Measurement of Top Quark Mass and Production Cross-Section using $t\bar{t}$ pairs with Semi-Leptonic Decay Channel

by

Qamar-ul-Hassan

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
INTERNATIONAL ISLAMIC UNIVERSITY
ISLAMABAD, PAKISTAN
JULY 2017

CERTIFICATE

The undersigned hereby certify that the research work presented in this thesis titled **"Measurement of Top Quark Mass and Production Cross-Section using $t\bar{t}$ pairs with Semi-Leptonic Decay Channel"** by **Qamar-ul-Hassan** is carried out under my supervision in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Supervisor:

Prof. Dr. Hafeez R. Hoorani
Director General
National Center for Physics
Quaid-i-Azam University Campus
Islamabad, Pakistan

Co-supervisor:

Dr. Waqar Adil Syed
International Islamic University
Islamabad

Submitted Through:

Prof. Dr. Mushtaq Ahmad
Chairman
Department of Physics
International Islamic University
Islamabad, Pakistan

Dedicated to my NANA and my family

Abstract

A measurement of the $t\bar{t}$ production cross section at $\sqrt{s} = 13$ TeV is presented using proton-proton collisions, corresponding to an integrated luminosity of $2.2\,\mathrm{fb}^{-1}$, collected with the CMS detector at the LHC. Final states with one isolated charged lepton (electron or muon) and at least one jet are selected and categorized according to the accompanying jet multiplicity. From a likelihood fit to the invariant mass distribution of the isolated lepton and a jet identified as coming from the hadronization of a bottom quark, the cross section is measured to be $\sigma(t\bar{t}) = 889 \pm 2\,(\mathrm{stat})^{+26}_{-28}\,(\mathrm{syst}) \pm 20\,(\mathrm{lumi})\,\mathrm{pb}$, in agreement with the standard model prediction. Using the expected dependence of the cross section on the pole mass of the top quark (m_t), the value of m_t is found to be 170.6 $\pm$ 2.7 GeV.

Acknowledgement

The beauty of doing research in particle physics is that you are not alone, therefore, this thesis also includes contribution of many people in many different ways. I would like to thank my colleagues for their advice, help and support, to whom I feel greatly indebted, who made this work very enjoyable.

First of all, a big thank to my supervisor, Prof. Dr. Hafeez ur Rehman Hoorani, whose keen guidance was always there for me. When I started, I knew nothing about particle physics but their friendly supervision, inspiring enthusiasm, time for endless discussions and direct collaboration enabled me to acquire necessary knowledge to conduct this research. Then a very big thank to Pedro da Silva. This work would not be possible to the finishing line without his support and unconditional help. I am glad to be supervised by both of you. All of my technical and conceptual difficulties were solved by you and I find it hard to imagine a better supervisor.

I am also grateful to chairman, Department of Physics, IIU, Islamabad Dr. Waqar Adil Syed and Prof. Dr. Ehsan ullah Khan to solve my university issues that helped me to concentrate on my research work.

While my stay at CERN, many of my foreign colleagues and friends including, T. Arndt, M. Ivova and A. Anuar (Tag and Probe), M. Aldaya, M. Soares, J. Keveaney, E. Yazgan, M. Seidel (TOP group convenors). J. Kieseler (TOP++) and A. David (Higgs combination tool) are specially thanked for their help and support.

Many special thanks to my collegues at NCP, including Izhar, Gul, Dr. Ahmad Abbasi, Dr. Irfan Asghar, Dr. Asif Siddique, Dr.Wajid Ali Khan Dr. Taimoor Khurshid, Waqs and Shoaib, they were always there to guide me in any situation.

At the end I would like to thank Higher Education Commission (HEC), Pakistan for their continuous funding for this research. My friends at HEC were always there to solve all my financial and academic issues and queries.

Qamar-ul-Hassan

Table of Contents

Abstract **iv**

Acknowledgement **v**

Table of Contents **vi**

List of Tables **ix**

List of Figures **x**

1 Introduction **1**

2 Theory **3**

2.1 Standard Model of Particle Physics 3

2.1.1 The unit system in the SM 3

2.1.2 Basic building blocks of The Standard Model 4

2.1.3 Extension of The Standard Model 6

2.2 Top quark phenomenology 8

2.2.1 Top quark production 9

2.2.2 Top quark decay 12

2.2.3 Top quark mass 13

2.2.4 Other top quark properties 14

2.2.5 Top quark as calibration tool 15

3 The Machine **16**

3.1 The Large Hadron Collidor (LHC) 16

3.1.1 Design Parameters of LHC 18

3.1.2 The experiments at LHC 19

3.2 Compact Muon Solenoid (CMS) 22

3.2.1 Coordinate system 22

3.2.2 Tracker 23

3.2.3 Calorimeters 25

3.2.4 Superconducting magnet 29

3.2.5 Muon system 29

3.2.6 Trigger 31

3.2.7 The LHC computing grid 31

3.2.8 2015 data recorded by CMS 33

4 Event generation and simulation 34
4.1 Event generators 34
4.1.1 MadGraph and aMC@NLO 35
4.1.2 powheg and aMC@NLO 36
4.1.3 COMPHEP 36
4.1.4 pythia 36
4.1.5 herwig ++ 37
4.1.6 SHERPA 37
4.2 Event simulation 37
4.2.1 Parton density functions 37
4.2.2 Hard matrix element 39
4.2.3 Parton Shower 40
4.2.4 Matching matrix element and parton shower 42
4.2.5 Hadronization 43
4.2.6 Underlying event 44
4.2.7 Color reconnection 45

5 Event reconstruction 46
5.1 Vertex Reconstruction 46
5.2 Particle flow reconstruction algorithim 47
5.3 Lepton Reconstruction 48
5.4 Jet Reconstruction 51
5.5 Missing Transverse Momentum 57

6 Event selection 58
6.1 Datasets 58
6.1.1 Corrections applied to the simulation 60
6.2 Event selection 64
6.2.1 Object selection 65
6.2.2 Lepton selection efficiency 67
6.3 Event Rate Cross Check 70
6.4 Signal discriminating variables 76
6.5 Pre-fit distributions 77

7 Systematic uncertainties 80
7.1 Experimental uncertainties 80
7.2 Theory uncertainties 83

8 Measurement of the Top-Quark Pair Production Cross Section 87
8.1 Strategy 87
8.2 Background estimations 88
8.2.1 QCD multijet estimation 89
8.2.2 Controlling W+jets contamination 93
8.3 Method used for the statistical analysis 96
8.3.1 Fit setup 98
8.3.2 Expected performance 100

8.4 Extrapolation to full phase space 115
8.4.1 Major uncertainties 117
8.5 Cross section extraction 121

9 Top quark mass measurement **123**

Conclusion **128**

Bibliography **132**

A Fit results for trigger efficiency **152**

B H_T distributions in the W+jets simulation **155**

C Control distributions **157**

D List of Publications **179**

List of Tables

2.1 The Standard Model Forces with the corresponding bosons and their properties . . . 6
2.2 Decay modes of $t\bar{t}$ pairs . . . 13

3.1 The designed parameters of the LHC accelerator . . . 20

6.1 Primary datasets used . . . 59
6.2 Simulation samples . . . 61
6.3 Simulation samples used for systematics . . . 62

8.1 Scale factors applied to the data-based estimate of the QCD multijet backgrounds . . . 93
8.2 Scale factors derived from MadGraph and MG5_aMC@NLO samples for W+jets . . . 96
8.3 Expected and observed fit results in cut-in-categories . . . 101
8.4 List of nuisances with their impact on the fit in the cut-in-categories analysis . . . 105
8.5 Expected and observed fit results in shape analysis . . . 108
8.6 List of nuisances with larger impact on the fit in the shape analysis 114
8.7 Summary of the uncertainties considered in the extrapolation to the full phase space . . . 116
8.8 Estimated acceptance . . . 119
8.9 Acceptance and uncertainties . . . 120

9.1 The source and value of the systematic uncertainties in the measurement of the pole mass of top quark (m_t). . . . 127
9.2 Summary of the cross sections measured in this analysis. . . . 128
9.3 Summary of the cross sections measured in this analysis [209]. . . . 129

List of Figures

2.1 The fundamental particles of the Standard Model 4
2.2 LO feynman diagrams of the $t\bar{t}$ pair production 9
2.3 LO feynman diagrams of the single top 10
2.4 Summary of the LHC measurements of $t\bar{t}$ production cross section at different center-of-mass energies 11
2.5 NLO feynman diagrams of the tW production 11
2.6 Different channels of top pairs decay 13
2.7 Relationship between pole mass scheme and $\overline{MS}$ scheme 14

