uncertainty. This uncertainty also covers the differences between the calculated ε_{QCD} for each jet bin and the value used in the Matrix Method for the events with two, three and four or more jets. Hence, the total uncertainty assigned to ε_{QCD} is a combined binomial uncertainty and the systematic uncertainty. A standard error propagation is used to derive the uncertainties on the number of W -like and QCD multijet events. Table 5.8 shows the scale factors as a function of jet multiplicity. The scale factors were calculated for both QCD and W -like events, as being the ratio of the number of events estimated in data with the Matrix Method to the number of events predicted by the Monte Carlo simulation.

Table 5.7: The results of the Matrix Method. The numbers of W -like and QCD multijet events in the candidate sample are determined with the W and Z boson contamination subtracted in the $\cancel{E}_T < 10$ GeV region. The ε_{QCD} value measured for the events with at least two jets is used for the events with two, three and four or more jets. Both statistical and systematic uncertainties are shown for ε_{QCD} .

N_{jets}	N_{sig}	N_{QCD}	ε_{sig}	ε_{QCD}
1	18358 ± 483	1654 ± 461	0.98 ± 0.02	$0.189 \pm 0.002 \pm 0.042$
2	4113 ± 192	393 ± 180	0.97 ± 0.02	$0.151 \pm 0.006 \pm 0.058$
3	1003 ± 60	108 ± 49	0.97 ± 0.02	$0.151 \pm 0.006 \pm 0.058$
≥ 4	426 ± 26	33 ± 15	0.96 ± 0.02	$0.151 \pm 0.006 \pm 0.058$

Table 5.8: Scale factors for the estimated number of QCD multijet events, denoted SF_{QCD} , and W -like events, denoted SF_{signal} , calculated with respect to Monte Carlo predictions.

N_{jets}	SF_{signal}	SF_{QCD}
1	0.92 ± 0.01	2.54 ± 0.12
2	0.97 ± 0.02	2.10 ± 0.19
3	1.08 ± 0.05	2.57 ± 0.47
≥ 4	0.96 ± 0.07	3.98 ± 1.55

5.4 Comparison between Data and Monte Carlo Predictions

To test how well the Monte Carlo modeling represents the data, a direct comparison of various kinematic distributions between the data and the [Monte Carlo](#) (MC) predictions with the relevant scale factors derived in the previous sections applied to the background Monte Carlo samples has been performed. For simplicity, we construct the W -like sample from the following simulated events: W , Z/γ^* , $t\bar{t}$, single top, and diboson events. As the QCD multijet processes are not well modeled by the PYTHIA Monte Carlo simulation, a data-driven QCD template is adopted in replacement of the QCD simulated background.

5.4.1 Data-driven QCD multijet template

The data-driven QCD template is obtained from the ‘orthogonal’ data sample which is composed of events that have a muon relative isolation satisfying the condition $0.15 < I_{rel}^\mu < 0.7$ or a muon transverse impact parameter calculated with respect to the beamspot satisfying $|d_0| > 0.025$ cm. An orthogonal QCD Monte Carlo sample is also derived in the same way. Figure 5.8 shows the distributions of the transverse mass of the W boson, reconstructed from the muon momentum and the \cancel{E}_T , for the events with at least one jet in the orthogonal data sample and the orthogonal QCD [Monte Carlo](#) sample. The transverse mass is calculated as

$$M_T = \sqrt{2 p_T^\ell p_T^\nu (1 - \cos(\Delta\varphi^{\ell\nu}))}, \quad (5.6)$$

where $p_T^\nu = \cancel{E}_T$ and $\Delta\varphi^{\ell\nu}$ (or $\Delta\phi(\ell, \cancel{E}_T)$) is the opening angle between the lepton, muon in our case, and the missing transverse energy. For both samples one can see a small bump

around 40 GeV/c², which consists mostly of the events that have a muon pointing in the opposite direction of the \cancel{E}_T in the transverse plane, i.e., those events have large $\Delta\phi(\mu, \cancel{E}_T)$ values. This effect is enhanced due to the muon p_T and \cancel{E}_T thresholds both at 20 GeV/c. One can also observe that the bump in the [Monte Carlo](#) sample is significantly smaller than the one in the data.

The events with back-to-back muon and \cancel{E}_T contributing to the region with a transverse mass around 40 GeV/c² can be filtered mostly by requiring $\Delta\phi(\mu, \cancel{E}_T) < 2$ rad. Figure 5.8 also shows the effect of this requirement in data and [Monte Carlo](#), respectively. The resulting distribution (blue) has a falling spectrum similar to the simulated QCD multijet events with the full event selection (see Figure 5.9). No significant effect is observed when the same requirement is applied to the signal QCD [Monte Carlo](#) sample. Therefore, we use the orthogonal data sample with the requirement of $\Delta\phi(\mu, \cancel{E}_T) < 2$ rad to extract the template for the QCD multijet background.

5.4.2 Data/Monte Carlo Comparisons

Figures 5.10 and 5.11 show comparisons between the QCD multijet shapes obtained from data and the QCD shapes simulated with Monte Carlo for the muon transverse momentum, the missing transverse energy, and the transverse mass of the W boson for the one- and two-jet samples and the three-jet and four-or-more-jet samples, respectively. The [Monte Carlo](#) QCD shapes have poor statistics and do not describe the data well. Therefore, we choose to use the data-driven QCD multijet shape from the orthogonal data sample for the rest of the analysis.

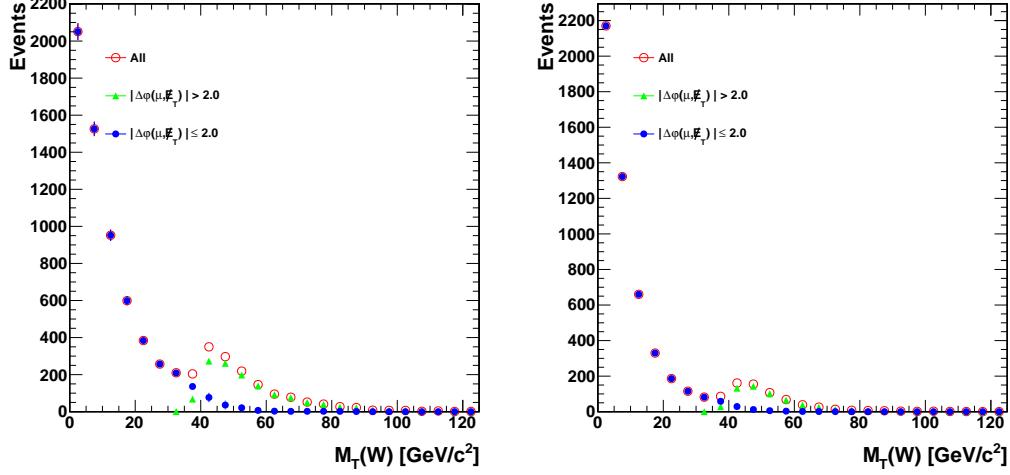


Figure 5.8: Transverse mass of the reconstructed W boson for events with at least one jet in the orthogonal data sample (left) and in the orthogonal QCD Monte Carlo sample (right). A small bump can be observed around $40 \text{ GeV}/c^2$. Most of the events that contribute to the bump have a muon which is opposite to the direction from the E_T in the transverse plane, i.e., those events have large $\Delta\phi(\mu, E_T)$ values. This effect is enhanced due to the muon p_T and E_T thresholds both at $20 \text{ GeV}/c$.

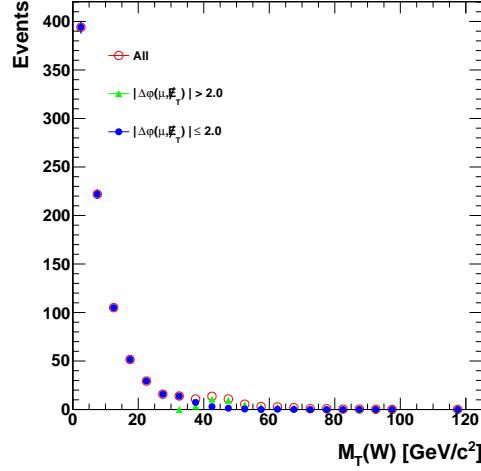


Figure 5.9: Transverse mass of the W boson using the QCD Monte Carlo sample with at least one jet. The requirement of $\Delta\phi(\mu, E_T) < 2 \text{ rad}$ has a small effect on the QCD Monte Carlo sample.

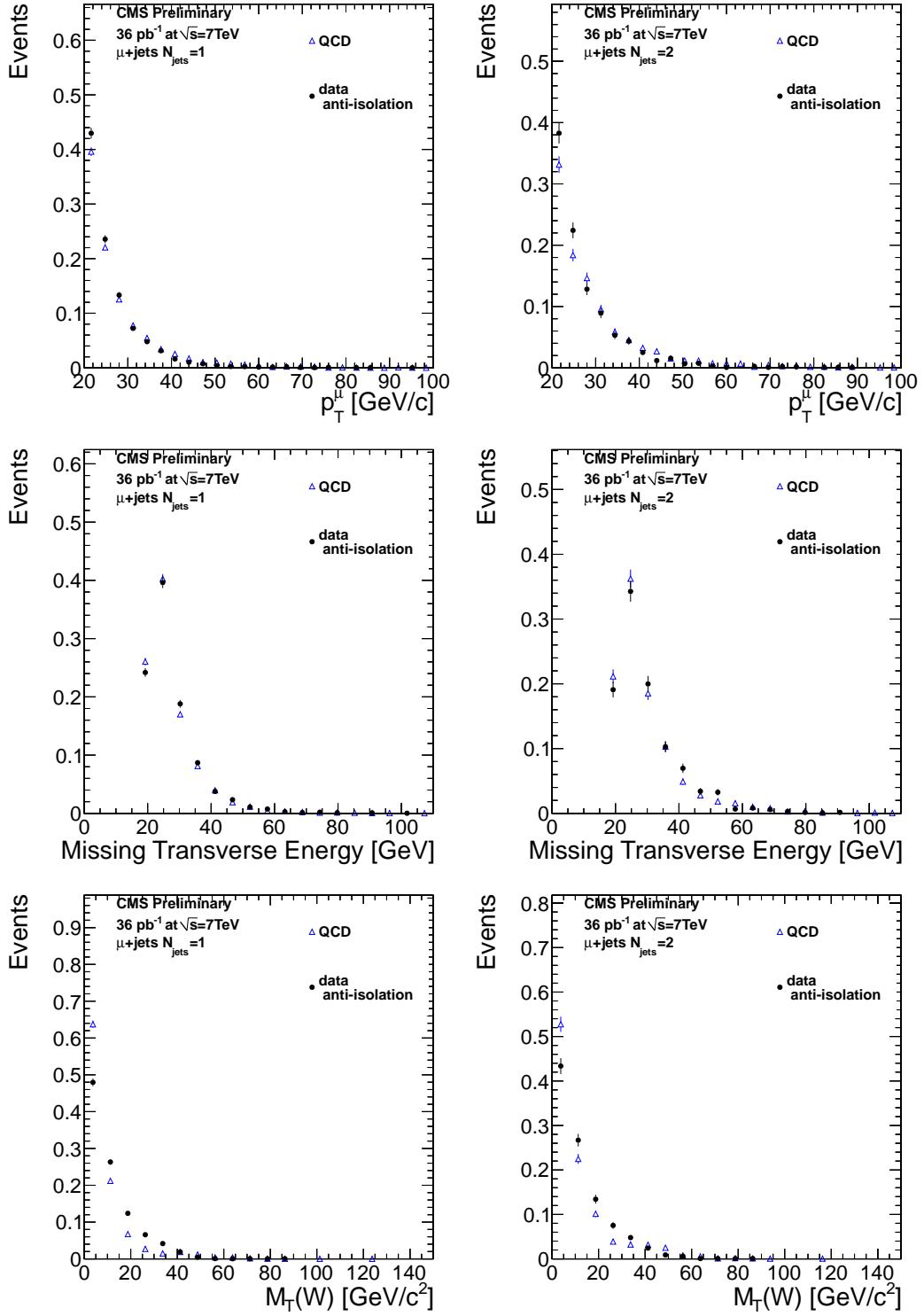


Figure 5.10: Shape comparisons between QCD Monte Carlo and the orthogonal sample for muon p_T (upper plots) and W transverse mass (lower plots) for events with 1 (left) and 2 (right) jets.

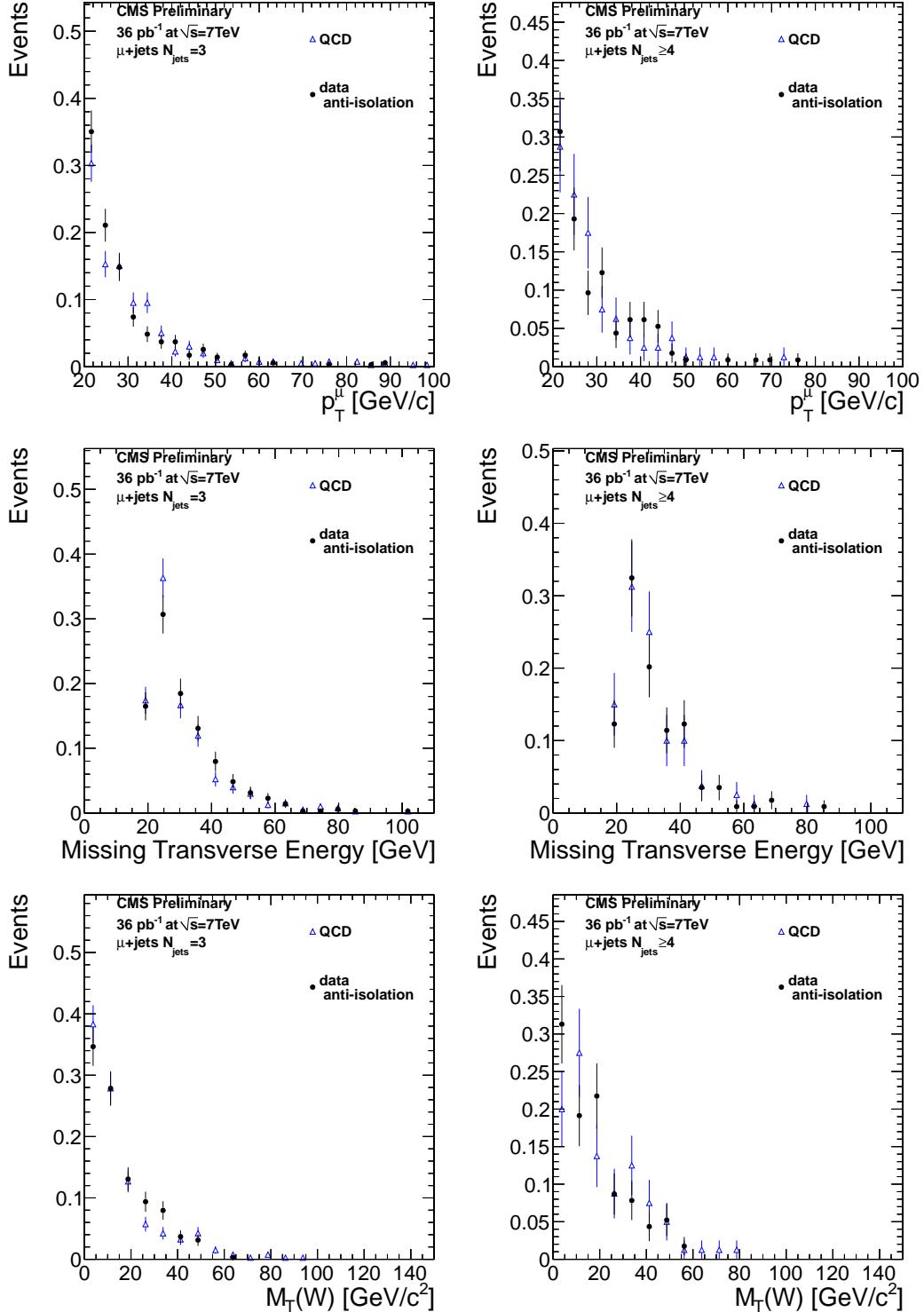


Figure 5.11: Shape comparisons between QCD Monte Carlo and the orthogonal sample for muon p_T (upper plots) and W transverse mass (lower plots) for events with 3 (left) and 4 or more (right) jets.

To compare the kinematic distributions between data and the [Monte Carlo](#) predictions, the data-driven QCD multijet shape is normalized to the prediction of the MM given in Table 5.7; the $W + \text{jets}$ and $Z/\gamma^* \rightarrow \ell\ell$ [Monte Carlo](#) samples are scaled according to the MM scale factors shown in Table 5.8; and the $t\bar{t}$ [Monte Carlo](#) is normalized to the NLO theoretical prediction of $\sigma(t\bar{t}) = 157.5^{+23.2}_{-24.4} \text{ pb}$ [79–81]. This procedure is adopted only for the purpose of comparing data and [Monte Carlo](#) predictions. To extract the cross section we use only the prediction of the MM for the QCD multijet background to normalize the shape taken from the orthogonal sample (see Section 6).

Figure 5.12 shows the distribution of the jet multiplicity for data compared to the [Monte Carlo](#) prediction. We also include in the plot a 45% uncertainty assigned to the QCD multijet sample from the results of the MM. The ratio between data and [Monte Carlo](#) is shown in the bottom plot for each jet bin with only the statistical uncertainty shown in this plot. One can see that the data and the [Monte Carlo](#) predictions are in good agreement.

The comparison between the data and the Monte Carlo predictions is shown for the following distributions: muon transverse momentum (Figs. 5.13 and 5.14), missing transverse energy (Figs. 5.15 and 5.16), transverse mass of the W boson (Figs. 5.17 and 5.18), and the mass distribution of the combination of three jets with the highest sum \vec{p}_T (Fig. 5.19). The data/Monte Carlo ratio plots for each distribution are also shown. Overall the [Monte Carlo](#) predictions are fairly consistent with the data.

In summary, from Figs. 5.12 to 5.19, we observe that the agreement between the data sample and the Monte Carlo predictions is very good within the 45% uncertainty assigned to the QCD contribution (obtained from data).