3.1 The injection chain of the LHC accelerator complex 17
3.2 Cross section of the LHC double bore magnet 19
3.3 The CMS detector, a longitudinal view 22
3.4 A slice of the Compact Muon Solenoid detector 23
3.5 Schematic overview of the CMS pixel detector 24
3.6 Cross section of the CMS tracker detector 25
3.7 Cross section of a quarter of the ECAL 26
3.8 Detailed view of a quarter of the HCAL 28
3.9 A slice of one quadrant of muon system in Compact Muon Solenoid experiment 30
3.10 A schematic view of the L1 Trigger 32
3.11 Data recorded by CMS as a function of time 33

4.1 *pp* collision simulation 35
4.2 The CT14 parton distribution functions 38
4.3 An example for a color reconnection 45

5.1 The concept of particle flow reconstruction 48
5.2 The graphical representation of the sensitivity of different jet clustering algorithms 52
5.3 The steps involved in the CMS jet energy correction scheme 53
5.4 L1 pileup correction 54
5.5 Corrections for the jet energy resolution 55

5.6 Secondary vertex displaced with a distance of L_{xy} in x-y plane with respect to the primary vertex 56
5.7 Schematic diagram of missing transverse energy 57

6.1 Distributions of the number of primary vertices in the combined ℓ+jets events 63
6.2 Muon trigger efficiencies 68
6.3 Muon identification and isolation efficiencies 69
6.4 Electron trigger efficiencies 71
6.5 Electron identification and isolation efficiencies 72
6.6 Median and relative variaton of the rate per run 73
6.7 Observed rate per run after the selection of muon 74
6.8 Observed rate per run after the selection of electron 75
6.9 Event yields from data and the expected $t\bar{t}$ signal and backgrounds for each of the 11 independent categories 77
6.10 Distributions in the observables used to fit the data with 2 and 3 jets 78
6.11 Distributions in the observables used to fit the data with 1 and at least 4 jets 79

7.1 Distributions for signal shapes with the b-tagging related uncertainties 82
7.2 Distributions for the signal shapes with the QCD-scale choice related uncertainties 84
7.3 Distributions for the W+jets shapes with the QCD-scale choice related uncertainties 84
7.4 Distributions for the signal shapes with the parton shower scale, hadronization, NLO generator choice, and top p_T uncertainties . . . 85
7.5 Distributions for tW shapes with the QCD-scale choice and tW/$t\bar{t}$ -related uncertainties 86

8.1 The probability of MC simulations of W+jets and $t\bar{t}$ 89
8.2 Distributions for the SIP3d variable 91
8.3 Distributions for the $\not{E}_T$ variable 94
8.4 Comparison between MadGraph and MG5_aMC@NLO for $\not{E}_T$ distributions 95
8.5 Comparison between MadGraph and MG5_aMC@NLO for $M(\ell, b)$ and min $M(\ell, b)$ variables 96
8.6 Comparison of the post-fit distributions in cut-in-categories 102
8.7 Summary of the fit results to the Asimov dataset in cut-in-categories 103
8.8 Summary of the fit results for data in cut-in-categories 104
8.9 Distribution of the inter-nuisances correlation in cut-in-categories analysis 106
8.10 Postfit correlation matrix in cut-in-categories analysis 107

8.11 Comparison of the post-fit distributions in shape analysis 109
8.12 Summary of the fit results using shape analysis 110
8.13 Summary of the fit results for data using the shape analysis 111
8.14 Distribution of the inter-nuisances correlation using shape analysis . 112
8.15 Postfit correlation matrix using shape analysis 113
8.16 Acceptance as function of the inclusive jet multiplicity 117
8.17 Observed and expected variation of the likelihood 121
8.18 Summary of the impact of the most relevant nuisance parameters . 122

9.1 Evolution of the $t\bar{t}$ cross section with the top quark mass 124
9.2 Model used for the extraction of m_t from the cross section measurement 125
9.3 Dependence of the likelihood on the top quark pole mass 126
9.4 Comparison of different PDF sets using cut-in-categories and shape analysis 126
9.5 Two dimensional contours of the PLR 127
9.6 Summary of top quark pair production cross section 129
9.7 Summary of top quark pair production cross section 130

A.1 Tag and probe fits in data $Z \to \mu\mu$ 152
A.2 Tag and probe fits in simulated $Z \to \mu\mu$ 153
A.3 Tag and probe fits in data $Z \to ee$ 153
A.4 Tag and probe fits in simulated $Z \to ee$ 154

B.1 Distributions for the H_T variable 156

C.1 Distributions for the lepton pt variable for μ events 158
C.2 Distributions for the lepton pt variable for e events 159
C.3 Distributions for the lepton eta variable for μ events 160
C.4 Distributions for the lepton eta variable for e events 161
C.5 Distributions for the H_T variable for μ events 162
C.6 Distributions for the H_T variable for e events 163
C.7 Distributions for the M($t\bar{t}$) variable for μ events 164
C.8 Distributions for the M($t\bar{t}$) variable for e events 165
C.9 Distributions for the CSV discriminator for μ events 166
C.10 Distributions for the CSV discriminator for e events 167
C.11 Distributions for the $\not{E}_T$ in μ events 168
C.12 Distributions for the $\not{E}_T$ in e events 169
C.13 Distributions for the number of b-tags in μ and e events 170
C.14 Distributions for the $M(\ell b)$ variable in μ events with 1, 2, 3 or at least 4 jets and one b-tag 171
C.15 Distributions for the $M(\ell b)$ variable in e events with 1, 2, 3 or at least 4 jets and one b-tag 172

C.16 Distributions for the $\min M(\ell b)$ variable in μ events with 2, 3 or at least 4 jets and at least two b-tags 173
C.17 Distributions for the $\min M(\ell b)$ variable in e events with 2, 3 or at least 4 jets and at least two b-tags 174
C.18 Distributions for the $\Delta R(\ell b)$ variable in μ events with 1, 2, 3 or at leat 4 jets and one b-tag 175
C.19 Distributions for the $\Delta R(\ell b)$ variable in e events with 1, 2, 3 or at leat 4 jets and one b-tag 176
C.20 Distributions for the $\Delta R(\ell b)$ variable in μ events with 2, 3 or at leat 4 jets and at least two b-tags 177
C.21 [Distributions for the $\Delta R(\ell b)$ variable in e events with 2, 3 or at leat 4 jets and at least two b-tags 178

Chapter 1

Introduction

The top quark (t) is the heaviest elementry particle known yet. Its discovery completed the third generation of quarks listed in standard model (SM). The top quark physics is a rich and still developing field. Since the mean lifetime of top quark (5×10^{-25} s) is very small as compared to the hadronization time, therefore it is the only quark used to study the bare quark properties. At Large Hadron Collider (LHC), the top quark is mainly produced in pairs via strong interactions. It can also be produced singly, at a smaller rate, via electroweak interactions.

The precise measurement of the top quark cross section opens the possibility to perform precision measurements of Parton Distribution Functions (PDFs) and the extrapolation to top quark pole mass (m_t) [1]. Also it provides an opportunity to search for new physics, the physics beyond the SM and precise measurement of the strong coupling constant (α_S).

As LHC is now running at $\sqrt{s} = 13$ TeV (Run-II), searching for new physics is particularly relevant. A precise measurement of the $t\bar{t}$ production cross section is expected to signal the presence of new physics from a deviation in the ratio of the cross section at two different center of mass energies of proton-proton (pp) collision at LHC [2]. Also a deviation in the branching ratios of different final states (lepton universality) [3] and production of supersymmetric t partners [4] are examples of potential new physics processes.

In this thesis, the study of a measurement of top quark pair production cross-section and its extrapolation to pole mass is focused, by using pp collision at the center-of-mass energy of 13 TeV recorded by the Compact Muon Solenoid (CMS) detector at the LHC during the year 2015.

The top quark being extremely elusive object, decays most frequently into a bottom quark and a W boson, $t \rightarrow Wb$ [5] and the W boson can further decay either to leptons (leptonic decay) or quarks (hadronic decay). Therefore, final states are expected to contain leptons, jets and b-jets. In this measurement, the data events with an isolated lepton (only muon or electron) and jets in final state are chosen, to study the semileptonic $t\bar{t}$ decay having a typical signature of two heavy flavored b-jets, two jets from W boson decay, missing energy and one isolated lepton (only muon or electron). The selection, however, is relaxed in order to control the main backgrounds W+jets, multijet quantum chromodynamics (QCD) and Drell-Yan (DY) from sidebands. Different techniques are explored to measure the $t\bar{t}$ mass and cross-section employing different categories and distributions. A statistical analysis based on a simultaneous profile likelihood fit to the distributions of interest is employed to extract the signal strength. After extrapolating to the full kinematic phase space, the mass and cross-section is measured.

The thesis is organized as follows: Chapter 2 gives an overview of the physics basics behind the analysis, including introduction to the SM describing the fundamental particles and interactions, the role of top quark in the SM and its phenomenology. The experimental setup, the LHC accelerator and the CMS detector is depicted in Chapter 3. Chapter 4 is devoted to explain the event generation and detector simulation procedure. Event reconstruction is detailed in Chapter 5. Event selection, trigger efficiency and identification and isolation efficiencies are discussed in Chapter 6. All the systematice uncertainties are described in Chapter 7. The measurement of top quark cross-section is described in Chapter 8. The results are extrapolated to extract pole mass of the top quark in Chapter 9, while the results are presented in Chapter 10.

Chapter 2

Theory

The measurement of $t\bar{t}$ production cross section and mass is the main topic of this thesis. This chapter is dedicated to describe the SM of elementry particle physics and some important properties of the top quark. A brief overview of the SM is highlighted in Section 2.1 while top quark phenomenology, its role in SM and the characteristics of its production and decay is explained in Section 2.2.

2.1 Standard Model of Particle Physics

The SM is a simple and comprehensive theory that embodies what the matter is and what holds it together. It was finalized in early 1970's after experimental confirmation of the existance of quarks. Despite the limitations of the SM, the experiments have established that the SM is a well-tested physics theory over the time.

2.1.1 The unit system in the SM

Since in particle physics we deal with very small particles so normal System International (SI System) units are inconveniently large to use. In high energy physics, the unit of energy of particles is electron volt (eV), so if an electron is accelerated through a potential difference of 1 volt, the kinetic energy gained by electron is: 1eV $= 1.6 \times 10^{-19}$ joules. The eV is a small unit. Other higher units used by particle physicsts are MeV (10^6 eV), GeV (10^9 eV) and TeV (10^{12} eV). This system of units is referrd to as Natural Units.

In natural units, we set reduced Plank's constant $\hbar$ and speed of light c equal to 1. So the relativistic energy-momentum relationship $E^2 = p^2c^2 + m^2c^4$ becomes $E^2 = p^2 + m^2$. Thus momenta and masses are also expressed in the units of eV. Similarly the length ($\hbar$/mc)and time ($\hbar$/mc^2) also have a unit of inverse energy.

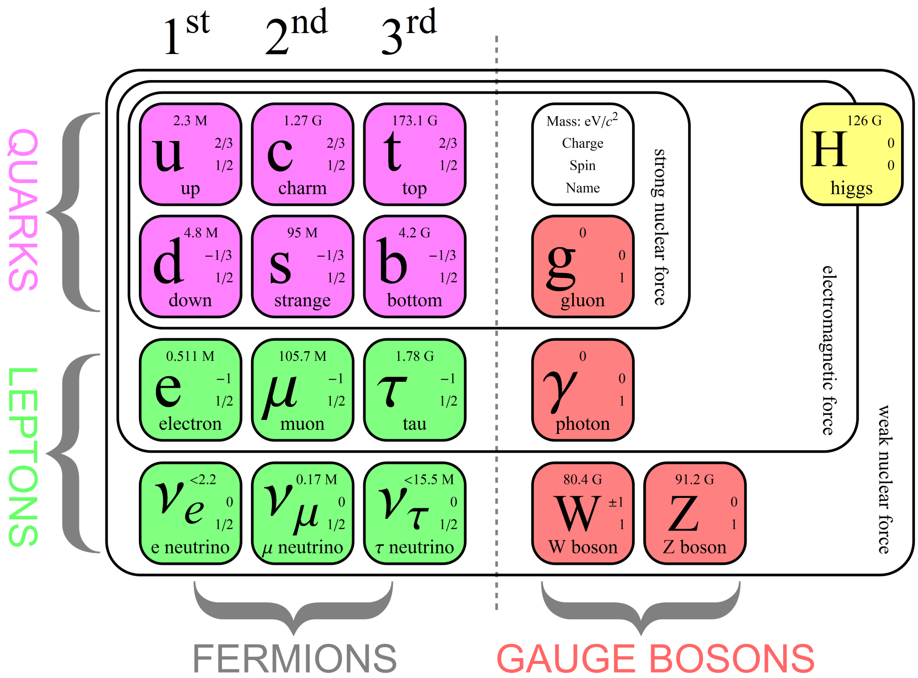


Figure 2.1: The fundamental particles of the Standard Model [6].

2.1.2 Basic building blocks of The Standard Model

The fundamental particles of nature can be categorized as fermions and bosons. The fermions are usually associated with matter while bosons, generally called force carries, are associated to the forces that govern the interactions between particles.

The particles with half-intergral spin are called fermions, which have no further sub-structures. Electrons, protons, neutrons, quarks, neutrinos are all fermions. Fermions were first predicted by Wolfgang Pauli in 1925 [7] and then in 1926, Enrico Fermi and Paul Dirac independently developed their statistical models to predict the fermions. They obey pauli exclusion principle and are characterized by Fermi-Dirac statistics. According to the SM of Particle Physics, all matter can be built up from a collection of 12 fermions and their corresponding anti-particles. The bare quarks do not exist (except top quark which will be explained in section 2.2) and they form bound states with other quarks. These bound states are known as Hadrons.

The fermions are categorized in three generations. Each generation consists of two quarks, a charged lepton and a neutrino associated with it. All ordinary matter is composed of the first generation particles, it may be named as ordinary generation. In this generation, particles can combine to form hadrons like protons, neutrons and stable atoms, having effectively infinite life spans. The particles of the second generation (may be called as exotic particles) and third generation (may be called as very exotic particles) are unstable, and therefore form unstable hadrons. They are only produced by cosmic rays in upper atmosphere or in the collider experiments. The longest-lived hadron containing a second generation quark is the lambda (λ) particle (made up of up, down and strange quarks). It has a mean lifetime less than a billionth of a second, which is considered long-lived for an unstable hadron. In third generation, the bottom quark is not much stranger than a strange quark but the top quark is so short-lived that it decays before making any bound state. The neutrinos, so small and precious, are an important subgroup within the leptons. They have very little mass and interact so weakly with other particles that they are exceptionally difficult to detect. To satisfy the Pauli Exclusion principle, quarks are given color quantum numbers (red, blue, green) [8–10]. All hadrons are colorless particles, and this is known as color confinement.

The Hadrons are further classified into Baryons and Mesons. Baryons are massive particles and consist of three quarks. Each quark is of different color. Each baryon has corresponding anti-baryon where quarks are replaced by their corresponding anti-quarks. Mesons are composed of quark anti-quark pair. Each quark is of different color but compositly mesons are colorless.

The particles with integer spin are called bosons. Gluons, photons, W, Z and

Higgs, all are bosons. Gluons, photons, W and Z have spin 1 while Higgs boson corresponds to a scalar field, so it has spin 0. The graviton, not dicovered yet, may correspond to a tensor field that could have spin 2. The statistical rules for bosons were first described by Satyendra Bose and Albert Einstein [7]. Bosons do not obey Pauli exclusion principle but follow Bose-Einstein statistics. Photon, a force carrier for electromagnetic force is an example of a massless and chargeless boson. The $W^{\pm}$ and Z are massive bosons and carry the weak interactions. The strong interactions are carried by eight massless gluons. These particles are summarised in Table 2.1.

Force	Boson	Symbol	Charge	Mass (GeV)
Electromagnetic	Photon	γ	0	Massless
Weak	W	$W^{\pm}$	±1	80.398$\pm$0.0025
Weak	Z	Z^0	0	91.188$\pm$0.002
Strong	8 gluons	g	0	Massless

Table 2.1: The Standard Model Forces with the corresponding bosons and their properties [11].

2.1.3 Extension of The Standard Model

The SM is one of the most tested theories in physics and is able to explain many observations in the electroweak and strong sector. Numerous precision measurements have been carried out to test the validity of the SM and so far so good. The outstanding successes of the SM include determination of the anomalous magnetic moment of the electron [12], the discovery of a scalar Brout-Englert-Higgs (BEH)-like boson in 2012 [13] and the prediction of the top quark in 1995 [14,15]. Despite its great success, the SM has some theoretical and experimental shortcomings. A few of them are discussed in the following.