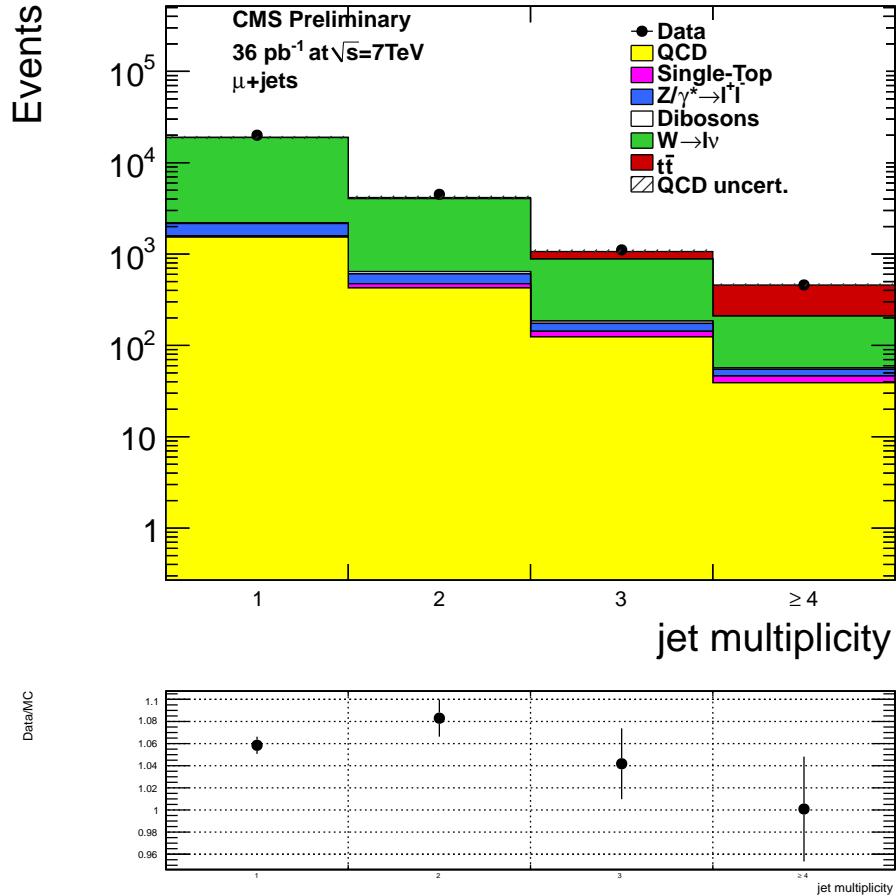


Figure 5.12: Data/Monte Carlo comparisons of the jet multiplicity. The QCD multijet contribution is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z/\gamma^*)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

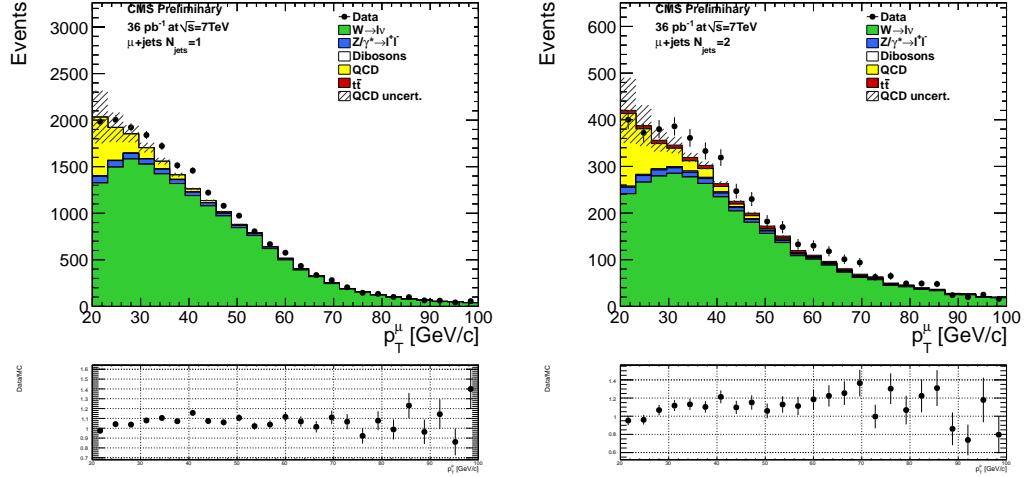


Figure 5.13: Data/Monte Carlo comparisons of the muon p_T for the 1-jet (left) and 2-jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)+\text{jets}$ Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

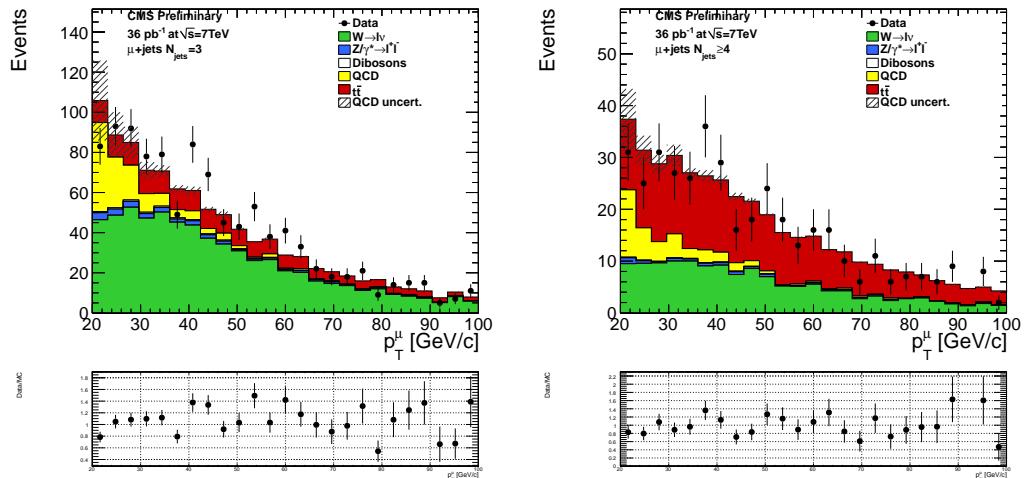


Figure 5.14: Data/Monte Carlo comparisons of the muon p_T for the 3-jet (left) and inclusive four jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)+\text{jets}$ Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

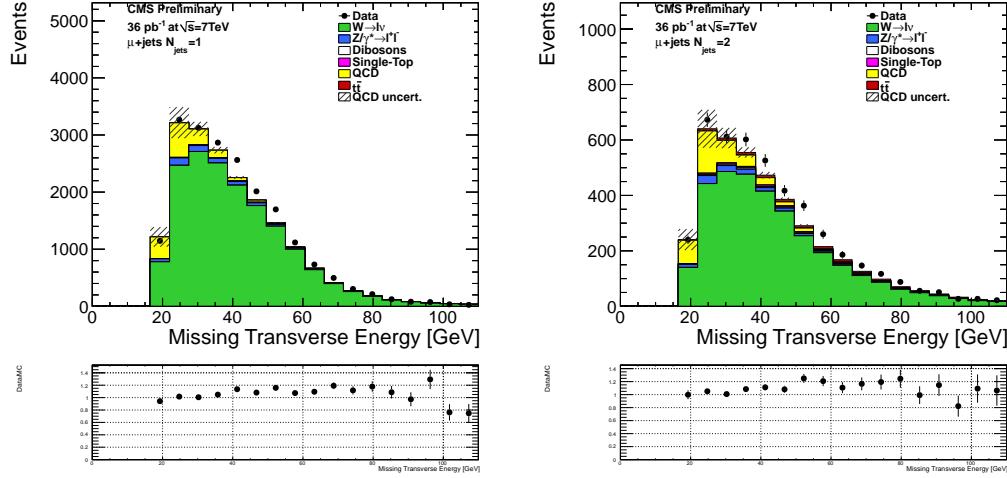


Figure 5.15: Data/Monte Carlo comparisons of the missing transverse energy for the 1-jet (left) and 2-jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)+\text{jets}$ Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

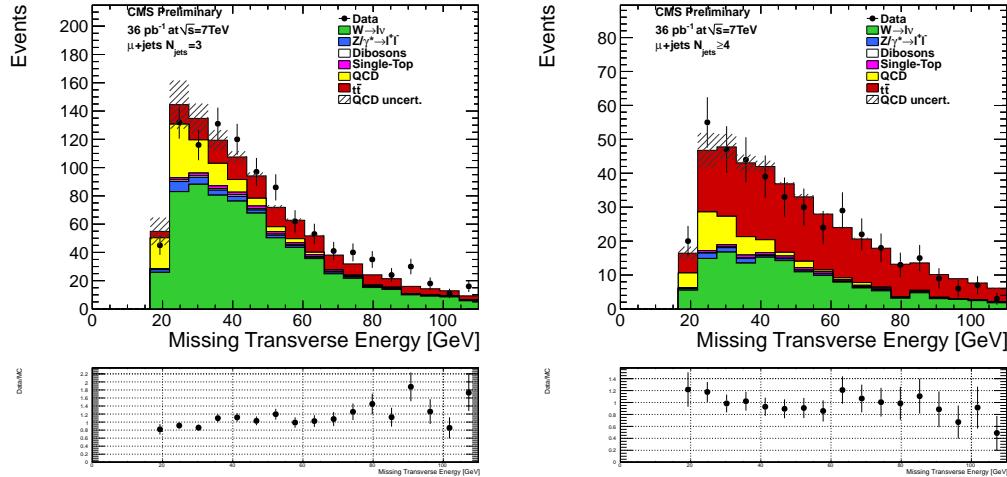


Figure 5.16: Data/Monte Carlo comparisons of the missing transverse energy for the 3-jet (left) and inclusive four jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)+\text{jets}$ Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

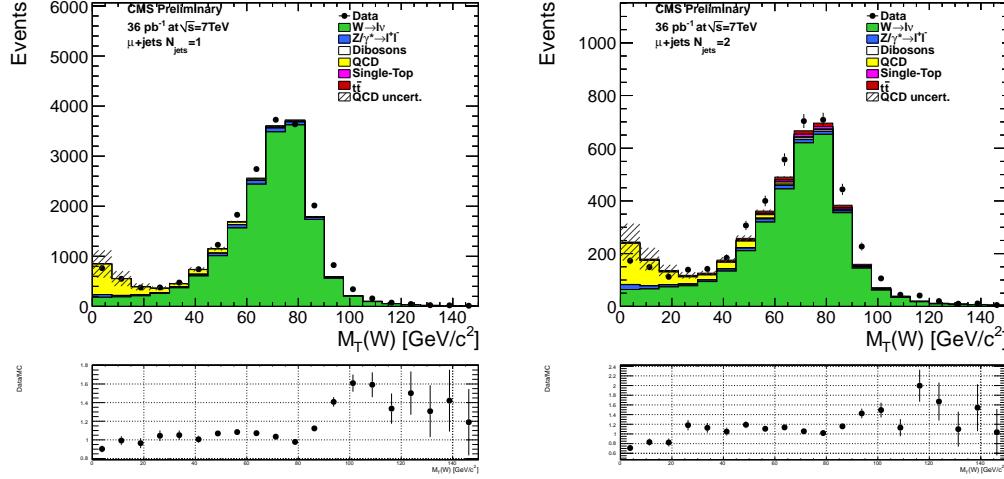


Figure 5.17: Data/Monte Carlo comparisons of the W M_T for the 1-jet (left) and 2-jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

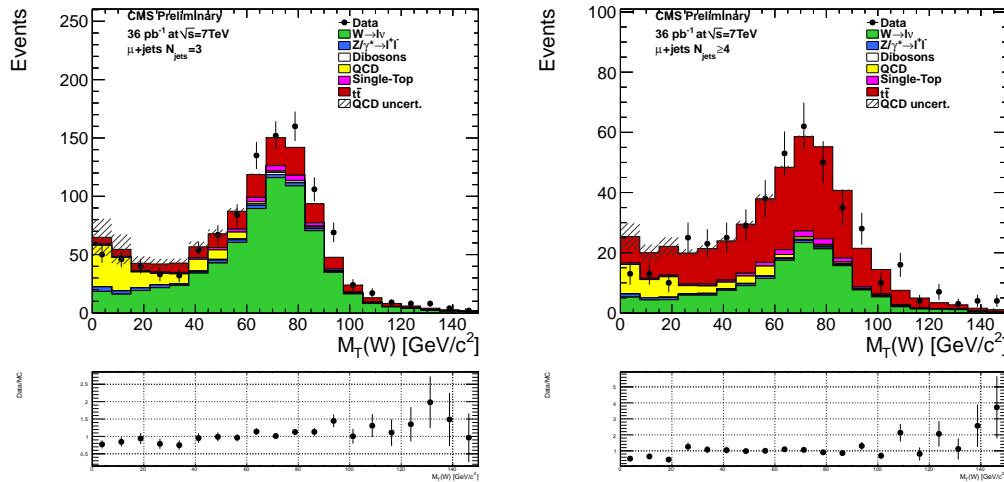


Figure 5.18: Data/Monte Carlo comparisons of the W M_T for the 3-jet (left) and inclusive four jet (right) samples. The shape for the QCD multijet sample is taken from the orthogonal sample and is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)$ +jets Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

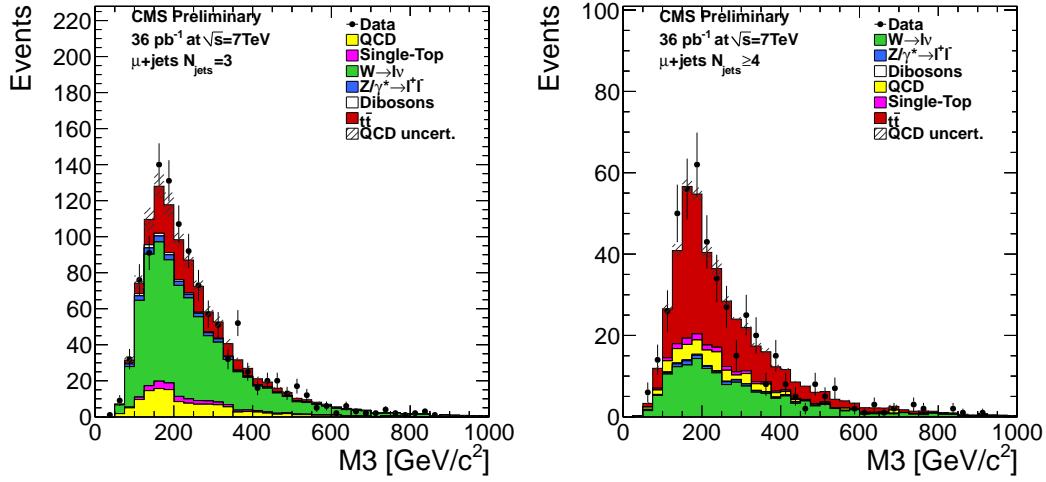


Figure 5.19: Data/Monte Carlo comparisons of M_3 for the 3-jet (left) and inclusive four jet (right) samples. M_3 is the invariant mass distribution of the combination of three jets with the largest magnitude of the vectorial sum of the jet p_T . The shape for the QCD multijet sample is taken from data with a reverse isolation criteria and the sample is normalized to the results from the MM. An uncertainty of 45% is assigned to the QCD contribution. The $W(Z)+\text{jets}$ Monte Carlo is also scaled to the predictions of the MM while the $t\bar{t}$ Monte Carlo is normalized to the SM theoretical prediction. The bottom plot shows the ratio of data to Monte Carlo, with only the statistical uncertainties shown.

Chapter 6

Measurement of the $t\bar{t}$ Cross Section

As described in Section 5.1, the preselected data sample is enriched with $t\bar{t}$ signal for events with at least three jets. We therefore include in this analysis the samples with exactly three jets ([3-jet sample](#)) and those with at least four jets ([4plus-jet sample](#)). To extract the $t\bar{t}$ production cross section, one needs to determine the composition of the contributing processes in the sample. The QCD multijet contribution has been estimated with the Matrix Method, as described in Section 5.3. The largest background from $W+jets$ events has yet to be determined. This chapter is dedicated to giving a complete account of measuring the $t\bar{t}$ production cross section. It is organized as follows: The method to extract the number of events for the $t\bar{t}$ signal is described in Section 6.1. The study of the fit stability using a [Monte Carlo](#) pseudo-experiments technique is discussed in Section 6.1.1, and the results of the method are summarized in Section 6.1.2. The estimation of the systematic uncertainties of the method is reported in Section 6.2.

6.1 The Method

To extract the number of $t\bar{t}$ events from the preselected data sample described in Section 5.1, we perform a simultaneous fit, which is an extended binned maximum likelihood fit using the **ROOFIT** [91] package in **ROOT** [92], on the tri-jet invariant mass (M_3) distribution for the [3-jet sample](#) and the [4plus-jet sample](#). M_3 is the invariant mass of the combination of three jets with the largest value of the vectorial sum of the jet \vec{p}_T . One of the advantages of using **ROOFIT** is that it handles the error propagation properly by using the full covariance matrix. The fit estimates the numbers of $t\bar{t}$, $W+\text{jets}$, $Z+\text{jets}$, diboson, single top, and QCD multijet events satisfying:

$$N_{data}^{n-\text{jets}} = N_{t\bar{t}}^{n-\text{jets}} + N_{W\text{jets}}^{n-\text{jets}} + N_{Z\text{jets}}^{n-\text{jets}} + N_{WW}^{n-\text{jets}} + N_{\text{single top}}^{n-\text{jets}} + N_{QCD}^{n-\text{jets}}, \quad (6.1)$$

where $n-\text{jets}$ is either the [3-jet sample](#) or the [4plus-jet sample](#). The single top and diboson contributions are fitted with a Gaussian constraint centered at the [Monte Carlo](#) prediction within a 30% uncertainty [82–84]. The QCD multijet contribution is also constrained with a Gaussian distribution with a mean given by the predictions of the Matrix Method and a width given by the estimated uncertainty of 45% (see Table 5.7). Another nuisance parameter is added to the fit for the $\sigma_{W+\text{jets}}/\sigma_{Z+\text{jets}}$ ratio with an uncertainty of 30% [93]. The ratios (Eq. 6.2) of the number of events in the four-or-more-jet sample to the number of events in the three-jet sample are taken from Monte Carlo and are left floating within a 30% uncertainty while estimating the systematics.

$$R_k^{4/3} = \frac{N_k^{n-\text{jets} \geq 4}}{N_k^{n-\text{jets} = 3}}, \quad k = t\bar{t}, W+\text{jets}, Z+\text{jets}. \quad (6.2)$$

The assumption is applied to the $t\bar{t}$, W +jets, and Z +jets events. The probability density functions (or templates) are taken from the Monte Carlo simulation except for the QCD multijet background, which is obtained from the orthogonal data sample as described in more detail in Section 5.4.1. The templates that were used in the simultaneous fit to extract the number of $t\bar{t}$ events are shown in Figure 6.1. One can see a clear difference between the $t\bar{t}$ templates and the others. The top quark pair production cross section is calculated using

$$\sigma(t\bar{t}) = \frac{N_{t\bar{t}}}{A \cdot \varepsilon \cdot L}, \quad (6.3)$$

where L is the total integrated luminosity, $N_{t\bar{t}}$ is the total number of fitted $t\bar{t}$ events in the sample with three or more jets obtained from the fit. The $A \cdot \varepsilon$ is the acceptance times the efficiency to select an isolated muon in the sample with three or more jets. Using the $t\bar{t}$ Monte Carlo sample, we obtain $A \cdot \varepsilon = 0.075 \pm 0.003$.

6.1.1 Study of the Fit Stability

The stability of the simultaneous fit has been studied using pseudo-samples. We generate 1000 pseudo-experiments (PEs) with the total number of events distributed following a Poisson distribution. In each PE, the pseudo-sample is fitted and the results are stored in order to build the pull distributions of the fit parameters. The fit results for the number of $t\bar{t}$ events in the 3-jet sample and $R_{t\bar{t}}^{4/3}$ (defined in Eq. 6.2) are shown in the upper set and the lower set of plots, respectively, in Fig. 6.2. The left column of plots shows the distribution of the fitted values. The plots in the middle column are the distributions of

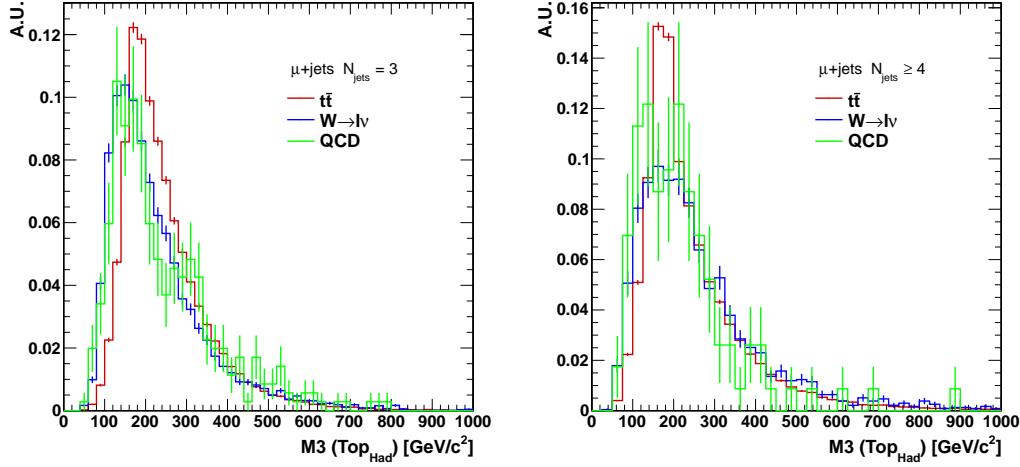


Figure 6.1: The probability density functions, i.e., templates for M_3 , used in the simultaneous fit. On the left, the shapes of the templates in the sample with three jets are shown. On the right, the template shapes for the sample with at least four jets are shown. The shapes for the QCD multijet events are taken from the orthogonal data sample as described in Section 5.4.1. The shapes of the $t\bar{t}$ events are clearly distinguishable from other processes.

the fit uncertainties. The plots on the right are the pull distributions defined as

$$\text{pull} = \frac{N_{fit} - N_{MC}}{\sigma_{fit}}. \quad (6.4)$$

The pull distribution is fitted with a Gaussian distribution to determine its mean and variance. The mean is close to zero, which means the fit itself gives an unbiased central value. However, the variance is less than one, indicating an overestimation of the fit errors. Therefore, the statistical uncertainties of the fit results have been corrected accordingly.

6.1.2 The Results

Figure 6.3 shows the results of the simultaneous fit of the M_3 distributions for the three-jet and four-or-more-jet samples. The template for each background contribution has been normalized to the results of the fit. The bottom plot shows the visualization of the

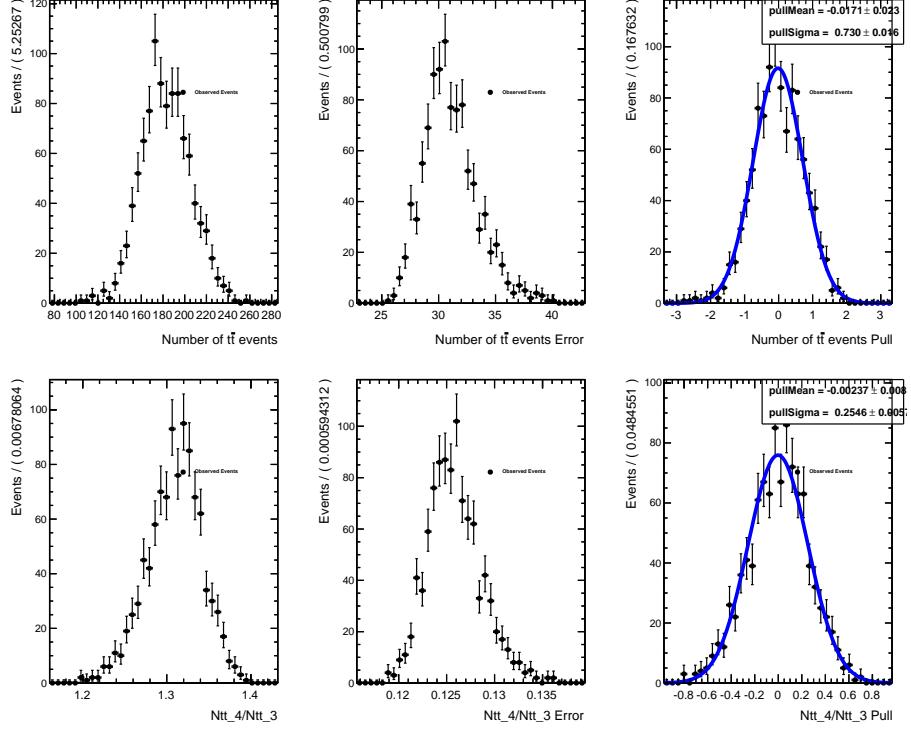


Figure 6.2: The stability of the fit is studied using 1000 pseudo-experiments. The upper set of plots are the fit value for the total number of $t\bar{t}$ events in the three-jet sample; the lower plots are the fit value for the ratio of the number of four-or-more-jet events to the number of three-jet events ($R_{t\bar{t}}^{4/3}$). The plots in the middle column are the distributions of the fit errors on the number of fitted $t\bar{t}$ events. The plots on the right are the pull distributions fitted separately with a Gaussian distribution to determine the mean and variance. The mean is close to zero, which means the fit itself gives an unbiased central value. However, the variance being lower than one indicates that the errors are overestimated.

covariance matrix. The number of $t\bar{t}$ events is highly anti-correlated with the number of $W+jets$ events. The fitted numbers of events and cross section are summarized in Table 6.1.

Table 6.1: Results of simultaneously fitting the M3 distributions.

$N_{t\bar{t}}$	N_{W+jets}	N_{Z+jets}	$\sigma \pm \Delta\sigma$ (stat.) [pb]
424.07 ± 33.73	867.01 ± 38.29	47.17 ± 16.32	157.09 ± 12.49

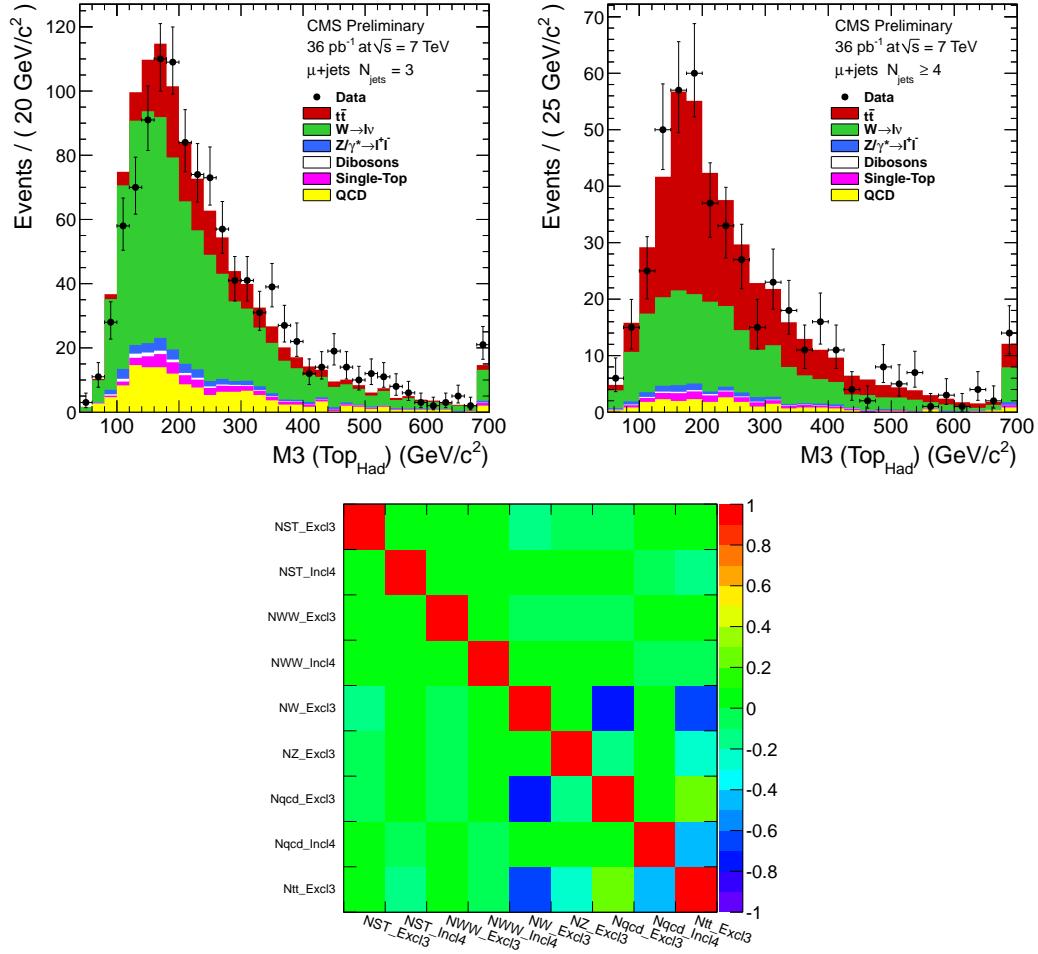


Figure 6.3: The results of the M_3 simultaneous fit to the three-jet (top-left figure) and four-or-more-jet (top-right figure) samples. The MC templates are normalized to the results of the fit. The bottom plot shows the visualization of the covariance matrix.

6.2 Estimation of the Systematic Uncertainties

It is important to understand the systematic impact on extracting the number of $t\bar{t}$ events due to the uncertainties from the theoretical modeling and the experimental setup. We have studied various sources of systematic uncertainties: absolute [Jet Energy Scale \(JES\)](#) in Section 6.2.1, [Jet Energy Resolution \(JER\)](#) in Section 6.2.2, factorization scale (μ_F) in Section 6.2.3, Matrix Element-Parton Shower ([ME-PS](#)) matching threshold in Section 6.2.4, initial- and final-state radiation in Section 6.2.5, [Underlying Event \(UE\)](#) modeling in Section 6.2.6, and the parton distribution function ([PDF](#)) systematic uncertainty in Section 6.2.7.

Unless stated otherwise, the common procedure we adopted to estimate the effect of different sources of systematic uncertainty is as follows: First, for each source of systematic error we construct the M3 templates by linearly interpolating [94] the templates from the systematic and nominal [Monte Carlo](#) samples (systematic MC samples are generated with control parameters varied from the default values used in generating the nominal samples, e.g., $\mu_F = m_{\text{top}}/2$ instead of $\mu_F = m_{\text{top}}$, or re-reconstructed from the nominal samples with different calibration values, such as a 10% increase in the [JES](#)); this step is done with the class “RooIntegralMorph” [95] of [ROOFIT](#) as follows:

- Define a nominal template $f_1(x)$ with a cumulative distribution function ([c.d.f.](#)) $F_1(x)$ and a systematic template (from the corresponding systematic sample) $f_2(x)$ with a [c.d.f.](#) $F_2(x)$.
- Find a value ‘y’ of both [c.d.f.s](#) and determines the corresponding x values (x_1, x_2) at which $F_1(x_1) = F_2(x_2) = y$.

- The value of the interpolated template $\overline{f(x)}$ is then calculated as:

$$\overline{f(x')} = \frac{f_1(x_1) \times f_2(x_2)}{\alpha f_2(x_2) + (1-\alpha)f_1(x_1)}; \quad x' = \alpha x_1 + (1 - \alpha)x_2, \quad (6.5)$$

where the templates $f_1(x)$ and $f_2(x)$ are always taken to be the end-points of the α .

The QCD multijet template is taken from the orthogonal data sample. Second, we throw 1000 pseudo-experiments (**PEs**) with the new templates to produce the pseudo-data for each source. For each **PE** the value of α is chosen randomly between zero (corresponding to the systematic templates) and one (corresponding to the nominal templates). Third, the pseudo-data are fitted by the nominal templates to extract the cross sections. Finally, the distributions of the extracted cross sections are fitted with a Gaussian distribution, denoted by $G(\mu, \sigma^2)$; the width of the distribution is taken as the 1σ uncertainty. For each uncertainty determination, the cross section is determined for an upper and a lower variation of the source. Assuming the upper variation gives a $G(\mu_1, \sigma_1^2)$ and the lower variation gives a $G(\mu_2, \sigma_2^2)$, then the upper- and lower-uncertainty (σ^+ and σ^-) are taken as the boundaries of the union of the variations:

$$\sigma^+ = MAX(\mu_1 + \sigma_1 - \mu_0, \quad \mu_2 + \sigma_2 - \mu_0), \quad (6.6)$$

$$\sigma^- = MIN(\mu_1 - \sigma_1 - \mu_0, \quad \mu_2 - \sigma_2 - \mu_0), \quad (6.7)$$

where μ_0 is the input cross section of 157.5 pb [79–81]. The systematic uncertainties are summarized in Section 6.2.8.

6.2.1 Jet Energy Scale

In order to estimate the effect on the cross section measurement due to the limited knowledge of the jet energy scale in the calorimeters, we derive the systematic templates for the $t\bar{t}, W + \text{jets}$, $Z + \text{jets}$, and single top samples by varied the four-momentum of jets with jet-by-jet corrections. The overall correction for each jet includes several contributions: First, a 1σ uncertainty calculated according to the p_T - and η -dependent jet energy uncertainties. Second, a 1.5% flat uncertainty for calibration changes. Third, a p_T -dependent pile-up uncertainty of $1.32/p_T$. Fourth, if the jet matches to a b quark on the parton level, we use an uncertainty of 2% if the jet satisfies $50 < p_T < 200$ GeV/c and $|\eta| < 2.0$, or an uncertainty of 3% for all other jets matched to b quarks. All these uncertainties are added together in quadrature to give a total correction for each jet. The corrections are then propagated to \cancel{E}_T according to: $\Delta \cancel{E}_T = - \sum \Delta \cancel{P}_{T,jet}$. The correction on the \cancel{E}_T affects the event yields slightly; however, its effect on the shape of the M3 distributions is not significant. After applying all the corrections, the full event selection is applied to derive the M3 templates from all the contributing Monte Carlo samples. The templates are added together according to the theoretical predictions (see Table 4.5) to create an overall template for the systematic pseudo-data. We follow the procedure described in the previous section to estimate the systematic uncertainties due to JES, resulting in +4.5% and -8.7% uncertainties. The interpolated templates are shown in Figure 6.5 for varying the JES correction up and down. Figure 6.6 shows the selected pseudo-experiments and the corresponding fit results.

6.2.2 Jet Energy Resolution

Jet asymmetry measurements suggest that the jet energy resolutions (JER) in data are about 10% worse than in the simulation [96]. The uncertainty on this measurement is of the same order (10%). To account for this, all jets in the [Monte Carlo](#) simulated samples have to be scaled accordingly and the effect has to be propagated to \cancel{E}_T properly. The templates (Figure 6.7) used for the pseudo-experiments are produced as follows: For the selected jets matched to generator jets, the difference of the p_T of the reconstructed and generated jet is multiplied with a scale factor (f_{JER}) and then divided by the reconstructed jet p_T ; the resulting factor is propagated to rescale the jet four-momenta. The f_{JER} used for the nominal templates for the measurement in data is 0.1. The impact of this uncertainty on our measurement is determined by evaluating the change in cross section when jets are rescaled by 0% or 20%, respectively, to account for a shift of $\pm 10\%$. The overall uncertainties caused by the JER are $+6.9\%$ and -6.5% . Figure 6.8 shows the selected pseudo-experiments and the corresponding fit results.

6.2.3 Factorization Scale

To estimate the uncertainty due to the factorization scale, the pseudo-data is drawn from the templates (Figures 6.9 and 6.10) obtained from the samples where the scale μ_F has been modified by a factor of 0.5 and 2.0 with respect to the nominal value. For the systematic effect due to this variation, we treat the $W + \text{jets}$ and $Z + \text{jets}$ processes as correlated, and thus we simultaneously shift the factorization scale up or down for both processes, while $t\bar{t}$ is considered uncorrelated and is shifted independently. This results in $+4.9\%$ and -8.3% for varying the $t\bar{t}$ sample, and $+8.9\%$ and -10.3% for varying the

$W(Z)+\text{jets}$ samples. Figures 6.11 and 6.12 show the selected pseudo-experiments and the corresponding fit results for $t\bar{t}$ and $W(Z)+\text{jets}$, respectively.

6.2.4 Matching Threshold

We have also investigated the impact of varying the **ME-PS** jet matching threshold for $t\bar{t}$, $W+\text{jets}$, and $Z+\text{jets}$ events. The threshold used for generating the $t\bar{t}$ MC sample has been changed to 10 and to 40 GeV/c (the nominal threshold of 20 GeV/c .) In the case of the $W(Z)+\text{jets}$ samples, the threshold has been changed to 5 and 20 GeV/c compared to the nominal value of 10 GeV/c . Again $W+\text{jets}$ and $Z+\text{jets}$ are treated as fully correlated and are varied simultaneously while $t\bar{t}$ is considered uncorrelated and therefore varied independently. This results in systematic errors of +4.6% and -8.3% for varying the $t\bar{t}$ sample, and +10.0% and -7.8% for varying the $W(Z)+\text{jets}$ samples. The templates used for generating pseudo-experiments are shown in Figures 6.13 and 6.14. Figures 6.15 and 6.16 show selected pseudo-experiments and the corresponding fit results for $t\bar{t}$ and $W(Z)+\text{jets}$, respectively.

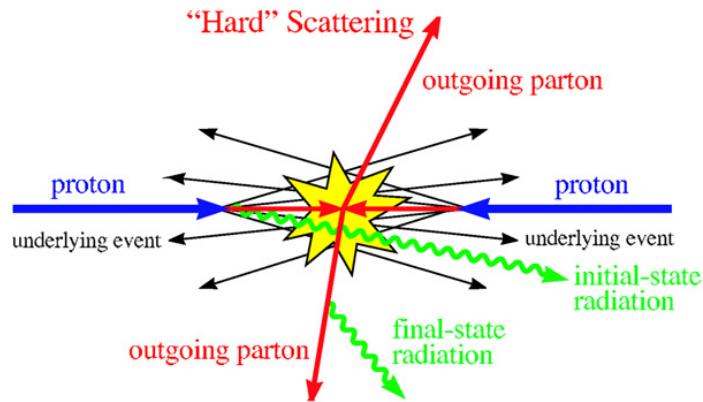


Figure 6.4: Schematic cartoon of $2 \rightarrow 2$ hard-scattering event.

6.2.5 Initial and Final State Radiation

As shown in Figure 6.4, a hard-scattering at a hadron collider can produce possible gluon radiation from both the incoming and outgoing partons [97], called initial-state (ISR) and final-state radiation (FSR), respectively. The impact on the cross section measurement due to a larger or smaller amount of ISR and FSR are estimated to be +3.5% and -8.1%. Figures 6.17 and 6.18 show the interpolated templates used to generate pseudo-data and the results of the simultaneous fit with the nominal templates, respectively.