Unification of forces

The biggest success of the SM till date is the unification of electromagnetic and weak forces, so-called electroweak force. In 19th century, electric and magnetic forces were unified in the theory of electromagnetism by Maxwell. In 20th century, a theory was developed to unify electromagnetic and weak forces to the electroweak force but there is no strong theory about the unification of the strong force. Nevertheless, it is expected that at some very high energy, the forces are indeed unified and this unification breaks spontaneously at the energy scales where the forces appear to be distinct.

Gravitational force

The general theory of relativity describing gravity can not be described in SM because a quantisation of the field of gravity leads to non-renormalisable divergences [16]. This is one of the main issues that prevents the SM from becoming the unified theory explaining all phenomena in the universe.

The hierarchy problem

The famous hierarchy problem relates the large discrepancy between the electroweak energy scale, on the order of 10^2 GeV and the Plank scale around 10^{19} GeV where gravity starts playing a crucial role. In this large gap between the two energy scales no new phenomena are predicted. Therefore the relatively light mass of the BEH boson induces very large corrections on the parameter of the SM. The SM does not explain this scientific puzzle.

Dark matter

Two different aspects that are observed in our universe and can not be explained by SM, are dark matter and dark energy. The dark matter does not interact electromagnetically with matter and it does not absorb, emit or reflect light. So it is very difficult to spot it. In our universe, the speed of rotation of galaxies does not support the generated gravity. So it is not possible to hold them togather with that gravity. The scientists believe that some undetectable matter is giving extra mass to these galaxies, generating the extra gravity to hold them together. This unkown matter is called as dark matter. It is measured through the effect of weak gravitational lensing, that mass of the galaxies consists of 80% [17] dark matter and the remaining 20% consists of ordinary matter. Studies of type 1a supernovae reveals the expanding universe which leads to the need of dark energy. Approximately 68.3% [18] of the content of universe is made up of dark energy.

Neutrinos: mixing and masses

The SM does not incorporate neutrino masses and oscillations. SM only contains left-handed massless neutrinos which are classified into three generations or flavors. So neutrinos can not form Yukawa couplings and thus not obtain their mass through the BEH mechanism [19].

2.2 Top quark phenomenology

The discovery of the Upsilon as a bound state of a bottom and anti-bottom quark at Fermilab in 1977 [20], confirmed the existance of third generation quarks. The expectations to find its generation partner, the top quark, remained alive. The top quark was firstly discovered at the Tevatron by the collaborations of CDF and D⊘ experiments [21, 22] in 1995. As the top quark is the heaviest elementry particle known till date, offers very interesting fields of study. Also the lifetime of top quark is very small [23] and is given as:

$$\tau_t = \frac{1}{\Gamma_t} \sim 0.5 \times 10^{-24} s < \frac{1}{\Lambda_{QCD}} < \frac{m_t}{\Lambda^2_{QCD}} \sim 3 \times 10^{-21} s, \tag{2.1}$$

where $\frac{1}{\Lambda_{QCD}}$ is the hadronization time scale and $\frac{m_t}{\Lambda^2_{QCD}}$ is spin decorrelation time scale. So the top quark decays before any hadronization or bound state and all the information concerning the properties of top quark is directly transferred to its decay products. The study of top quark provides possibilities to observe a quark largely unaffected by colour-confinement and these properties are crucial to test the SM predictions for the behaviour of a free quark or bare quark. The top quark is likely to be a sensitive probe of the electroweak (EWK) symmetry breaking [24,25]. In addition, top quark is a major background to many important searches Beyond the Standard Model (BSM) processes. A precise measurement of top quark properties is still of particular interest, even after the 20 years of its discovery.

2.2.1 Top quark production

At LHC, top quarks are produced by two different mechanisms. Top pairs ($t\bar{t}$) are produced via strong interaction while single tops (either t or $\bar{t}$) are produced via weak interactions. The leading-order (LO) diagrams for the production of $t\bar{t}$ pairs and single top are shown in Figure 2.2 and Figure 2.3, respectively. The LO pair production modes are either induced by the fusion of two gluons, ($gg \to t\bar{t}$) or by the annihilation of quark and antiquark ($q\bar{q} \to t\bar{t}$) [26]. The required energy to produce a $t\bar{t}$ pair is at least double the top quark mass. In the LHC, at center-of-mass energy of 13 TeV and at the corresponding partonic momentum fraction, the gluon-gluon luminosity is higher than the quark luminosity [27]. Therefore at LHC, being a proton-proton collider, gluon fusion is the more dominant production mechanism where $\sim 95\%$ of $t\bar{t}$ pairs are produced by gluon-gluon fusion [26, 27].

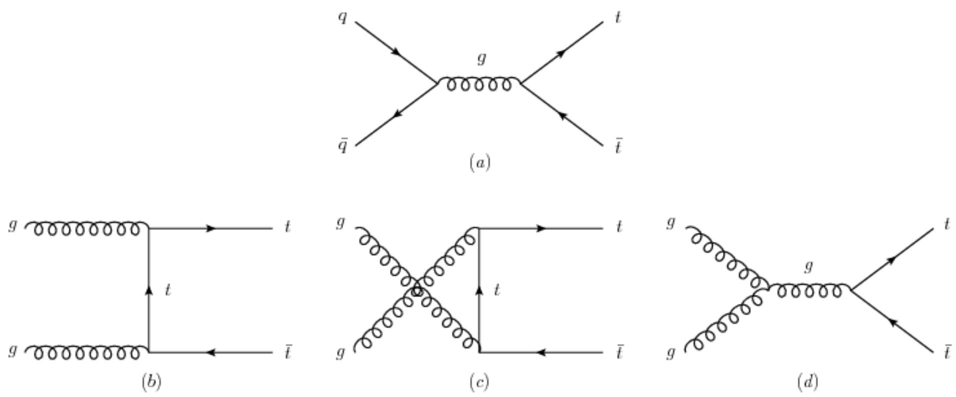


Figure 2.2: LO feynman diagrams of the $t\bar{t}$ pair production (a) via quark anti-quark annihilation and (b-d) via gluon fusion.

The constituents of a proton play an important role in the production of any particle. The content of hadrons is usually characterised by the parton distribution functions (PDFs), which is written as [28]:

$$\sigma^{t\bar{t}}_{h_1h_2}(s, m_t) = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \hat{\sigma}_{ij}, \tag{2.2}$$

where $\hat{\sigma}_{ij}(\hat{s} = x_1 x_2 s, m_t, \alpha_s(\mu_R))$ is the cross section for the partonic process $ij \to t\bar{t}$ (with $i, j = g, q, \bar{q}$) convolved with the parton distribution functions $f_i(x, \mu_F)$ of

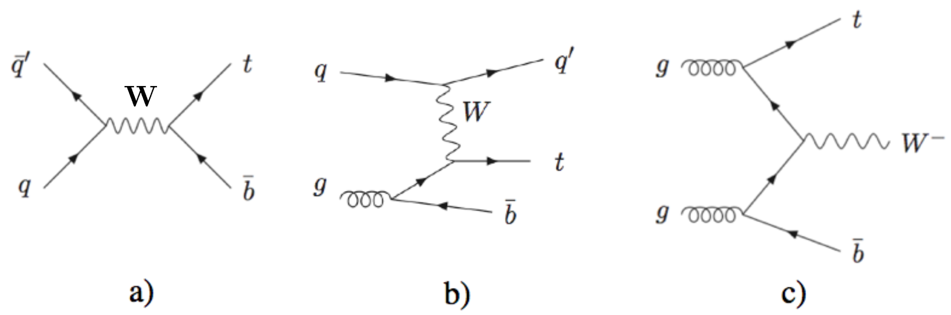


Figure 2.3: LO feynman diagrams of the single top (a) s-channel, (b) t-channel and (c) for tW-channel.

the proton. At LO, the parton distribution functions describe the probability for finding a parton i with momentum fraction x inside proton, at a factorization scale μ_F. Like the renormalization scale μ_R, the factorization scale is set to a typical scale associated with the process. A suitable choice is $\mu_F = \mu_R = m_t$. The dependence of the cross section on these artificial scales is a remnant of fixed order perturbation theory. The theory of uncertainty is evaluated by varying the scales by factors of 1/2 and 2. The uncertainty decreases with the inclusion of higher order corrections, where the state-of-the-art is next-to-next-to-leading order (NNLO) [29].