6.2.6 Underlying Event Modeling

As shown in Figure 6.4, along with the hard-scattering there are softer interactions from the remaining partons in the colliding hadrons at a hadron collider, called underlying events. An uncertainty due to different underlying event (UE) tunes, D6T [66] and Z2 [98], has been investigated. The default underlying event modeling uses the D6T tune. We studied the effect on our measurement by using the Z2-tune Monte Carlo $t\bar{t}$, W +jets, and Z +jets samples. The UE uncertainty is estimated to be less than 1%.

6.2.7 Parton Distribution Function

One of the major uncertainties is the limited knowledge of the parton distribution functions (PDFs) of the colliding protons. This systematic uncertainty is evaluated following the re-weighting method described in [99], using the alternative PDF set, CTEQ6.6 [100], via the LHAPDF (Les Houches Accord Parton Distribution Function) library [101]. The event weight is applied to the $t\bar{t}$ and W +jets simulated samples using positive and negative variations for each of the 22 eigenvectors of the CTEQ6.6 PDF, resulting in a total of 44

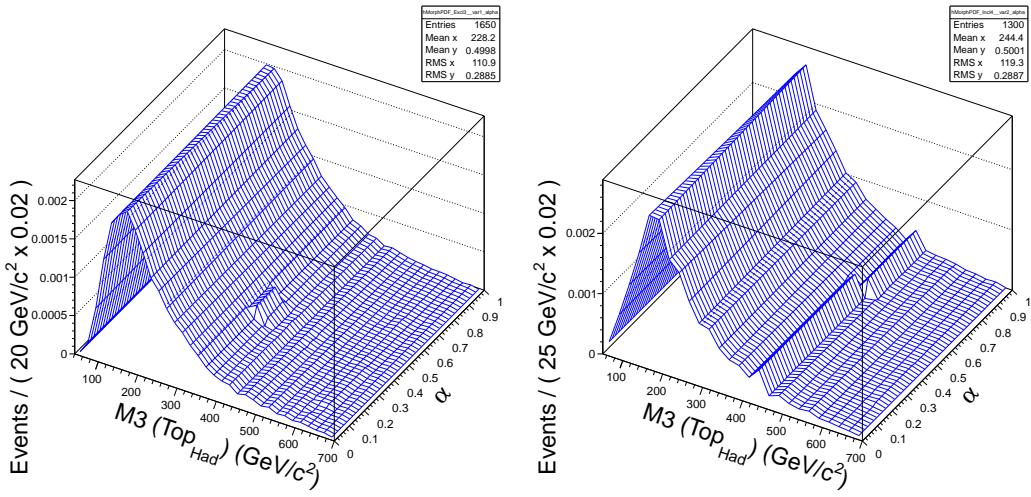
different templates for each sample, which are then used to fit the data directly to estimate the impact on our measurement. The resulting uncertainties are slightly less than 3%.

6.2.8 Summary of the Systematic Uncertainties

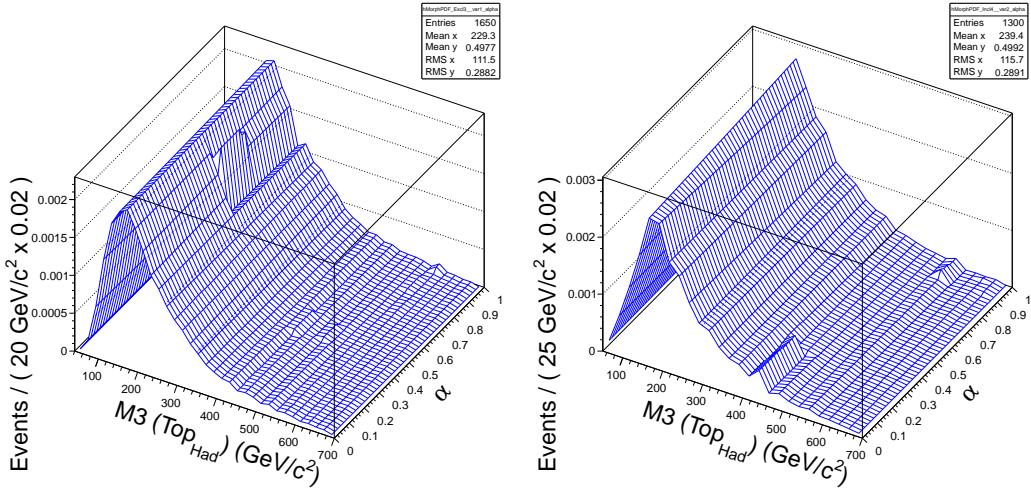
We have investigated the impact of various sources of systematic uncertainties on the $t\bar{t}$ cross section measurement. The total systematic uncertainties are calculated to be $+17.7\%$ and -22.3% by summing all of the uncertainties in quadrature. The results of the systematic uncertainty determinations are summarized in Table 6.2. Taking into account all the systematic uncertainties, the measured cross section is $157 \pm 12(\text{stat.})^{+28}_{-35}(\text{syst.}) \pm 6(\text{lumi.}) \text{ pb}$.

Table 6.2: Summary of the systematic uncertainties in percentages of the top pair production cross section.

Type	Uncertainty in %	
JES	-8.7	+4.5
JER	-6.5	+6.9
Factorization Scale ($t\bar{t}$)	-8.3	+4.9
Factorization Scale (W/Z)	-10.3	+8.9
Matching Threshold ($t\bar{t}$)	-8.3	+4.6
Matching Threshold (W/Z)	-7.8	+10.0
ISR/FSR	-8.1	+3.5
PDF	± 3	
Underlying Event	± 0.5	
Total	-22.3	+17.7

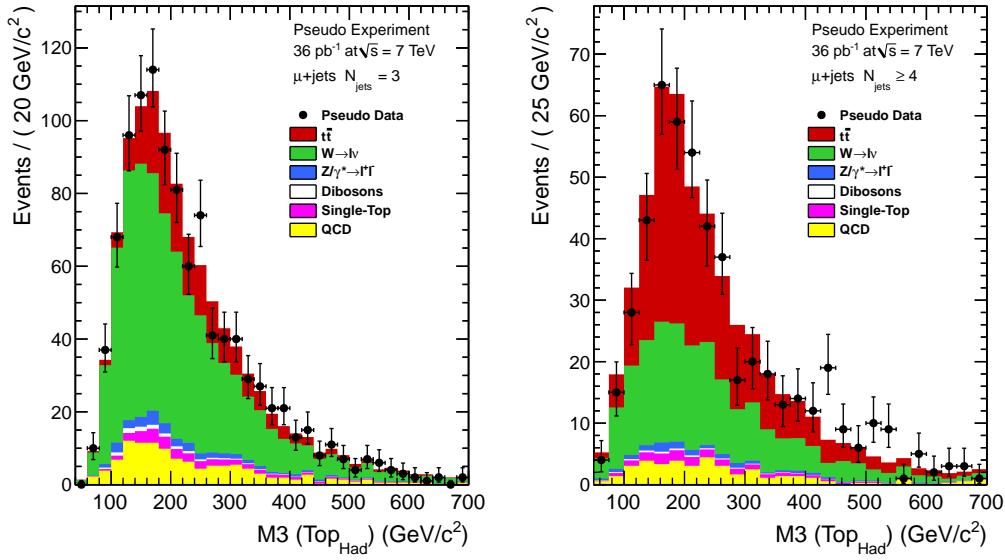


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

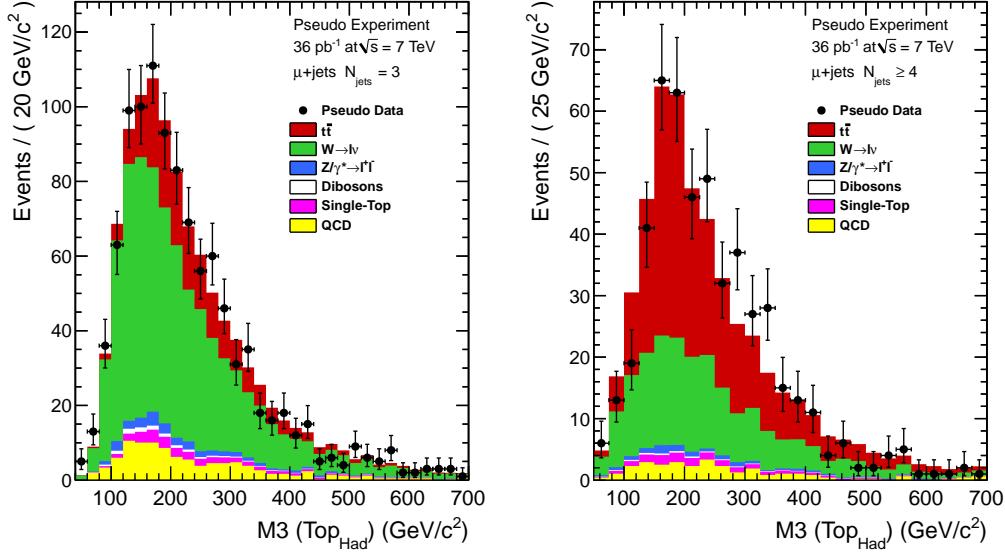


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.5: The interpolated templates for the M_3 distribution as a function of α for JES systematic error determination: (a) JES scaled up and (b) JES scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.

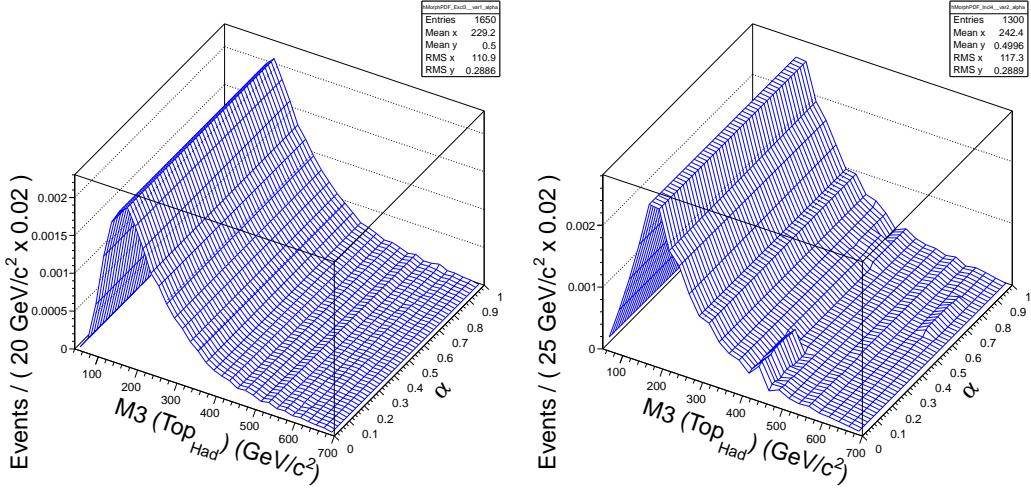


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

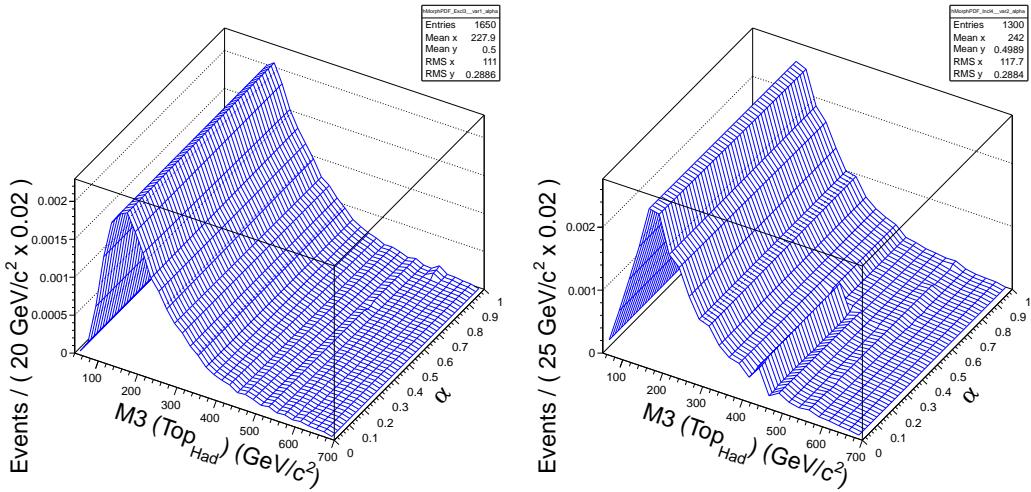


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.6: Selected MC pseudo-experiments and the fit results for the JES systematic error determination: (a) JES scaled up and (b) JES scaled down.

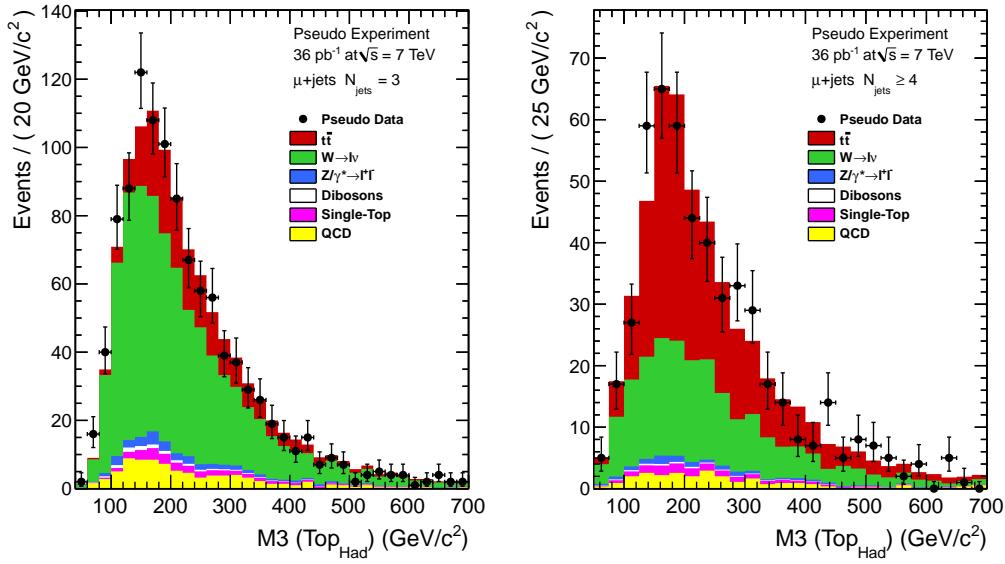


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

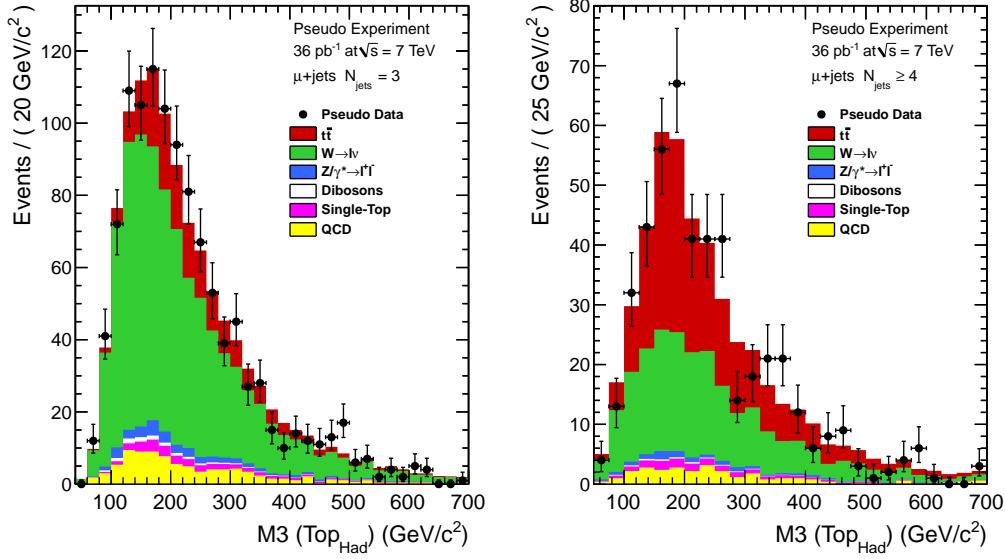


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.7: The interpolated templates for the M_3 distribution as a function of α for JER systematic error determination: (a) JER scaled up and (b) JER scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.



(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.



(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.8: Selected MC pseudo-experiments and the fit results for the JER systematic error determination: (a) JER scaled up (+20%) and (b) JER scaled down (+0%).

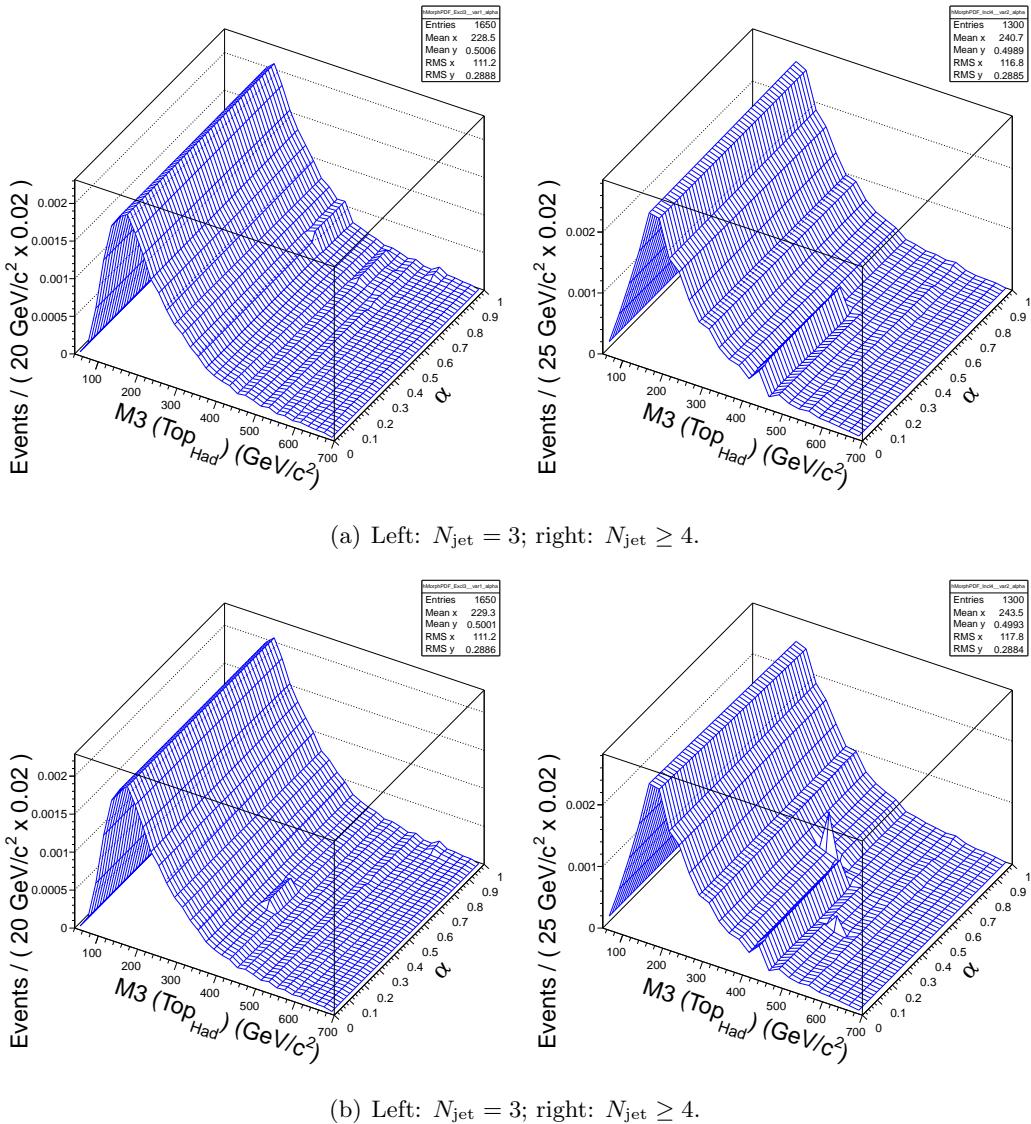
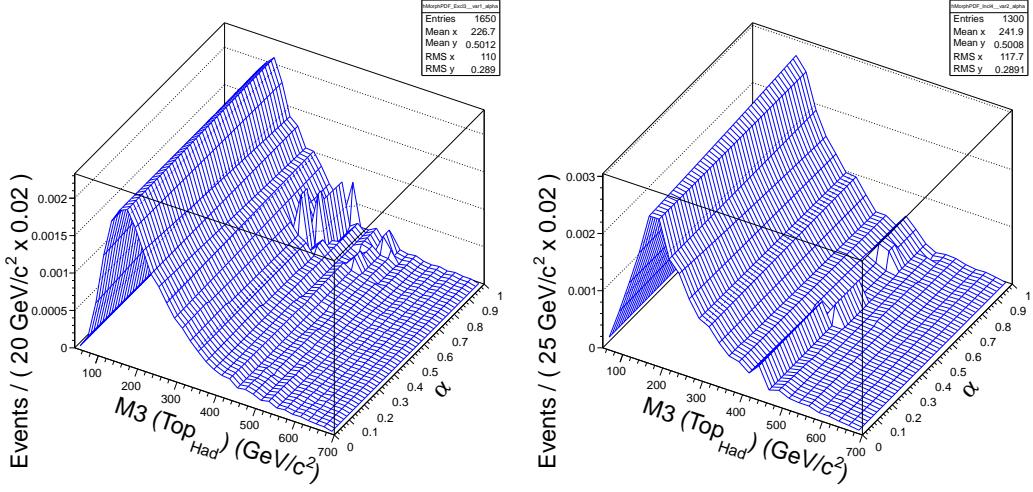
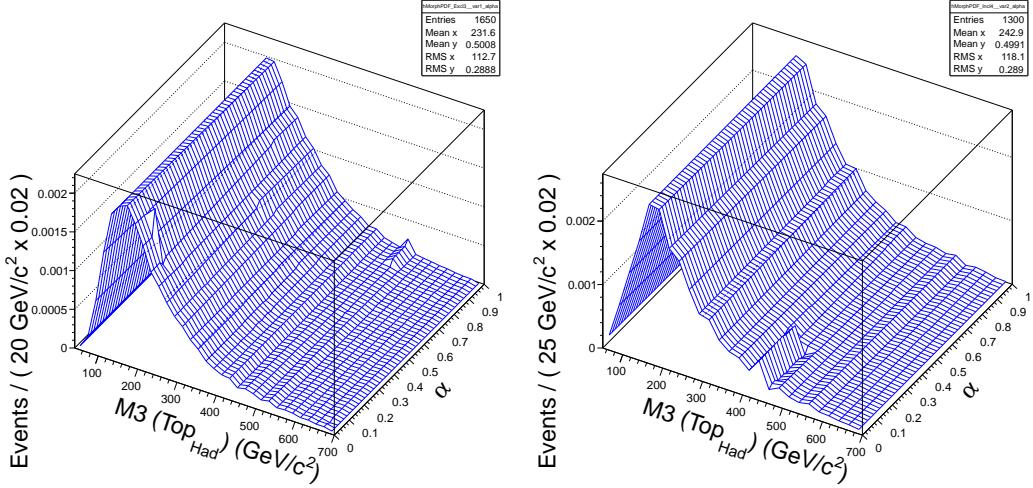


Figure 6.9: The interpolated templates for the M3 distribution as a function of α for the $t\bar{t}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.

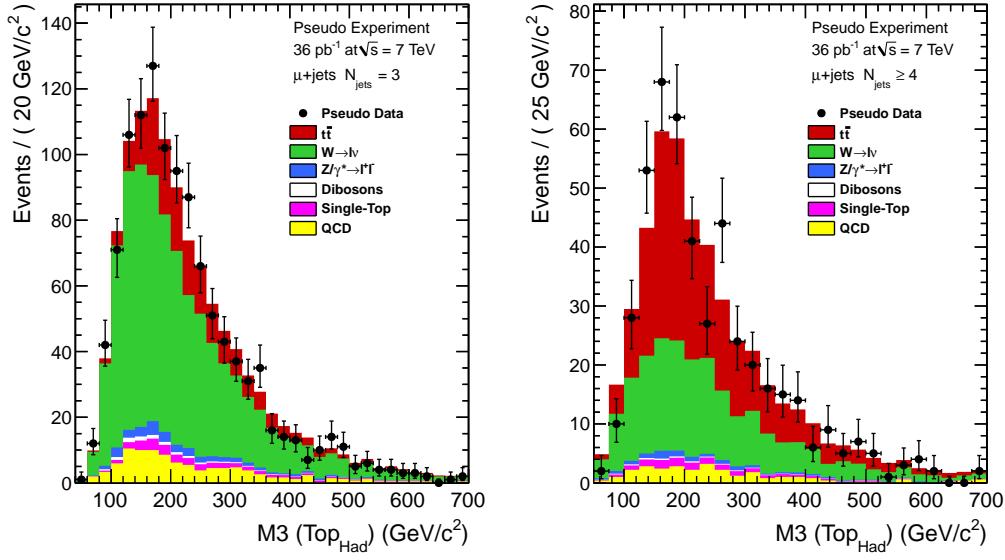


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

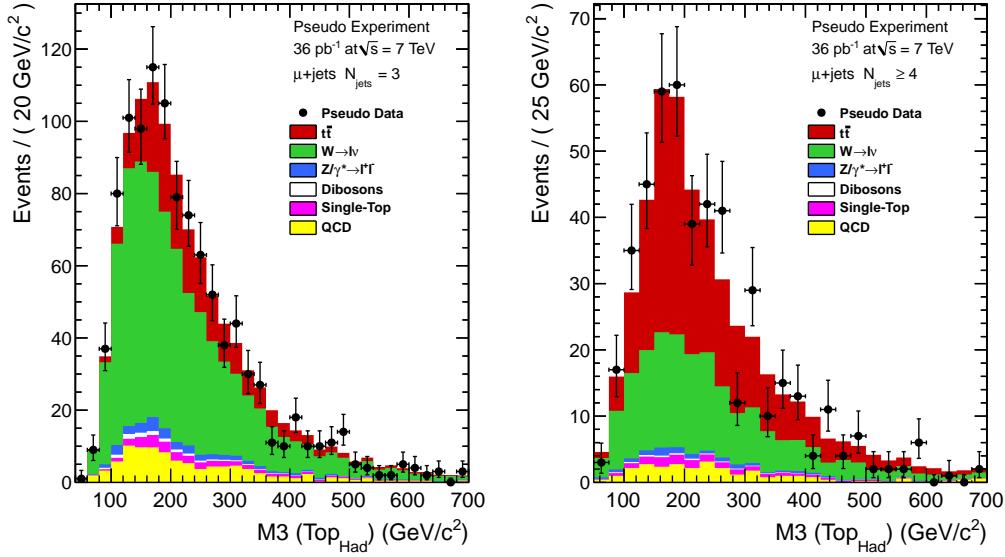


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.10: The interpolated templates for the M_3 distribution as a function of α for the $W(Z)+\text{jets}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.

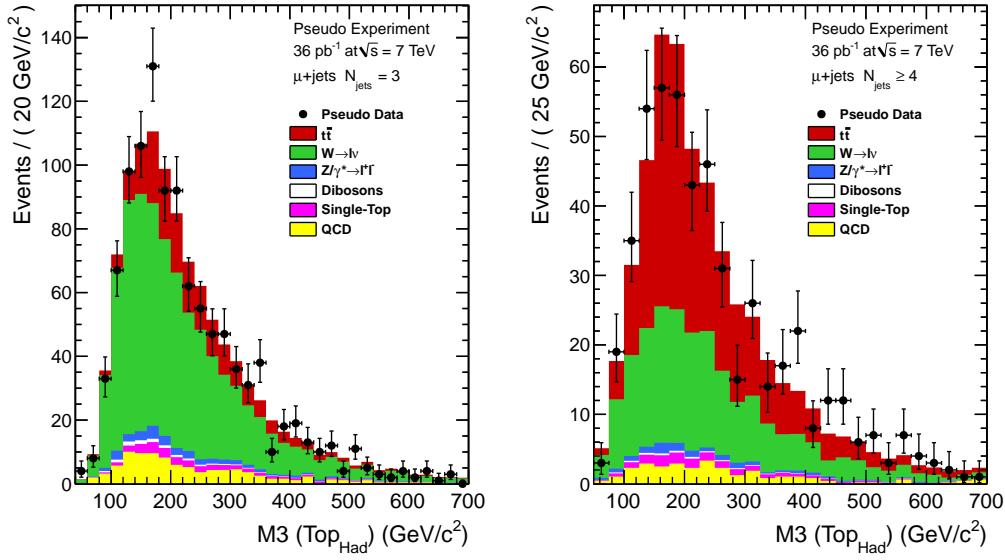


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

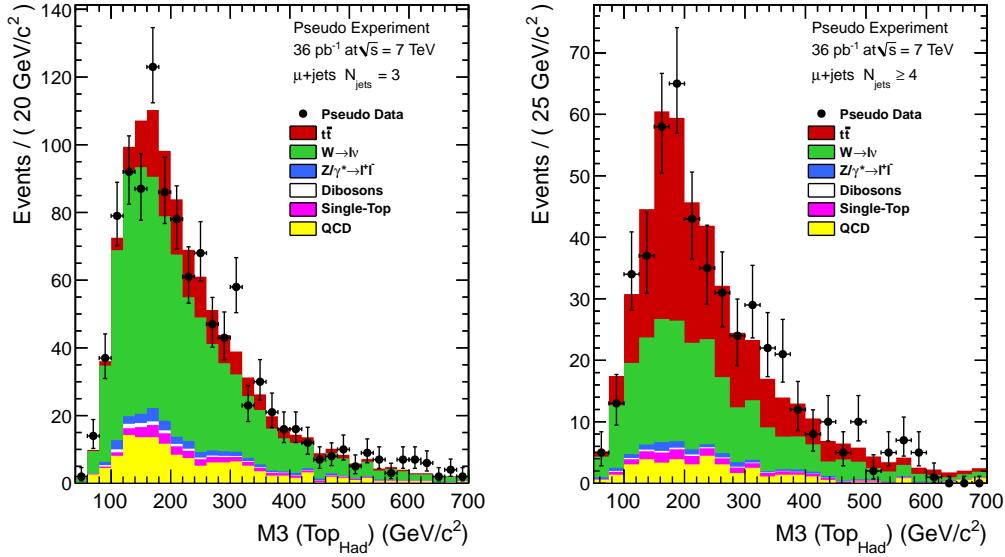


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.11: Selected MC pseudo-experiments and the fit results for the $t\bar{t}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down.

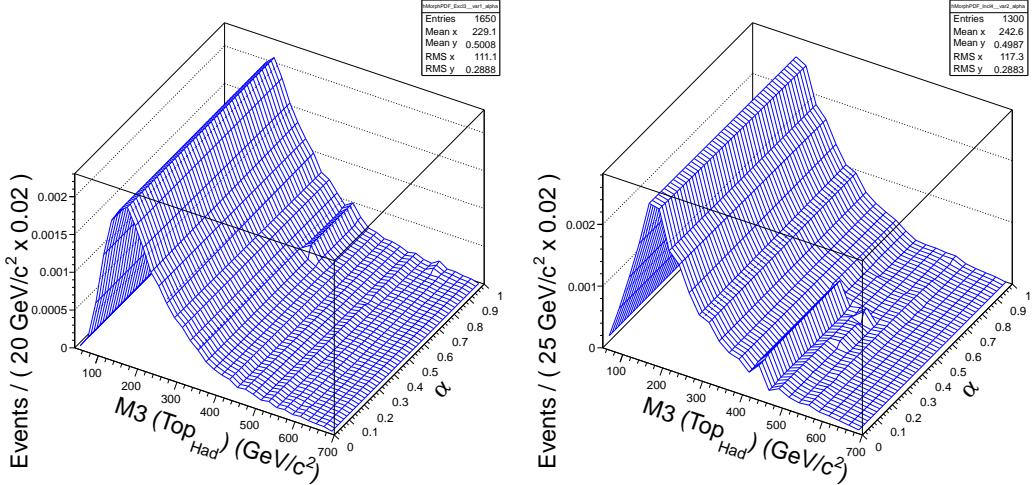


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

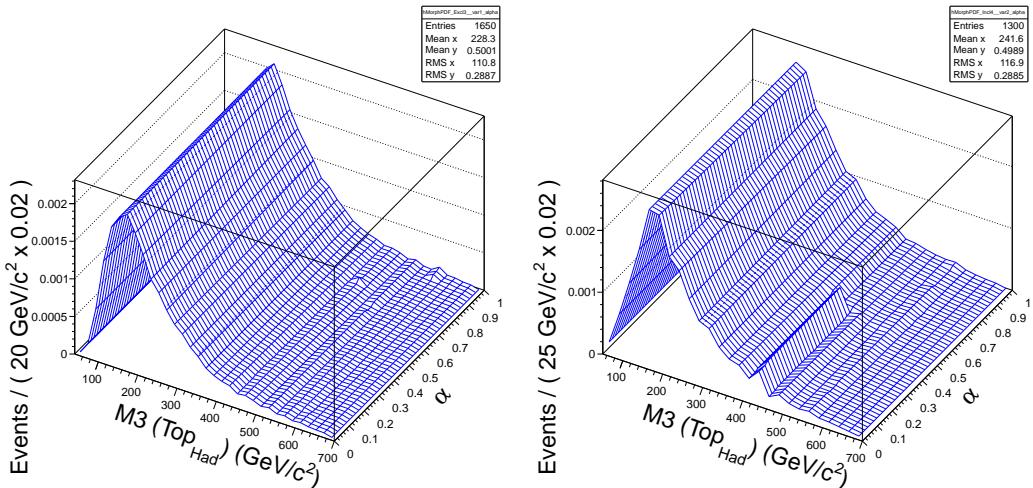


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.12: Selected MC pseudo-experiments and the fit results for the $W(Z)+\text{jets}$ factorization (μ_F) systematic error determination: (a) μ_F scaled up and (b) μ_F scaled down.

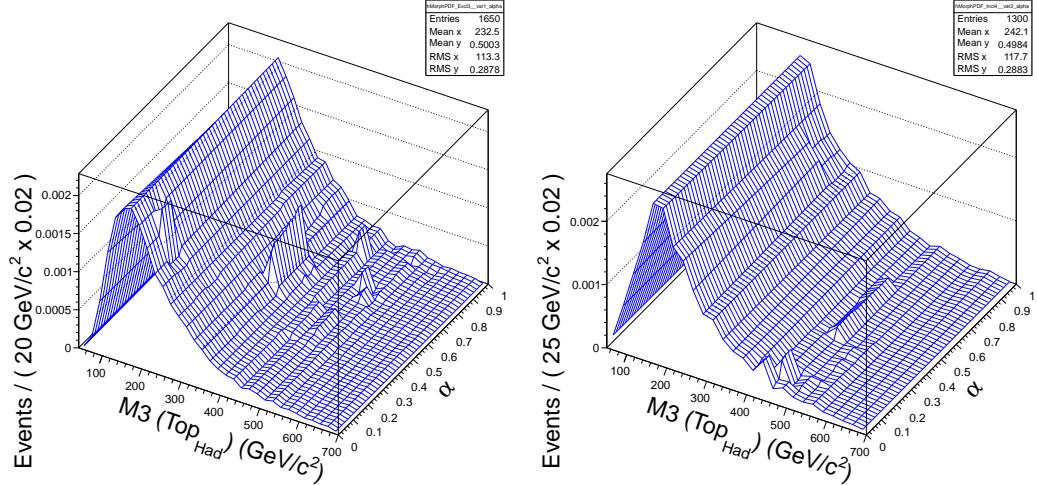


(a) Left: $N_{jet} = 3$; right: $N_{jet} \geq 4$.

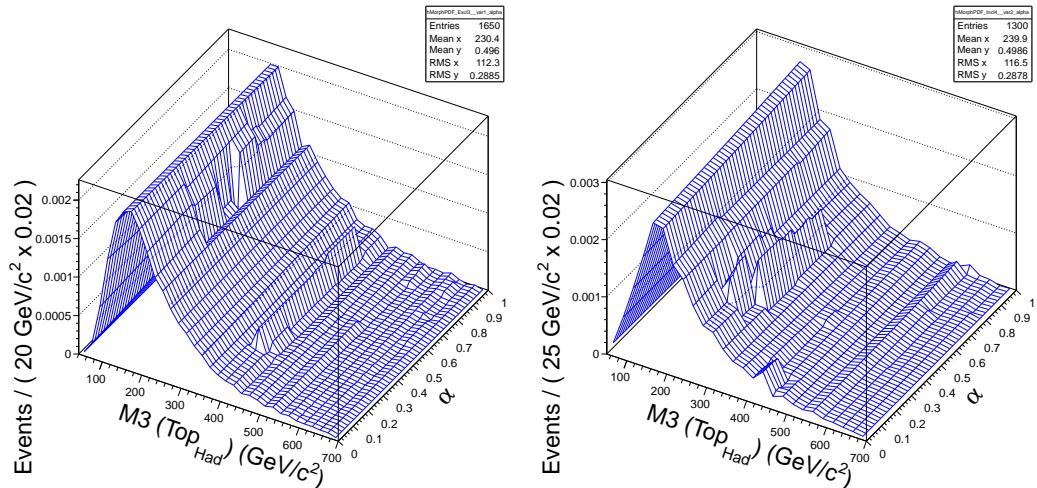


(b) Left: $N_{jet} = 3$; right: $N_{jet} \geq 4$.

Figure 6.13: The interpolated templates for the M_3 distribution as a function of α for the $t\bar{t}$ ME-PS jet matching threshold systematic error determination: (a) 40 GeV/c and (b) 10 GeV/c. The nominal threshold is 20 GeV/c. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.