At the LHC, in addition to $t\bar{t}$ producction, also single top production via electroweak interaction is possible with much smaller cross sections. Based on the way the *W-boson* is involved in the reaction, the single-top production can be categorized into three production categories i.e. s-channel, t-channel and tW-associated production. The t and s-channels have a similar structure with different virtuality of the mediating W-boson. The s-channel production of the single top has the smallest cross section while the t-channel has the largest production cross section. In the tW-channel, the top quark is produced in association with a W-boson, requiring a bottom quark in the initial state. The LO production of the tW, where $\sigma_{tW} < \sigma_{t\bar{t}}$, is well defined but at NLO, it is very difficult to separate tW production from $t\bar{t}$ production. At NLO, σ_{tW} gets huge virtual and real corrections [31] due to contamination from $t\bar{t}$. Figure 2.6 shows the contributions to tW production

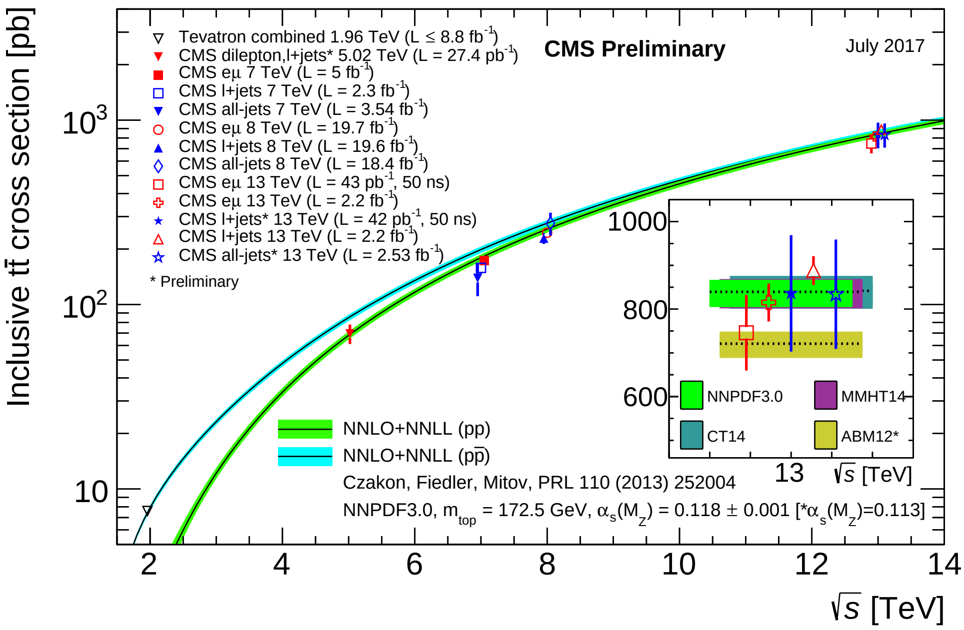


Figure 2.4: Summary of the LHC measurements of $t\bar{t}$ production cross section at different center-of-mass energies [30].

at NLO. So at NLO QCD, tW production is expected to interfere with $t\bar{t}$ production because of the similar initial and final states of some diagrams [32]. This interference contribution is large when $m_{bW} \to m_{top}$.

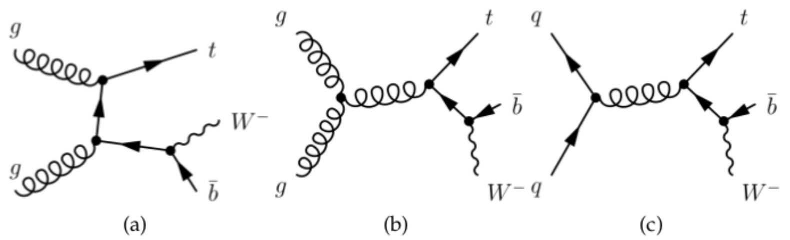


Figure 2.5: NLO feynman diagrams of the tW production.

Two schemes for defining the tW signal and distinguishing it from $t\bar{t}$ production have been compared in this thesis [33]. The difference between them measures the tW-$t\bar{t}$ interference.

In the DR method, all diagrams that enter $t\bar{t}$ production at the amplitude level are removed. Thus the interference with LO top pair production is completely removed. The removal of these diagrams can lead to the violation of electroweak and QCD gauge invariance.

In DS method, a gauge-invariant subtraction term modifies the NLO tW cross section to locally cancel the contribution from $t\bar{t}$. This method is based on the narrow approximation.

2.2.2 Top quark decay

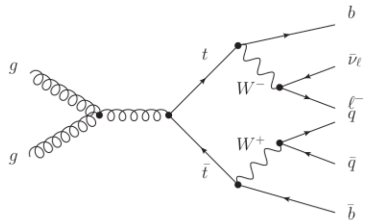
The life time of top quark is about 20 times shorter than the typical timescale for the strong interaction [23], so it is the only quark in the SM that can decay through the weak interaction. Also the large mass of the top quark distinguishes it from other quarks and results into a rich phenomenology of the top quark decays [34]. As the mass of the top quark exceeds the W-boson mass, it decays weakly via $t \rightarrow Wq$. Due to large value of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element V_{tb} [35], the top quark decays mostly to a W-boson and a b-quark [36]. The relative probabilities of the decay modes are given by the squares of the corresponding elements of the CKM matrix,

$$BR(t \rightarrow dW^+) \approx 1 \times 10^{-4},$$
$$BR(t \rightarrow sW^+) \approx 1.6 \times 10^{-3},$$
$$BR(t \rightarrow bW^+) \approx 0.998.$$

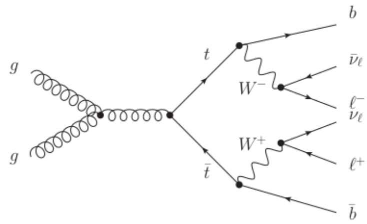
The subsequent decay of the W-boson can occur in the following decay modes:

$$BR(W^+ \rightarrow l^+ v_e) = 0.108, l = e, \mu, \tau,$$
$$BR(W^+ \rightarrow q\bar{q}) = 0.337 \times |V_{q\bar{q}}|^2.$$

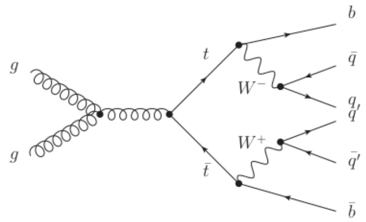
Depending on the decay modes of the W-boson [37], top quark pair events can be classified into three event categories where W-boson decays either in zero lepton (all-jets), one lepton (lepton+jets) or two leptons (dilepton). All-jets final state has the largest branching fraction. It has no "loss" by decays into τ leptons and over constrained kinematics since no neutrino from W-boson decay is present. However, it suffers from the non-resonant multijet background. The dileptonic final state, after suppression of the Drell-Yan background is the cleanest channel because it is characterized by two isolated leptons with high transverse momentum. But it also suffers with small branching ratio and event reconstruction is difficult due to the presence of two hard neutrinos. The lepton+jets channel (usually electron+jets and muon+jets) having large branching fraction, one isolated high transverse momentum lepton, at least four jets, two of which are b-quark jets, and constrained kinematics, makes it a good candidate for experimental analysis [37]. Figure 2.6 shows all categories of the top quark decay.



(a) semileptonic



(b) dileptonic



(c) full-hadronic

Figure 2.6: Different channels of top pairs decay.

$BR(t\bar{t}) \to b\bar{b} + X)$	$e^+ v_e$	$\mu^+ v_\mu$	$\tau^+ v_\tau$	$q\bar{q}$
$e^- \bar{v}_e$	1/81	1/81	(1/81)	6/81
$\mu^- \bar{v}_\mu$	1/81	1/81	(1/81)	6/81
$\tau^- \bar{v}_\tau$	(1/81)	(1/81)	(1/81)	(6/81)
$q\bar{q}$	6/81	6/81	(6/81)	(36/81)

Table 2.2: Decay modes of $t\bar{t}$ pairs, depending on the approximate W boson branching ratio. Decay modes containing τ leptons pose additional experimental challenges so that they are rarely used in the analysis of top quark properties.

2.2.3 Top quark mass

The top quark mass m_t is the most interesting parameter to dsiscuss in top quark physics. The mass of the top quark is a free and key parameter of the guage theory and SM. The two well defined schemes are the pole mass scheme and $\overline{MS}$ scheme.