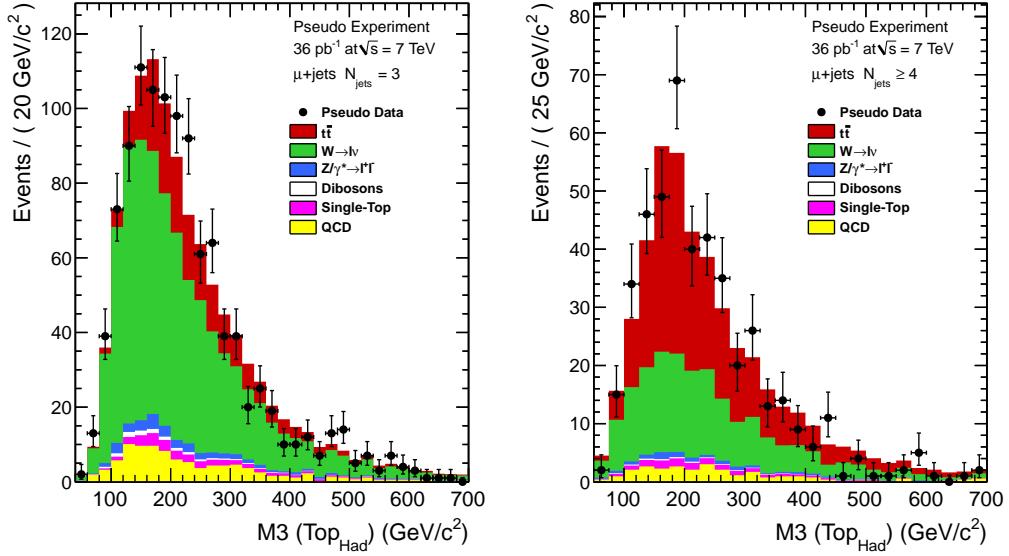


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

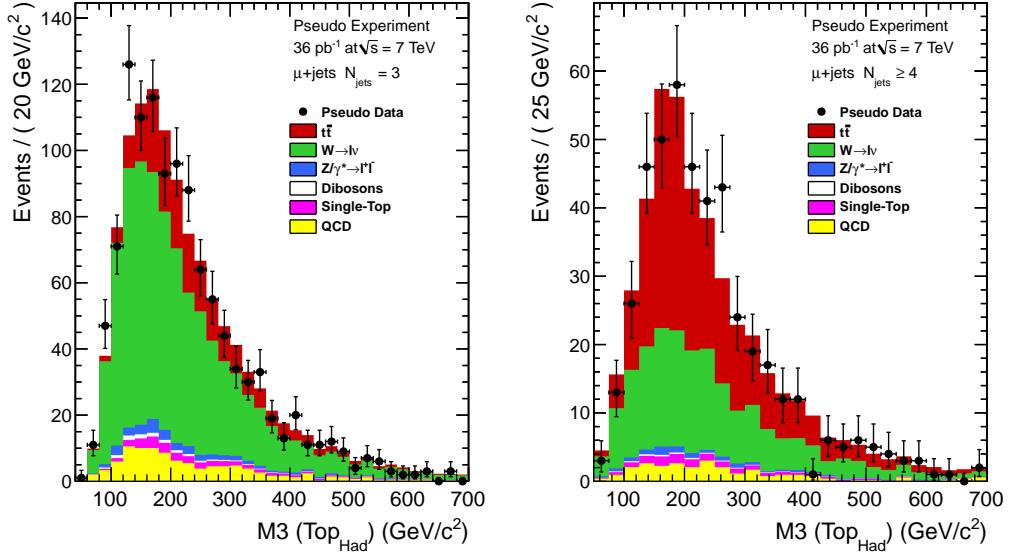


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.14: The interpolated templates for the M_3 distribution as a function of α for the $W(Z)+\text{jets}$ ME-PS jet matching threshold systematic error determination: (a) 20 GeV/c^2 and (b) 25 GeV/c^2 . The nominal threshold is 10 GeV/c^2 . $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.

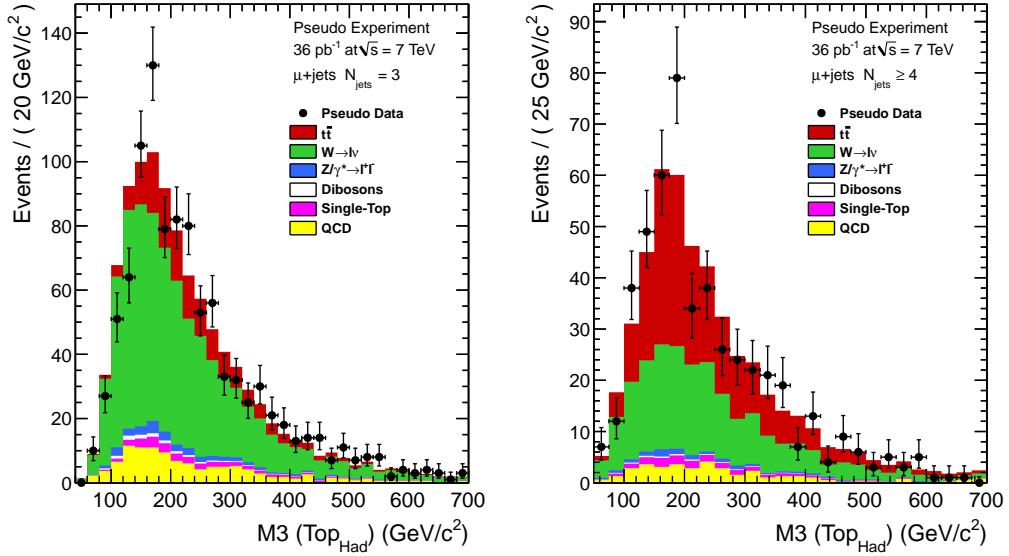


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

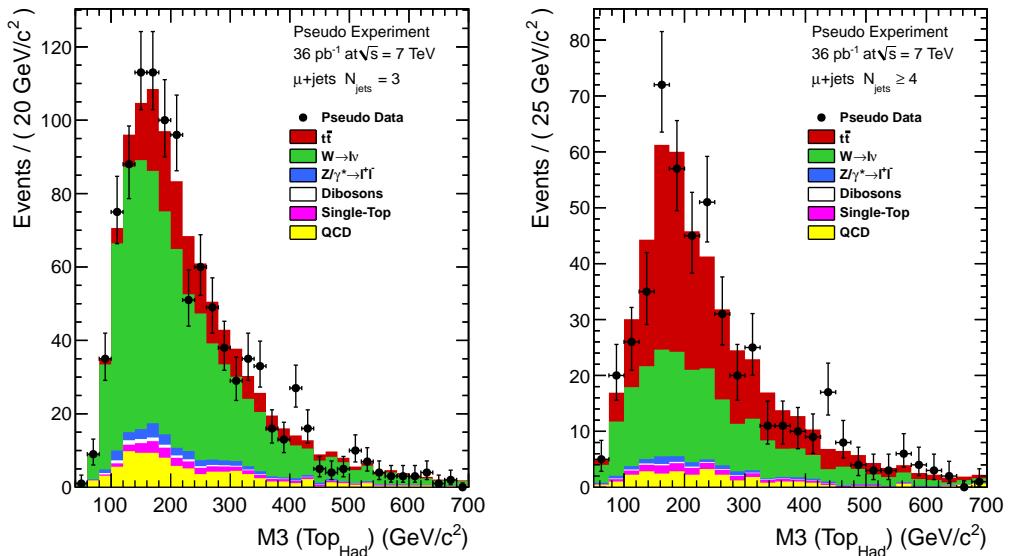


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.15: Selected MC pseudo-experiments and the fit results for the $t\bar{t}$ ME-PS jet matching threshold systematic error determination: (a) 40 GeV/ c and (b) 10 GeV/ c . The nominal threshold is 20 GeV/ c .

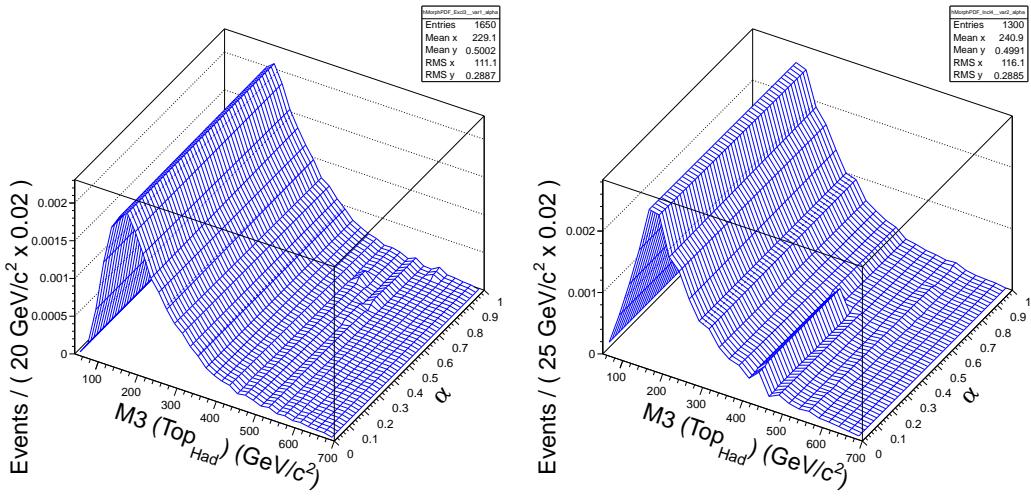


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

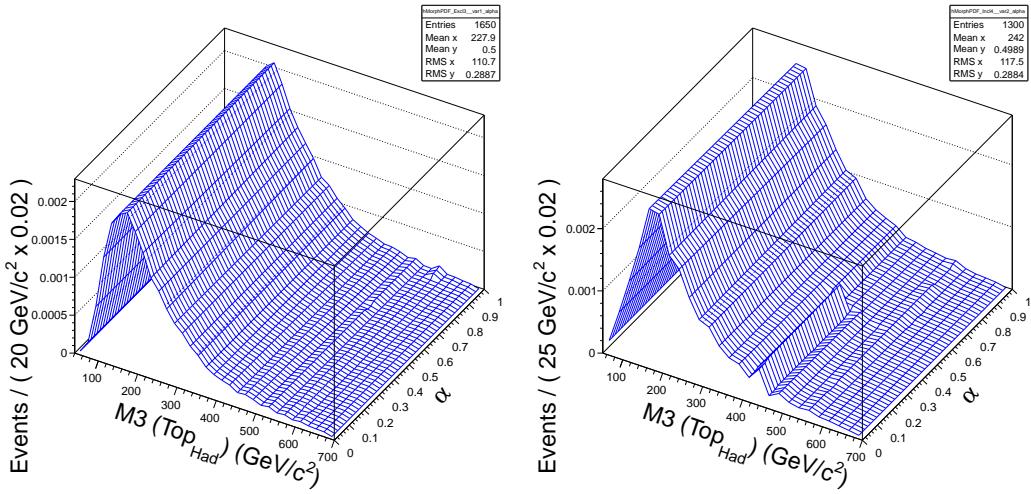


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.16: Selected MC pseudo-experiments and the fit results for the $W(Z)+\text{jets}$ ME-PS jet matching threshold systematic error determination: (a) 20 GeV/c and (b) 5 GeV/c. The nominal threshold is 10 GeV/c.

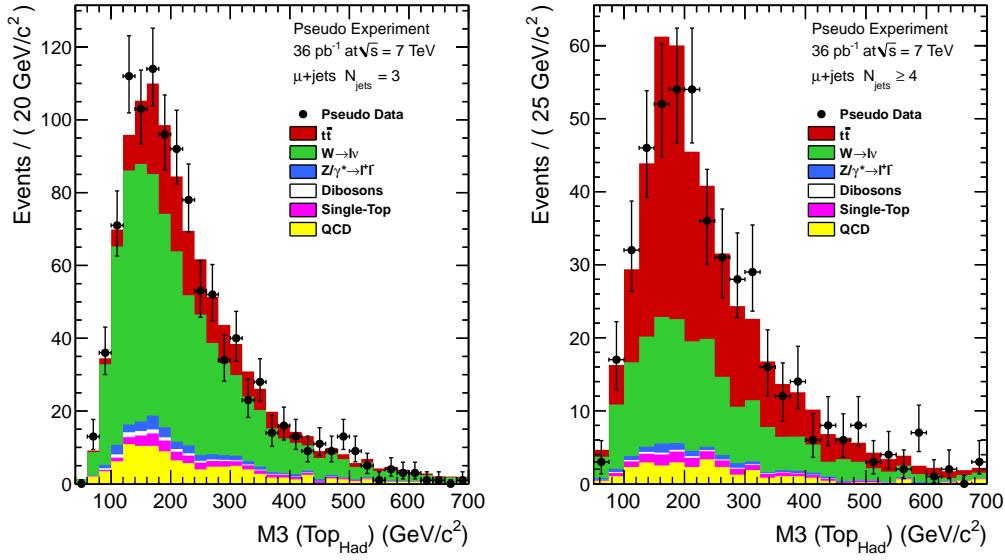


(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

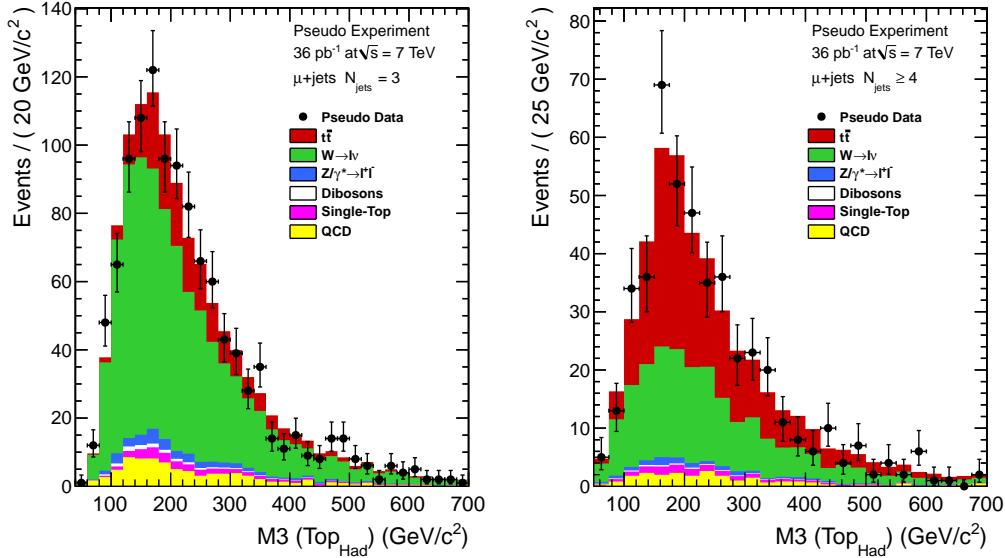


(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.17: The interpolated templates for the M_3 distribution as a function of α for ISR/FSR systematic error determination: (a) larger ISR/FSR and (b) smaller ISR/FSR. $\alpha = 1$ corresponds to the nominal template, and $\alpha = 0$ the systematic template.



(a) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.



(b) Left: $N_{\text{jet}} = 3$; right: $N_{\text{jet}} \geq 4$.

Figure 6.18: Selected MC pseudo-experiments and the fit results for the ISR/FSR systematic error determination: (a) larger ISR/FSR and (b) smaller ISR/FSR.

Chapter 7

Conclusions

We perform a measurement of the top quark pair-production cross section in pp collisions at a center-of-mass energy of 7 TeV at the Large Hadron Collider. The analysis is based on a 36 pb^{-1} dataset collected with the CMS experiment in 2010. We study the kinematic characteristics of events with one isolated muon, hadronic jets, and substantial missing transverse energy to distinguish the $t\bar{t}$ signal from the electroweak and QCD multijet backgrounds. A dedicated data-driven method to estimate the QCD multijet background is also used. The $t\bar{t}$ signal is extracted by simultaneously fitting the tri-jet invariant mass distributions in events with exactly three jets and at least four jets. The invariant mass distributions of the combinations of the three jets with the largest magnitude of the vectorial sum of the jet \vec{p}_T are fitted to obtain the $t\bar{t}$ signal and the contributions from backgrounds. The measured cross section is $157 \pm 12(\text{stat.})^{+28}_{-35}(\text{syst.}) \pm 6(\text{lumi.}) \text{ pb}$, which is in good agreement with the theoretical value of $157.5^{+23.2}_{-24.4} \text{ pb}$ [79–81].

Bibliography

- [1] LEP Working Group for Higgs boson searches Collaboration, “Search for the standard model Higgs boson at LEP,” *Phys. Lett. B* **565** (2003) 61–75, [arXiv:hep-ex/0306033](https://arxiv.org/abs/hep-ex/0306033). [doi:10.1016/S0370-2693\(03\)00614-2](https://doi.org/10.1016/S0370-2693(03)00614-2).
- [2] CDF and DØ Collaboration, “Combined CDF and DØ Upper Limits on Standard Model Higgs- Boson Production with up to 6.7 fb^{-1} of Data,” [arXiv:1007.4587](https://arxiv.org/abs/1007.4587).
- [3] G. Sterman, “Summation of large corrections to short-distance hadronic cross sections,” *Nucl. Phys. B* **281** (1987).
- [4] R. Bonciani, S. Catani, M. Mangano *et al.*, “NLL resummation of the heavy-quark hadroproduction cross-section,” *Nucl. Phys. B* **529** (1998) 424–450, [arXiv:hep-ph/9801375v2](https://arxiv.org/abs/hep-ph/9801375v2). [doi:10.1016/S0550-3213\(98\)00335-6](https://doi.org/10.1016/S0550-3213(98)00335-6).
- [5] M. L. Mangano, “The $t\bar{t}$ cross-sections in hadronic collisions,” *Nuovo Cimento B Serie* **123** (2008) 1021–1027. [doi:10.1393/ncb/i2008-10659-6](https://doi.org/10.1393/ncb/i2008-10659-6).
- [6] S. L. Glashow, “Partial Symmetries of Weak Interactions,” *Nucl. Phys. B* **22** (1961) 579–588. [doi:10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [7] A. Salam and J. C. Ward, “Electromagnetic and weak interactions,” *Phys. Lett.* (1964).
- [8] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.* **19** (1967) 1264–1266. [doi:10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [9] S. Weinberg, “The Quantum Theory of Fields, Volume 1: Foundations”. Cambridge University Press, 2005.
- [10] A. Zee, “Quantum field theory in a nutshell”. Universities Press, 2003.
- [11] B. Roe, “Particle physics at the new millennium”. Springer-Verlag, 1996.
- [12] K. Nakamura *et al.* (Particle Data Group), “Review of Particle Physics,” *J. Phys. G* **37** (2010) 075021. [doi:10.1088/0954-3899/37/7A/075021](https://doi.org/10.1088/0954-3899/37/7A/075021).

- [13] CDF Collaboration, “Observation of top quark production in $\bar{p}p$ collisions,” *Phys. Rev. Lett.* **74** (1995) 2626, [arXiv:hep-ex/9503002](https://arxiv.org/abs/hep-ex/9503002).
[doi:10.1103/PhysRevLett.74.2626](https://doi.org/10.1103/PhysRevLett.74.2626).
- [14] DØ Collaboration, “Observation of the top quark,” *Phys. Rev. Lett.* **74** (1995) 2632, [arXiv:hep-ex/9503003](https://arxiv.org/abs/hep-ex/9503003). [doi:10.1103/PhysRevLett.74.2632](https://doi.org/10.1103/PhysRevLett.74.2632).
- [15] CDF and DØ Collaboration, “Combination of CDF and DØ Results on the Mass of the Top Quark,” [arXiv:1007.3178v1](https://arxiv.org/abs/1007.3178v1).
- [16] J. H. Kuhn, “Theory of top quark production and decay,” [arXiv:hep-ph/9707321](https://arxiv.org/abs/hep-ph/9707321).
- [17] M. Jeabek and J. H. Khn, “Semileptonic decays of top quarks,” *Phys. Lett. B* **207** (1988) 91 – 96. [doi:10.1016/0370-2693\(88\)90892-1](https://doi.org/10.1016/0370-2693(88)90892-1).
- [18] M. Jezabek and J. H. Khn, “QCD corrections to semileptonic decays of heavy quarks,” *Nucl. Phys. B* **314** (1989) 1 – 6. [doi:10.1016/0550-3213\(89\)90108-9](https://doi.org/10.1016/0550-3213(89)90108-9).
- [19] J. R. Incandela *et al.*, “Status and Prospects of Top-Quark Physics,” *Prog. Part. Nucl. Phys.* **63** (2009) 239, [arXiv:0904.2499](https://arxiv.org/abs/0904.2499).
[doi:10.1016/j.ppnp.2009.08.001](https://doi.org/10.1016/j.ppnp.2009.08.001).
- [20] G. Hanson *et al.*, “Evidence for Jet Structure in Hadron Production by e^+e^- Annihilation,” *Phys. Rev. Lett.* **35** (1975) 1609–1612.
[doi:10.1103/PhysRevLett.35.1609](https://doi.org/10.1103/PhysRevLett.35.1609).
- [21] CMS Collaboration, “CMS physics Technical Design Report, Volume II: Physics Performance”. *J. Phys. G* **34** (2006) 995, CERN-LHCC-2006-021,
<http://cdsweb.cern.ch/record/942733>.
- [22] V. D. Barger and R. J. N. Phillips, “Collider Physics (Frontiers in Physics)”. Addison Wesley Publishing Company, 1987.
- [23] L. Evans and P. Bryant, “LHC Machine,” *JINST* **3** (2008) S08001.
[doi:10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [24] O. S. Brüning, P. Collier, P. Lebrun *et al.*, “LHC Design Report”. CERN-2004-003-V-1, <http://cdsweb.cern.ch/record/782076>.
- [25] ATLAS Collaboration, “The ATLAS experiment at the CERN LHC,” *JINST* **3** (2008) S08003. [doi:10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [26] CMS Collaboration, “The CMS experiment at the CERN LHC,” *JINST* **3** (2008) S08004. [doi:10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [27] C. Carli (ed.), “Proceedings of the 2010 Charmonix Workshop on LHC Performance,” 2010. [CERN-ATS-2010-026](https://cdsweb.cern.ch/record/1250026).
- [28] B. Goddard (ed.), “Proceedings of the 2010 Evian Workshop on LHC Commissioning,” 2010. [CERN-ATS-2010-028](https://cdsweb.cern.ch/record/1250028).

- [29] CMS Collaboration, “CMS Luminosity-Public Results”.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults2010>.
- [30] CMS Collaboration, “The CMS tracker system project: Technical Design Report,” 1997. [CERN-LHCC-98-006](#).
- [31] CMS Collaboration, “The CMS tracker: addendum to the Technical Design Report,” 2000. [CERN-LHCC-2000-016](#).
- [32] CMS Collaboration, “CMS: The Electromagnetic Calorimeter Technical Design Report”. CERN-LHCC-97-033, <http://cdsweb.cern.ch/record/349375>.
- [33] CMS Collaboration, “CMS: The Hadron Calorimeter Technical Design Report”. CERN-LHCC-97-031, <http://cdsweb.cern.ch/record/357153>.
- [34] R. Loos, P. Wertelaers, D. Barney *et al.*, “CMS ECAL Preshower and Endcap Engineering Design Review. v.2 - Preshower”. CERN-ECAL-EDR-4, <http://cdsweb.cern.ch/record/539819>.
- [35] A. A. Annenkov, M. V. Korzhik, and P. Lecoq, “Lead tungstate scintillation material,” *Nucl. Instrum. and Meth. A* **490** (2002), no. 1-2, 30 – 50.
[doi:10.1016/S0168-9002\(02\)00916-6](https://doi.org/10.1016/S0168-9002(02)00916-6).
- [36] X. D. Qu, L. Y. Zhang, and R. Y. Zhu, “Radiation induced color centers and light monitoring for lead tungstate crystals,” *IEEE Trans. Nucl. Sci.* **47** (2000) 1741–1747. [doi:10.1109/TNS.2000.914439](https://doi.org/10.1109/TNS.2000.914439).
- [37] P. Adzic *et al.*, “Energy resolution of the barrel of the CMS Electromagnetic Calorimeter,” *JINST* **2** (2007), no. 04, P04004.
[doi:10.1088/1748-0221/2/04/P04004](https://doi.org/10.1088/1748-0221/2/04/P04004).
- [38] CMS Collaboration, “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data,” *JINST* **5** (2010) T03012, [arXiv:0911.4991](https://arxiv.org/abs/0911.4991).
[doi:10.1088/1748-0221/5/03/T03012](https://doi.org/10.1088/1748-0221/5/03/T03012).
- [39] CMS Collaboration, “The CMS muon project: Technical Design Report”. CERN-LHCC-97-032, <http://cdsweb.cern.ch/record/343814>.
- [40] R. Santonico and R. Cardarelli, “Development of resistive plate counters,” *Nucl. Instrum. and Meth.* **187** (1981) 377. [doi:10.1016/0029-554X\(81\)90363-3](https://doi.org/10.1016/0029-554X(81)90363-3).
- [41] R. Cardarelli, A. D. Ciaccio, and R. Santonico, “Performance of a resistive plate chamber operating with pure CF3Br,” *Nucl. Instrum. and Meth. A* **333** (1993) 399. [doi:10.1016/0168-9002\(93\)91182-M](https://doi.org/10.1016/0168-9002(93)91182-M).
- [42] M. Abbrescia, G. Bruno, A. Colaleo *et al.*, “Beam test results on double-gap resistive plate chambers proposed for CMS experiment,” *Nucl. Instrum. and Meth. A* **414** (1998) 135. [doi:10.1016/S0168-9002\(98\)00571-3](https://doi.org/10.1016/S0168-9002(98)00571-3).

- [43] CMS Collaboration, “The TriDAS project. Technical Design Report, Volume 1: the trigger systems”. CERN-LHCC-2000-038, <http://cdsweb.cern.ch/record/706847>.
- [44] CMS Collaboration, “The TriDAS project. Technical design report, Volume 2: Data acquisition and high-level trigger technical design report”. CERN-LHCC-2002-026, <http://cdsweb.cern.ch/record/578006>.
- [45] CMS Collaboration, “Measurement of CMS Luminosity,” *CMS PAS EWK-10-004* (2010).
- [46] CMS Collaboration, “Absolute Calibration of Luminosity Measurement at CMS,” *CMS AN AN-11-093* (2011).
- [47] CMS Collaboration, “CMS Physics Technical Design Report Volume I: Detector Performance and Software”. CERN-LHCC-2006-001, <http://cdsweb.cern.ch/record/922757>.
- [48] T. Miao, H. Wenzel, and F. Yumiceva, “Beam Position Determination Using Tracks,” *CMS Note 2007/021* (2007).
- [49] W. Adam, B. Mangano, T. Speer *et al.*, “Track Reconstruction in the CMS tracker,” Technical Report CMS-NOTE-2006-041. CERN-CMS-NOTE-2006-041, CERN, Geneva, Dec, 2006. http://cdsweb.cern.ch/record/934070/files/NOTE2006_041.pdf.
- [50] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Basic Eng.* **82** (1960) 35–45.
- [51] CMS Collaboration, “CMS tracking performance results from early LHC operation,” *Eur. Phys. J. C* **70** (2010) 1165–1192. [doi:10.1140/epjc/s10052-010-1491-3](https://doi.org/10.1140/epjc/s10052-010-1491-3).
- [52] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting,” *CMS AN 2006/104* (2006).
- [53] CMS Collaboration, “Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV,” *CMS Physics Analysis Summary MUO-10-002* (2010). <http://cms-physics.web.cern.ch/cms-physics/public/MUO-10-002-pas.pdf>.
- [54] S. Baffioni, C. Charlot, F. Ferri *et al.*, “Electron Reconstruction in CMS,” Technical Report CMS-NOTE-2006-040. CERN-CMS-NOTE-2006-040, CERN, Geneva, Feb, 2006. http://cdsweb.cern.ch/record/934070/files/NOTE2006_040.pdf.
- [55] CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7$ TeV,” *CMS PAS EGM-10-004* (2010). <http://cms-physics.web.cern.ch/cms-physics/public/EGM-10-004-pas.pdf>.
- [56] R. Frühwirth, “Track fitting with non-Gaussian noise,” *Comput. Phys. Commun.* **100** (1997) 1 – 16. [doi:10.1016/S0010-4655\(96\)00155-5](https://doi.org/10.1016/S0010-4655(96)00155-5).

- [57] W. Adam, R. Frühwirth, A. Strandlie, T. Todorov, “Reconstructions of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC,” *J. Phys. G: Nucl. Part. Phys.* **31** (2005) N9. [doi:10.1088/0954-3899/31/9/N01](https://doi.org/10.1088/0954-3899/31/9/N01).
- [58] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm,” *JHEP* **04** (2008) 063. [doi:10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [59] CMS Collaboration, “Jet Performance in pp Collisions at $\sqrt{s} = 7$ TeV,” *CMS Physics Analysis Summary JME-10-003* (2010).
<http://cms-physics.web.cern.ch/cms-physics/public/JME-10-003-pas.pdf>.
- [60] CMS Collaboration Collaboration, “Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS,” *CMS PAS JME-09-002* (2009).
<http://cms-physics.web.cern.ch/cms-physics/public/JME-09-002-pas.pdf>.
- [61] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS,” *CMS Physics Analysis Summary PFT-09-001* (2009).
<http://cms-physics.web.cern.ch/cms-physics/public/PFT-09-001-pas.pdf>.
- [62] CMS Collaboration, “Missing Transverse Energy Performance in Minimum-Bias and Jet Events from pp Collisions at $\sqrt{s} = 7$ TeV,” *CMS Physics Analysis Summary JME-10-004* (2010).
<http://cms-physics.web.cern.ch/cms-physics/public/JME-10-004-pas.pdf>.
- [63] CMS Collaboration, “Missing E_T Performance in CMS,” *CMS Physics Analysis Summary JME-07-001* (2007).
<http://cms-physics.web.cern.ch/cms-physics/public/JME-07-001-pas.pdf>.
- [64] CMS Collaboration, “Track-corrected Missing Transverse Energy in CMS,” *CMS Physics Analysis Summary JME-09-010* (2009).
<http://cms-physics.web.cern.ch/cms-physics/public/JME-09-010-pas.pdf>.
- [65] G. Altarelli and G. Parisi, *Nucl. Phys.* **B126** (1977) 298; Yu. L. Dokshitzer, Sov. Phys. JETP **46** (1977) 641; L. N. Lipatov, *Yad. Fiz.* **20** (1974) 181; V. Gribov and L. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 675, 938.
- [66] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual,” *JHEP* **05** (2006) 026, [arXiv:hep-ph/0603175v2](https://arxiv.org/abs/hep-ph/0603175v2).
[doi:10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026).
- [67] B. Andersson *et al.*, “Parton fragmentation and string dynamics,” *Phys. Reports* **97** (1983) 31 – 145. [doi:10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- [68] T. Sjöstrand, “Jet fragmentation of multiparton configurations in a string framework,” *Nucl. Phys.* **B248** (1984) 469 – 502.
[doi:10.1016/0550-3213\(84\)90607-2](https://doi.org/10.1016/0550-3213(84)90607-2).
- [69] M. L. Mangano, F. Piccinini, A. D. Polosa *et al.*, “ALPGEN, a generator for hard multiparton processes in hadronic collisions,” *JHEP* (2003), no. 07, 001, [arXiv:hep-ph/0206293](https://arxiv.org/abs/hep-ph/0206293). [doi:10.1088/1126-6708/2003/07/001](https://doi.org/10.1088/1126-6708/2003/07/001).

- [70] F. Maltoni and T. Stelzer, “MadEvent: automatic event generation with MadGraph,” *JHEP* **02** (2003) 027, [arXiv:hep-ph/0208156](https://arxiv.org/abs/hep-ph/0208156).
doi:10.1088/1126-6708/2003/02/027.
- [71] J. Alwall *et al.*, “A standard format for Les Houches Event Files,” *Comput. Phys. Commun.* **176** (2007) 300 – 304, [arXiv:hep-ph/0609017](https://arxiv.org/abs/hep-ph/0609017).
doi:10.1016/j.cpc.2006.11.010.
- [72] GEANT4 Collaboration, “GEANT4—a simulation toolkit,” *Nucl. Instrum. and Meth. A* **506** (2003) 250 – 303. *doi:10.1016/S0168-9002(03)01368-8*.
- [73] GEANT4 Collaboration, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.* **53** (2006) 270. *doi:10.1109/TNS.2006.869826*.
- [74] J. Pumplin, D. R. Stump, J. Huston *et al.*, “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis,” *JHEP* **2002** (2002), no. 07, 012, [arXiv:hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195). *doi:10.1088/1126-6708/2002/07/012*.
- [75] S. Belov, L. Dudko, E. Galkin *et al.*, “LCG MCDB—a knowledgebase of Monte-Carlo simulated events,” *Computer Physics Communications* **178** (2008), no. 3, 222 – 229, [arXiv:hep-ph/0703287](https://arxiv.org/abs/hep-ph/0703287). *doi:10.1016/j.cpc.2007.08.010*.
- [76] N. Davidson *et al.*, “Universal Interface of TAUOLA Technical and Physics Documentation,” [arXiv:1002.0543](https://arxiv.org/abs/1002.0543).
- [77] M. L. Mangano, M. Moretti, F. Piccinini *et al.*, “Matching matrix elements and shower evolution for top- quark production in hadronic collisions,” *JHEP* **01** (2007) 013, [arXiv:hep-ph/0611129](https://arxiv.org/abs/hep-ph/0611129). *doi:10.1088/1126-6708/2007/01/013*.
- [78] S. D. Drell and T.-M. Yan, “Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies,” *Phys. Rev. Lett.* **25** (1970) 316–320.
doi:10.1103/PhysRevLett.25.316.
- [79] <http://mcfm.fnal.gov/>.
- [80] J. M. Campbell and R. Ellis, “MCFM for the Tevatron and the LHC,” *Nucl. Phys. B* **205-206** (2010) 10 – 15, [arXiv:1007.3492v1](https://arxiv.org/abs/1007.3492v1).
doi:10.1016/j.nuclphysbps.2010.08.011.
- [81] R. Kleiss and W. J. Stirling, “Top quark production at hadron colliders: Some useful formulae,” *Z. Phys. C* **40** (1988) 419–423. *doi:10.1007/BF01548856*.
- [82] J. M. Campbell, R. Frederix, F. Maltoni *et al.*, “Next-to-Leading-Order Predictions for t-Channel Single-Top Production at Hadron Colliders,” *Phys. Rev. Lett.* **102** (2009) 182003, [arXiv:0903.0005](https://arxiv.org/abs/0903.0005). *doi:10.1103/PhysRevLett.102.182003*.
- [83] J. M. Campbell, R. K. Ellis, and F. Tramontano, “Single top production and decay at next-to-leading order,” *Phys. Rev. D* **70** (2004) 094012,
[arXiv:hep-ph/0408158](https://arxiv.org/abs/hep-ph/0408158). *doi:10.1103/PhysRevD.70.094012*.

- [84] J. M. Campbell and F. Tramontano, “Next-to-leading order corrections to W t production and decay,” *Nucl. Phys. B* **726** (2005) 109–130,
`arXiv:hep-ph/0506289`. [doi:10.1016/j.nuclphysb.2005.08.015](https://doi.org/10.1016/j.nuclphysb.2005.08.015).
- [85] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through O(alpha(s)**2),” *Phys. Rev. D* **74** (2006) 114017,
`arXiv:hep-ph/0609070`. [doi:10.1103/PhysRevD.74.114017](https://doi.org/10.1103/PhysRevD.74.114017).
- [86] M. Czakon and A. Mitov, “Inclusive heavy flavor hadroproduction in NLO QCD: The exact analytic result,” *Nucl. Phys. B* **824** (2010) 111 – 135.
[doi:10.1016/j.nuclphysb.2009.08.020](https://doi.org/10.1016/j.nuclphysb.2009.08.020).
- [87] N. Kidonakis, “Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution,” *Phys. Rev. D* **82** (2010) 114030.
[doi:10.1103/PhysRevD.82.114030](https://doi.org/10.1103/PhysRevD.82.114030).
- [88] M. Cacciari, S. Frixione, M. L. Mangano *et al.*, “Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC,” *JHEP* **09** (2008) 127. [doi:10.1088/1126-6708/2008/09/127](https://doi.org/10.1088/1126-6708/2008/09/127).
- [89] N. Kidonakis, “Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production,” *Phys. Rev. D* (**in preparation**) (2011)
`arXiv:1103.2792v1`.
- [90] DØ Collaboration, “Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b tagging,” *Phys. Rev. D* **74** (2006), no. 11, 112004. [doi:10.1103/PhysRevD.74.112004](https://doi.org/10.1103/PhysRevD.74.112004).
- [91] A toolkit for modeling the expected distribution of events in a physics analysis.
<http://root.cern.ch/drupal/content/rootfit>.
- [92] An object-oriented data-analysis framework. <http://root.cern.ch>.
- [93] CMS Collaboration, “Rates of Jets Produced in Association with W and Z Bosons,” *CMS PAS EWK-10-012* (2011). In preparation.
- [94] A. L. Read, “Linear interpolation of histograms,” *Nucl. Instrum. and Meth. A* **425** (1999) 357–360. [doi:10.1016/S0168-9002\(98\)01347-3](https://doi.org/10.1016/S0168-9002(98)01347-3).
- [95] An implementation of the histogram interpolation technique.
<http://root.cern.ch/root/html/RooIntegralMorph.html>.
- [96] CMS Collaboration, “Jet Resolution Determination at $\sqrt{s} = 7$ TeV,” *CMS Physics Analysis Summary JME-10-014* (2010).
<http://cms-physics.web.cern.ch/cms-physics/public/JME-10-014-pas.pdf>.
- [97] J. M. Campbell, J. W. Huston, and W. J. Stirling, “Hard interactions of quarks and gluons: a primer for LHC physics,” *Rep. Prog. Phys.* **70** (2007) 89.
[doi:10.1088/0034-4885/70/1/R02](https://doi.org/10.1088/0034-4885/70/1/R02).