In pole mass (m_t^{pole}) scheme, also known as long distance mass scheme, the classical top mass is measured by taking the top quark to at infinite distance. In this scheme top mass is measured in isolation, without any external interaction. Therefore this scheme is only valid for colorless and stable particles. But all quarks are colored particles so it has certain theoretical limitations.

In $\overline{MS}$ scheme, also known as short distance mass scheme, main advantage is the cancellation of static interquark potential and divergences in the self-enegy. The perturbative series for inclusive $t\bar{t}$ cross section with $\overline{MS}$ converges faster than the pole mass. Figure 2.7 shows the relationship between pole mass scheme and $\overline{MS}$ scheme.

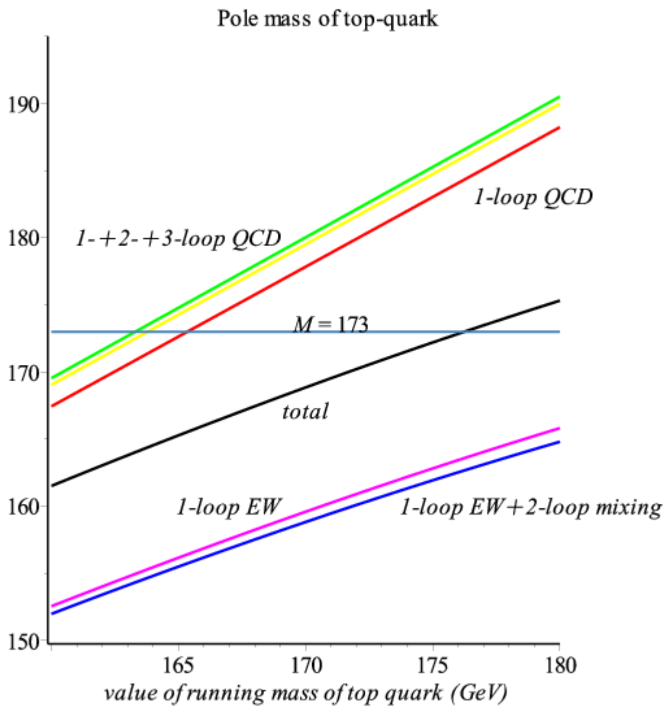


Figure 2.7: Relationship between pole mass scheme and $\overline{MS}$ scheme. It includes either QCD or electroweak corrections, or both [38].

2.2.4 Other top quark properties

A very short life time and the heaviest elementry particle known today, top quark is believed to play crucial role in electroweak symmetry breaking, serves as a window to access new physics beyond the SM and provides a unique environment to observe a free or bare quark. Few important parameters and physics searches related directly and indirectlyto the top quark study are following:

- top quark mass

- top anti-top mass difference
- $t\bar{t}$ forward backward charge asymmetry
- $t\bar{t}$ spin correlations
- ratio of branching fractions
- W helicity and anomalous couplings
- color flow in $t\bar{t}$ events
- search for flavor changing neutral currents (FCNC)

2.2.5 Top quark as calibration tool

As discussed earlier, top quarks are produced in millions at the LHC, so LHC can be considered as a factory of the top quark. The production of top quark pairs can be used to calibrate the LHC detectors in complex event-topologies. Large statistics samples of the top quark pair events are used to infer corrections to detector and reconstruction effects. Assuming the SM prediction $BR(t \to bW) = 1$, several techniques have been developed to calibrate the efficiency of b-tagging algorithms [39, 40] in both the l+jets and di-leptonic decay channels, which are used to estimate the light jet energy scale, missing energy, leptons and trigger in the l+jets channel.

Chapter 3

The Machine

The Standard Model (SM) as discussed in detail in previous chapter, has several shortcomings and limitations. Different theories have been developed to extend the Standrad Model but to test these theories experimently, we need high energy particle accelerators. A series of particle accelerators i.e. Stanford Linear Accelerator Center (SLAC), Tevatron, D0, Large Electron-Positron Collidor (LEP), Large Hadron Collidor (LHC) [41] were built to verify these theories. Currently the LHCis the world's biggest and most powerful machine where scientists, engineers and students from all over the world are involved in different experiments to study the SM and to explore Physics beyond The Standard Model.

In this chapter, Section 3.1 describes briefly the introduction of Large Hadron Collidor, its important parameters and the experiments. In section 3.2, The Compact Muon Solenoid (CMS) experiment [42] is explained in detail since all the study presented in this thesis is carried out on the data recorded by CMS during year 2015.

3.1 The Large Hadron Collidor (LHC)

The Large Hadron Collidor is the most powerful and complex particle collider ever made by mankind. It was designed to carry out proton-proton collisions at the center-of-mass energy of 14 TeV with the maximum luminosity of 10^{34} cm^2 s^{-1}. So each beam of protons have an energy of 7 TeV which is nearly 7 times more than the previous world's most energetic detector, Tevatron accelerator [43] which collides proton-anitproton at center-of-mass energy of 1.96 TeV. Although LHC is a proton-proton collidor but in some weeks of the year heavy ions (lead ions) are also collided for heavy ion physics.

The LHC was built by the European Organization for Nuclear Research, CERN, near the Swiss-French border at Geneva, Switzerland. The LHC tunnel which had a circumference of 26.7 km and is 175 meters under the surface of earth, was excavated in the 1980's for the installation of LEP. The LEP was shutdown in 2000 and the installation of LHC started in the same tunnel. The First ever proton-proton collisions at 7 TeV center-of-mass energy started in March 2010. The detector was shut down in March 2012 after achieving center-of-mass energy of 8 TeV and reopened in June 2015 for its Run-II with much higher energy ever attained by any particle accelerator of 13 TeV.

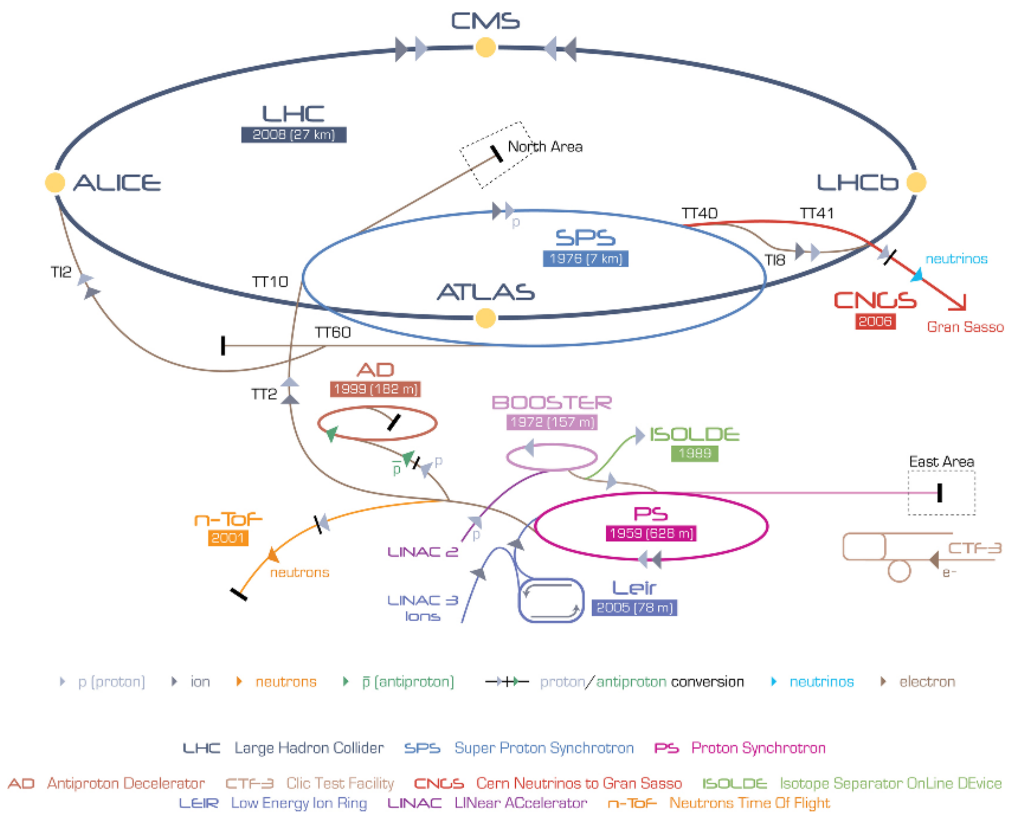


Figure 3.1: The injection chain of the LHC accelerator complex [44].

3.1.1 Design Parameters of LHC

Protons are obtained by stripping off the orbital electrons from hydrogen atoms. The protons are accelerated in several steps before being inserted to the LHC as is shown in Figure 3.1. These protons start their tour and are accelerated to first linar accelarator (LINAC 2) via a grid with constant negative charge. In LINAC 2, protons gain an energy of 50 MeV. From LINAC 2, protons are directed to a synchrotron, the proton synchrotron booster (PSB), where the protons are squeezed into bunches and their accelerating energy reaches to 1.4 GeV. From PSB, the beam is directed to a circular accelerator, Proton Synchrotron (PS). The circumference of PS is about 628.3 m. Here the beam gets its bunch structure and every bunch is spaced by 25 ns from the previous. In PS the protons are accelerated upto 26 GeV. After leaving PS, the protons enter super proton synchrotron (SPS) which is a 6 km long accelerator. In SPS, the protons are accelerated upto 450 GeV and the bunches finally enters the final LHC ring where they are accelerated to their projected center-of-mass energy , 14 TeV with in just 20 minutes.

For heavy-ion collisions, lead ions are produced using a source of vaporized lead. Then these ions are sent to linear accelerator (LINAC 3). After LINAC 3, ions are transferred to Low Energy Ion Ring (LEIR) and take the same route as the protons from the PS accelerator to the LHC final ring.

In LHC, protons are "bunched" togather rather than to have a continuous beam of particles while filling the LHC machine, a bunch seperation of 25 ns (for 2011 data taking period bunch seperation was 75 and 50 ns while it was 50 ns for 2012) is maintained with trains of 72 occupied and 12 empty bunches. Out of the 3564 bunch spaces available during each cycle, 2808 are filled. So each of the two beams of LHC, moving in opposite directions consists of upto 2808 bunches of proton with upto 1.7×10^{11} particles per bunch. During the circulation of the bunches in LHC machine, the filling is suspended. The beams remain in the machine to provide collisions for 10-15 hours and when the luminosity of beams is degraded, these are dumped and machine is refilled.

In order to keep protons on track and accelerate them to nearly speed of light, LHC uses different kinds of superconducting cavities (**SC**) and magnets (dipoles, quadropoles, sextopoles, decapoles etc) along the ring. Each of the two beams of LHC having the particles of the same charge need two opposite magnetic fields along their paths. The width of the tunnel is not sufficient to set up two different accelerators, so a special twin-bore magnet system was designed to generate the dipole field. Figure 3.2 shows the cross section of main LHC double bore magnet where two coils are integrated in one cryostat.

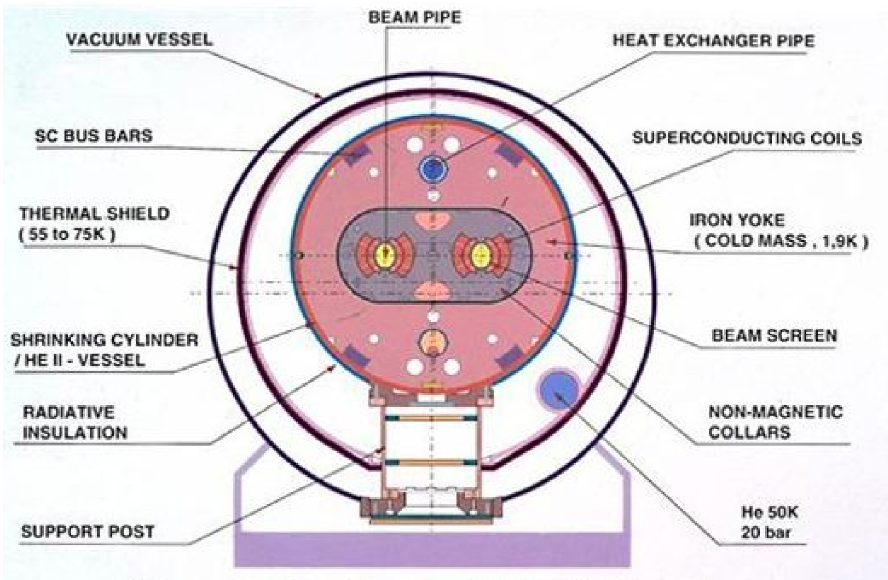


Figure 3.2: Cross section of the LHC double bore magnet [45].

The LHC consists of 1232 superconducting dipole magnets, 15 meter long and weighing over 27 tonnes each. These magnets produce a field of upto 8.36 Tesla for 7 TeV beams and are cooled by 37 million kilograms of superfluid helium to a temperature of 1.9 K to achieve superconductivity for the Niobium-Titanium windings of the magnets. Few quadrupole and sextupole magnets are also installed at several points along the ring to focus the beam and for orbital corrections.

3.1.2 The experiments at LHC

At the Large Hadron Collidor, the two beams travelling in opposite directions, collide with each other at four different places. At these places the curved shape of LHC tunnel is changed to a straight section. Along these interaction points

Parameter	Designed value
circumference [Km]	26.7
injection energy [TeV]	0.45
peak dipole field [T]	8.33
center-of-mass energy [TeV]	14
peak luminosity [cm^2s^{-1}]	10^{34}
circulating beam current [A]	0.58
number of bunches	2808
bunch population	1.1×10^{11}
bunch spacing [ns]	25
transverse normalized emittance [μm]	3.75
full crossing angle [μrad]	300
stored beam energy [MJ]	360
stored magnets energy [GJ]	11
energy loss per turn [KeV]	6.7
dipole perating temperature [K]	1.9

Table 3.1: The designed parameters of the LHC accelerator [46].

four detectors are installed. These all four detectors are different from each other depending on their specific physics programme.

A Toroidal LHC Appratus (ATLAS)

ATLAS [47, 48] is the largest detector of LHC with 46 meters long, 25 meters in diameter and 7000 tonnes weight. The volume of ATLAS is about twice of the CMS. It is a multi-layered experiment consisting of four sub-detectors. ATLAS has diverse physics programme. The purpose of this experiment is to search new physics, extra dimensions, supersymmetry (SUSY), precise study of the properties and characteristics of already discovered particles and cross check the results of Compact Muon Solenoid detector.

Compact Muon Solenoid Detector (CMS)

CMS [49,50] also has a vast physics programme. The main responsibilities of CMS is to probe the physics beyond standard model, precise and improved study of the known particles, search out the constituents of dark matter, extra dimensions, and verify and cross check the ATLAS results. A dtailed explanation of CMS is given in section 3.2.

A Large Ion Collider Experiment (ALICE)

Along with two huge general purpose detectors at LHC, there are some specific purpose experiments installed. ALICE [51,52] is also a specific purpose experiment that focuses on the study of quark-gluon plasma (QGP). The QGP is produced by heavy ion collisions. Along with proton-proton (pp) studies, main focus of ALICE is to study lead-lead and proton-lead collisions. Usually in November every year, LHC is switched to heavy ions (lead ions) instead of protons.

LHCb

This is also a specific purpose experiment which is designed to focus on heavy flavour physics, b-quark sector and CP violation that helps to investigate the differences between matter and antimatter. So the main physics objective of LHCb is to investigate the presence of antimatter in the universe. LHCb [53] experiment consists of a series of discrete detectors to detect particles in the range between 10 mrad to 300 mrad. collaboration does not cover the whole spatial range but it acts as a one-arm spectrometer which only focuses on the study of charm and beauty mesons which are predominantly produced in the forward and backword directions with small angles to the beam line.

LHCf

It is the smallest detector of LHC which is about 460 feet in front of the ATLAS collision point [54]. It is designed to perform precision measurements of the particles that are produced in the forward region to improve the understanding and stimulate the ultra high energetic cosmic rays in laboratory conditions.

Total Cross Section, Elastic Scattering and and Diffraction Dissociation Measurement (TOTEM)

It is installed close to the CMS interaction region and focuses on the forward particles that are produced in proton-proton collisions in CMS. TOTEM is specially designed to measure elastic proton scattering and total proton-proton cross-secion [55].

3.2 Compact Muon Solenoid (CMS)

CMS is located in 100 meters underground cavern near village Cessy, France. As mentioned earlier, it has a 4π geometry with diverse physics programme. The key feature of CMS is superconducting magnets that provide a magnetic field of 3.8 T. The CMS detector has multi-layered onion like structure where each layer is designed to measure the momentum and energy of specific particles. The layers inner tracking system, the electromagnetic calorimeter and hadron calorimeter are hosted inside the bore of superconducting magnet [42]. Muon detector system is located out side the magnet. It is compact in the sense that as compared to ATLAS detector, CMS having cylindrical shape, is only 22 meters long with overall diameter of 15 meters. But the weight of the CMS is 15 thousand tons.