- [98] R. Field, “Early LHC Underlying Event Data - Findings and Surprises,” [arXiv:1010.3558](https://arxiv.org/abs/1010.3558).
- [99] P. Biallass, T. Hebbeker, C. Hof *et al.*, “Parton Distribution Uncertainty Determination within CMSSW,” *CMS AN* **2009/048** (2009).
- [100] H.-L. Lai *et al.*, “Uncertainty induced by QCD coupling in the CTEQ-TEA global analysis of parton distributions,” [arXiv:1004.4624](https://arxiv.org/abs/1004.4624).
- [101] M. R. Whalley, D. Bourilkov, and R. C. Group, “The Les Houches Accord PDFs (LHAPDF) and Lhaglue,” [arXiv:hep-ph/0508110](https://arxiv.org/abs/hep-ph/0508110).
- [102] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>.
- [103] G. Crockford, K. Kostro, and A. Macpherson, “Data Exchange for the Fixed Display of Experimental conditions.”, [LHC-OP-ES-0018](https://cds.cern.ch/record/1250000).

Glossary

1-jet sample Pre-selected sample with exactly 1 jet.

2-jet sample Pre-selected sample with exactly 2 jets.

2plus-jet sample Pre-selected sample with at least 2 jets.

3-jet sample Pre-selected sample with exactly 3 jets.

4plus-jet sample Pre-selected sample with at least 4 jets.

BPix barrel pixel detector.

CDF Collider Detector Facility at the [Fermilab](#) Tevatron.

c.d.f. Cumulative Distribution Function.

CERN European Organization for Nuclear Research.

CMS Compact Muon Solenoid.

CSC Cathode Strip Chamber.

DØ DØ experiment at the [Fermilab](#) Tevatron.

DAQ Data Acquisition.

DGLAP Dokshitzer-Gribov-Lipatov-Altarelli-Parisi perturbative evolution.

DIP LHC Data Interchange Protocol.

DQM Data Quality Monitoring.

DT Drift Tube.

EB ECAL Barrel.

ECAL Electromagnetic Calorimeter.

EE ECAL Endcaps.

ES Preshower Detector.

Fermilab Fermi National Accelerator Laboratory.

FPix forward pixel detectors.

GCT Global Calorimeter Trigger.

GMT Global Muon Trigger.

GT Global Trigger.

HB HCAL Barrel.

HCAL Hadronic Calorimeter.

HE HCAL Endcaps.

HF Forward Hadronic Calorimeter.

HLT High Level Trigger.

HO Outer Hadronic Calorimeter.

HPD HCAL Hybrid Photodiode.

IP5 Interaction Point of CMS.

JER Jet Energy Resolution.

JES Jet Energy Scale.

LEP Large Electron Positron.

LHC Large Hadron Collider.

LHE Les Houches Events.

MCDB [LHC Computing Grid \(LCG\)](#) Monte Carlo Database.

ME Matrix Element.

MM Matrix Method.

Monte Carlo Monte Carlo simulation.

P5 Point 5 of the LHC.

PAG Physics Analysis Group.

PD Primary Dataset.

PDF Parton Distribution Function.

PE Pseudo Experiment.

PIXEL pixel detector.

pp proton-proton.

pp1 patch panel 1.

PS Parton Shower.

QA Quality Assurance.

QCD Quantum Chromodynamics.

RIB Rod-In-a-Box.

Rod A basic self-contained element populated with 6 to 12 TOB modules plural.

ROOFIT A toolkit for modeling the expected distribution of events in a physics analysis.

ROOT An object-oriented data-analysis framework.

RPC Resistivity Plate Chamber.

SM Standard Model.

SST silicon strip tracker.

TEC Tracker EndCaps.

Tevatron TeV-range proton-anti-proton accelerator at [Fermilab](#).

TIB Tracker Inner Barrel.

TID Tracker Inner Disks.

TIF Tracker Integration Facility.

TK Tracker.

TOB Tracker Outer Barrel.

TST Tracker Support Tube.

UE Underlying Event.

Appendix A

The L1 and HLT Triggers

Table A.1: Single muon based L1 trigger definition. H_T is the scalar sum of transverse energies of the leptons and the jets, and the missing transverse energy (\cancel{E}_T).

Algorithm Name	Definition
L1_SingleMu7	1 μ , $p_T > 7$ GeV/c, $ \eta < 2.45$
L1_DoubleMuOpen	2 μ , no p_T cut, $ \eta < 2.45$
L1_Mu3_Jet10	1 μ , $p_T > 3$ GeV/c, $ \eta < 2.45$ and (1 jet, $p_T > 10$ GeV/c, $ \eta < 5$ or 1 τ jet, $p_T > 10$ GeV/c, $ \eta < 3$)
(L1_SingleMu3) AND (L1_ETM20)	1 μ , $p_T > 3$ GeV/c, $ \eta < 2.45$ and $\cancel{E}_T > 20$ GeV/c
L1_HTT50	$H_T > 50$ GeV/c

Table A.2: Single muon based HLT trigger definition. Unless specified otherwise, the eta range is $|\eta| < 2.5$ for muons and $|\eta| < 3.0$ for electrons.

Algorithm Name	Definition
HLT_Mu9	1 μ , $p_T > 9$ GeV/c
HLT_Mu11	1 μ , $p_T > 11$ GeV/c
HLT_Mu13_v1	1 μ , $p_T > 13$ GeV/c
HLT_Mu15_v1	1 μ , $p_T > 15$ GeV/c
HLT_IsoMu9	1 isolated μ , $p_T > 9$ GeV/c
HLT_IsoMu11_v1	1 isolated μ , $p_T > 11$ GeV/c
HLT_IsoMu9_PFTau15_v1	1 isolated μ , $p_T > 9$ GeV/c and 1 PF τ , $p_T > 15$ GeV/c
HLT_Mu11_PFTau15_v1	1 μ , $p_T > 11$ GeV/c and 1 PF τ , $p_T > 15$ GeV/c
HLT_DoubleMu3_v2	2 μ , $p_T > 3$ GeV
HLT_DoubleMu5_v1	2 μ , $p_T > 5$ GeV
HLT_L2DoubleMu20_NoVertex_v1	2 μ , $p_T > 20$ GeV, no secondary vertex
HLT_Mu5_MET45_v1	1 μ , $p_T > 5$ GeV/c and $\cancel{E}_T > 45$ GeV
HLT_Mu5_Jet50U_v1	1 μ , $p_T > 5$ GeV/c and 1 jet, $p_T > 50$ GeV/c
HLT_Mu5_HT70U_v1	1 μ , $p_T > 5$ GeV/c and $H_T > 70$ GeV
HLT_Mu3_Ele8_HT70U_v1	1 μ , $p_T > 3$ GeV/c and 1 e , $p_T > 8$ GeV/c and $H_T > 70$ GeV/c

Appendix B

CMSSW

The CMS software framework (CMSSW) [102] is an overall collection of software built upon an Event Data Model (EDM), which allows all possible information belonging to a particular collision event to be stored collectively in one common data format. It provides essential services and tools for the simulation, calibration and alignment, and reconstruction modules that process event data for physics analysis. CMSSW provides several dedicated data formats to reduce computing time and storage space for the reconstruction and analysis steps. The RAW data format contains detector information, such as silicon strip tracker hits which are needed for the reconstruction of high-level objects such as electrons or muons. The RECO format contains high-level physics objects (e.g., tracks, primary vertices, and muons), a full record of the reconstructed hits and clusters used to produce them, and sufficient information allowing subsequent iteration of calibrations or algorithms without relying on RAW data. The Analysis Object Data (AOD) is a subset of the the RECO format designed to facilitate a wide range of physics analyses.

Appendix C

The CMS Tracker Integration at CERN

Described in this appendix are the integration activities for the CMS silicon strip tracker when the sub-detectors ([TIB](#), [TOB](#), and [TEC](#)) were transported to [CERN](#). The main work was carried out in the [Tracker Integration Facility \(TIF\)](#), which was setup specifically to accommodate all the sub-detectors of the tracking system. The goal of the integration is to make sure all the devices fully functional after inserted inside the [Tracker Support Tube \(TST\)](#), and to exercise the complete data acquisition workflow before the relocation to the final experiment site for commissioning.

C.1 The Tracker Outer Barrel Integration

The US CMS tracker group was responsible for the [TOB](#) integration, which began in March 2006 at [CERN](#). At that time different sub-detector groups were working in separate

laboratories. The integration activities later on converged to the **TIF** (which accommodates also **TIB** and **TEC**) in June 2006. Described here are the activities during the integration era. When separate shipments of the **TOB Rods** arrived at **CERN**, they were stored in the inventory and then tested independently on the XY-table by the **Quality Assurance (QA)** team. Those **Rods** validated by the **QA** tests were carried to the **TIF** for insertion into the **TST**. Next step was to validate the **TOB** segment by segment, which was defined according to the cooling pipes, electrical and optical cables, and relative geometry position inside **TST**. Divided along z direction are **TOB+** and **TOB-**, each half has six layers, each layer has either 3 or 4 segments (there are 44 segments in total). In this phase of integration, the technical team took charge in the installation. The procedure includes insertion of rods, routing and welding cooling pipes, and laying down electrical and optical connection. Meanwhile the **QA** team verified the installed sections accordingly. The **QA** team took care of tasks of testing electronic functionalities, measuring detector performance along with closely monitoring ambient temperature and humidity, and uncovering potential problems. When sufficient amount of the **TOB** segments were installed and validated, a sector test involving multi segments were carried out to examine the integrity and uncover potential problems. These steps assure the CMS experiment of a fully functional tracker that performs high quality data taking. The **TOB+** was fully validated before end of October, followed by completing the **TOB-** in November 2006. Followed by the **TIB** and **TID** insertion in February 2007, and the **TEC**+/-, the outermost parts in the z direction, were integrated into the **TST** in March.

C.2 The Rod-In-a-Box Project

The [Rod-In-a-Box \(RIB\)](#) project was carried out with a [TOB Rod](#) installed in a protective box at [P5](#) during June 2007 to March 2008. As the first piece of the silicon strip tracker installed in the experimental hall, it served the important purpose as a diagnostic device to identify possible hardware and software problems. It was setup to understand the noise environment at [P5](#) and test possible (grounding/shielding) solutions to get rid or minimize the so-called wing noise that affects the performance of the [TOB Rods](#). The wing noise was caused by the magnetic field generated by the imbalance of the currents in the power distribution lines in the RODs. The results were compared to the controlled tests carried out at the XY-table in similar hardware configuration. The [RIB](#) was also used to test local [Tracker \(TK\) DAQ](#) software within [P5](#) computing environment, and later on bridged with the central [DAQ](#) system. Furthermore, the [RIB](#) served as in-situ alarm detector to catch new noise source occurred during the installation process at [P5](#). The outcome of this project confirmed same behavior observed in previous studies and was considered as an indication of the high quality detector performance.

C.2.1 Setup of the RIB

First, an empty box, as property of the [RIB](#), was installed in the proposed setup location. Then the spare [Rod](#) was identified and tested on the XY-table before transportation to [P5](#). It was strapped on one of the [HCAL](#) wedges in the experimental cavern as shown in Fig. [C.1](#). First data successfully taken in August 2007. The [RIB](#) was moved into [ECAL](#) bore on Oct 8th, 2007 as shown in Fig. [C.2](#) to measure the effect of operating [ECAL](#)

system. The **RIB** was moved back to **HCAL** wedge for the tracker insertion (Dec 2007).

Finally, the **RIB** was retired on March 25th, 2008.



Figure C.1: Shown in this picture is the RIB in its position on top of a HCAL wedge in the experimental cavern at P5.



Figure C.2: Shown in this picture is the RIB in its position in the ECAL bore in the experimental cavern at P5.

C.2.2 Local Test

There were several local tests carried out by the RIB. One was the test of different grounding schemes analogous to the XY-table studies. We also studied different daughter card configurations at the patch panel 1 (pp1), which is a metallic box housing the power and optical interfaces between the detector front-end and the power supplies as well as the read-out electronics. A daughter card is a small piece of printed circuit board (PCB) designed to plug into the pp1 to provide different grounding schemes for the power supply lines. Figure C.3 shows the effect on noise performance of the RIB when applying daughter card at the pp1.

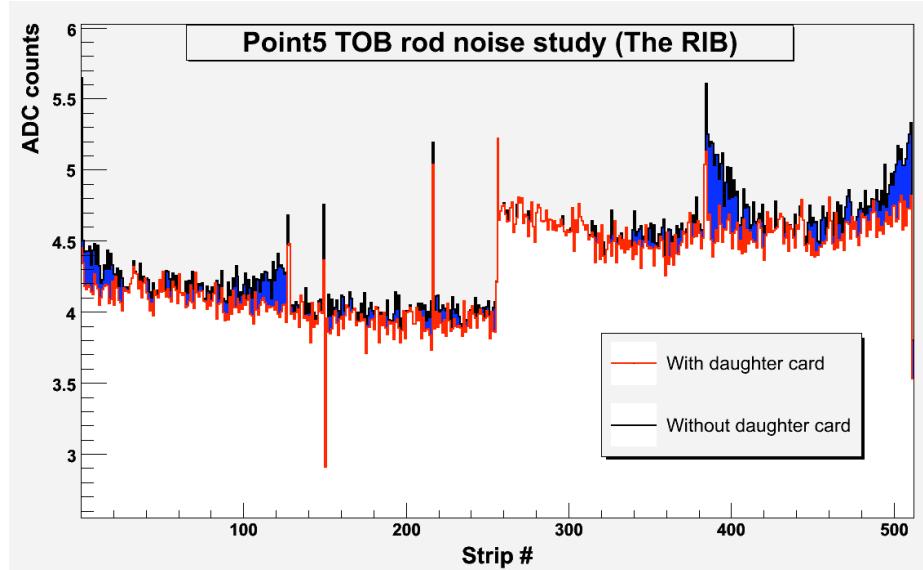


Figure C.3: Pedestal profiles of module 6 with (read) and without (black) pp1 daughter card. The blue shade displays wings. Module 6 is the module closest to the power supply cable and the ccu.

C.2.3 Global Test

To test the tracker online [DAQ](#) workflow, the [RIB](#) was included in the global runs with other subdetectors. The successful global runs with the [RIB](#) was marked as a milestone of the tracker detector being integrated with the Central [DAQ](#) and the global trigger system.

C.3 The CMS Tracker Online Software

The CMS tracking system relies not only on the sophisticated algorithms described in Section 4.2.1, but also on the quality of low-level reconstruction of the registered electronic signals. The tracker online software was designed for firsthand data acquisition and reconstruction. The validity of the online software is one of the most crucial parts of the whole tracking system. We took part in the validation task force of the Silicon Tracker Online [DAQ](#) software. The task force involved in the installation of the CMSSW software required by the [DAQ](#), the validation on new tracker software tools, managing all the CMSSW software packages used by tracking [DAQ](#), and coordinating the software releases. The validation activities were carried out on the cosmic rack and the XY-table at the [TIF](#) and the [RIB](#) at [P5](#) before the first part of the final tracker ready for the check out in mid February 2008.

Appendix D

CMS Beamspot Data Quality Monitoring

The CMS Beamspot Data Quality Monitoring ([DQM](#)), referred to as BeamMonitorDQM, make use of the $d_0 - \varphi_0$ correlation of tracks [48]. To first order, the impact parameter d_0 for tracks coming from the interaction vertex can be parameterized by

$$d_0(\varphi_0, z_p) = x_0 \cdot \sin \varphi_0 + \frac{dx}{dz} \cdot \sin \varphi_0 \cdot z_p - y_0 \cdot \cos \varphi_0 - \frac{dy}{dz} \cdot \cos \varphi_0 \cdot z_p, \quad (\text{D.1})$$

where x_0 and y_0 are the position of the beamspot at $z = 0$, and $\frac{dx}{dz}$ and $\frac{dy}{dz}$ are the x and y slopes of the beam. The $d_0 - \varphi_0$ fitter, referred to as BeamFitter, is a simple iterative χ^2 fitter. The contribution from each track is weighted by its error. The beamspot positions and slopes are extracted by minimizing the χ^2 distribution iteratively:

$$\chi^2 = \sum_{i=1}^{N_{\text{tracks}}} \left(\frac{d_{0i} - (x_0 \cdot \sin \varphi_{0i} + \frac{dx}{dz} \cdot \sin \varphi_{0i} \cdot z_{pi} - y_0 \cdot \cos \varphi_{0i} - \frac{dy}{dz} \cdot \cos \varphi_{0i} \cdot z_{pi})}{\sigma_i} \right)^2, \quad (\text{D.2})$$

where $\sigma_i^2 = \sigma_{d_0}^2 + \sigma_{Beam}^2$, and σ_{d_0} and σ_{Beam} are the impact parameter resolution and the average transverse beam width, respectively.

The workflow of the BeamMonitorDQM consists of the following steps: First, the BeamFitter is invoked to obtain fitted parameters of the beamspot. Second, the results are transferred to the CMS online DQM graphical user interface (GUI) to display the current beam conditions at [P5](#), and at the same time the results are communicated to the CERN Control Center (CCC) through the [LHC Data Interchange Protocol \(DIP\)](#) [103]. The [DIP](#) server also provides the functionality to save the results to the CMS online database in the format called scalers, which can then be fed to the online [HLT](#) system as a precise estimation of the interaction point, this is especially useful for the [HLT](#) paths that need a precise calculation of the impact parameters, e.g., b -tagging related [HLT](#). The workflow is summarized in Figure [D.1](#).

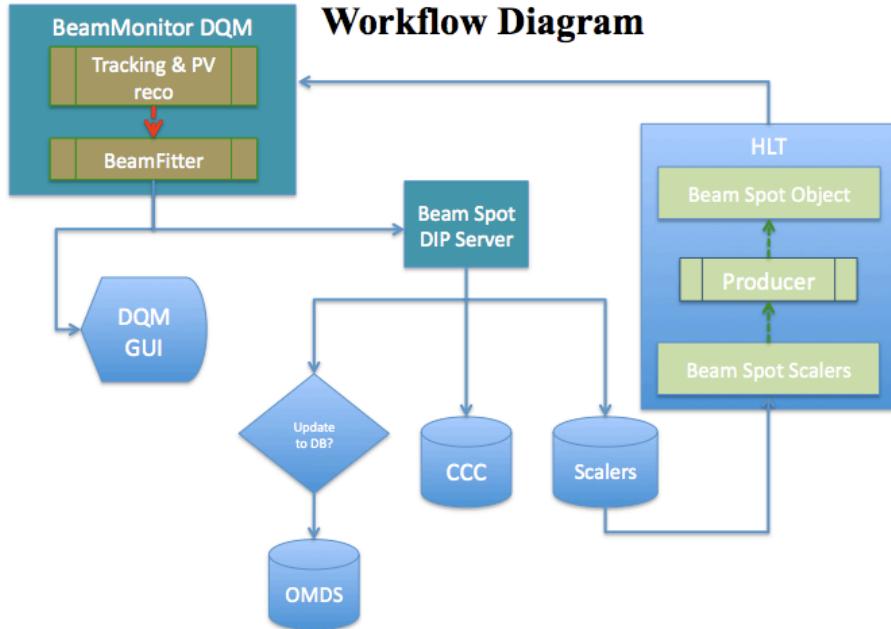


Figure D.1: Workflow diagram of the BeamMonitorDQM.