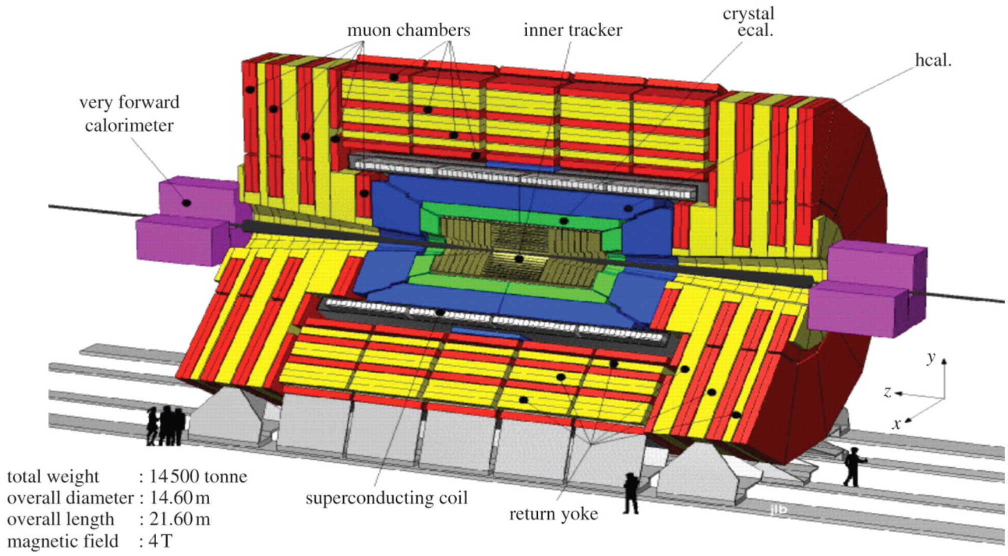


Figure 3.3: The CMS detector, a longitudinal view. The tracking system is located very close to the beam pipe which is surrounded by electromagnetic and hadronic calorimeters. The muon detection system is located out side the magnet [42].

3.2.1 Coordinate system

In order to locate a certain particle in the detector, a coordinate system is used by the CMS which is centered at the collision point. The x-axis is directed from origin to the center of the LHC while the y-axis upwards to ground surface and the z-axis is set perpendicular to x and y axis.

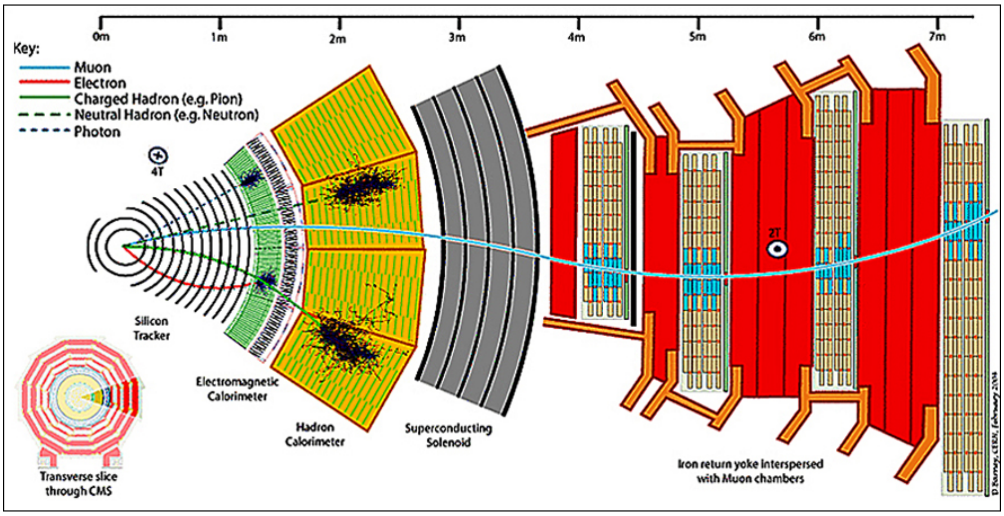


Figure 3.4: A slice of the Compact Muon Solenoid detector [42].

To locate the particles in detector, we normally use the azimuthal angle ϕ, the polar angle θ and radial coordinate 'r'. Azimuthal angle ϕ, is the angle between object and x-axis, and is measured in the x-y plane. It ranges from 0–2π. The polar angle θ (most oftenly used pseudorapidity instead of θ) is measured from the positive z-axis [?, 56]. The pseudorapidity is given as:

$$\eta = -\ln|tan(\frac{\theta}{2})| \tag{3.1}$$

3.2.2 Tracker

The CMS tracker is the innermost part of the detector and is closest to the beam pipe around the interaction point, receiving a very high particle flux. The length of this cylinderical shape detector is 5.8 meters and its diameter is 2.6 meters. The whole tracking system is immersed in a magnetic field of 3.8 T [57]. The tracking system is designed to provide an efficient and precise momentum resolution of the charged particles. Identification of primary and secondary vertices and reconstruction efficiency of charged particles, produced during high energy collisions is also performed by CMS traker system. The curvature of the tracks produced by charged particles is used to measure the momentum of the particles.

The CMS tracker system consists of two main detector technologies, the pixel detector and the silicon strip detector covering the pseudorapidity range $|\eta| < 2.5$.

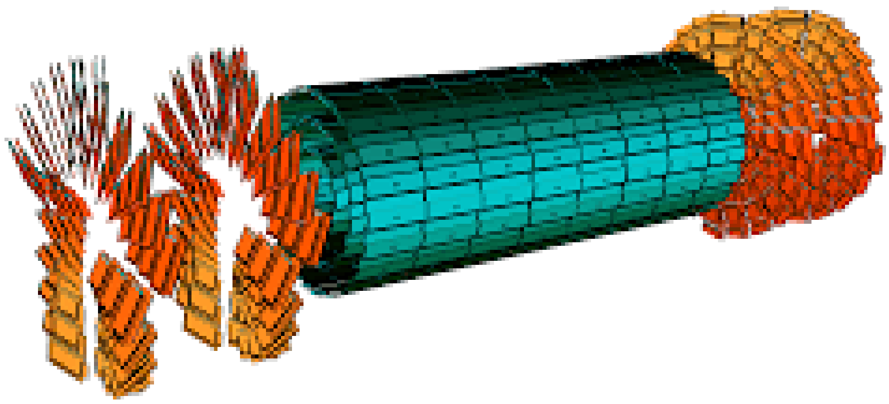


Figure 3.5: Schematic overview of the CMS pixel detector [58].

The pixel detector, as shown in Figure 3.5, is the first layer of the tracker and is located closest to the beam pipe at a radius of less than 10 cm from the beam axis. The pixel detector consists of three barrel layers installed at radius of 4.4 cm, 7.3 cm and 10.2 cm respectively and two disks on each side of the barrel are located at distances of ±32.5 cm and ±46.5 cm from the collision point. The cell size of the pixel detector is 100 × 150 μm^2. The pixel detector is organized in 1440 modules with 66 million hybrid pixel sensors.

The pixel detector is surrounded by silicon strip tracker where the particle density decreases. Figure 3.6 displays the CMS silicon strip detector. The silicon strip tracker consists of 9.6 million silicon strip sensors. It is 5.8 m long and ranges from 20 cm to 116 cm in radius. The silicon strip detector consists of several modules. The Tracker Inner Barrel (TIB) consists of four layers. The length of these layers ranges from 20 cm to 55 cm and covers the range $|z| < 65$ cm. Each side of TIB is complemented with three additional disks, the Tracker Inner Disks (TID) in the region of 65 cm $< |z| <$ 110 cm. Tracker Outer Barrel (TOB) surrounds the inner part. TOB consists of six layers of silicon strip detectors which extend from 55 cm to 116 cm and is in a range of $|z| < 118$ cm. At each side of TOB, nine end cap disks are installed. These disks are called Tracker Endcap disks (TEC) and they range from 124 cm $< |z| <$ 282 cm at a radius of 22.5 cm to 113.5 cm. The coverage of silicon strip tracker ranges up to $|\eta| < 2.5$. The inner part (TIB) provides the resolution for a single hit of 23 to 35 μm, while the outer barrel (TOB) provides the resolution of single hit between 35-53 μm. The details of the tracker system are listed at [57]

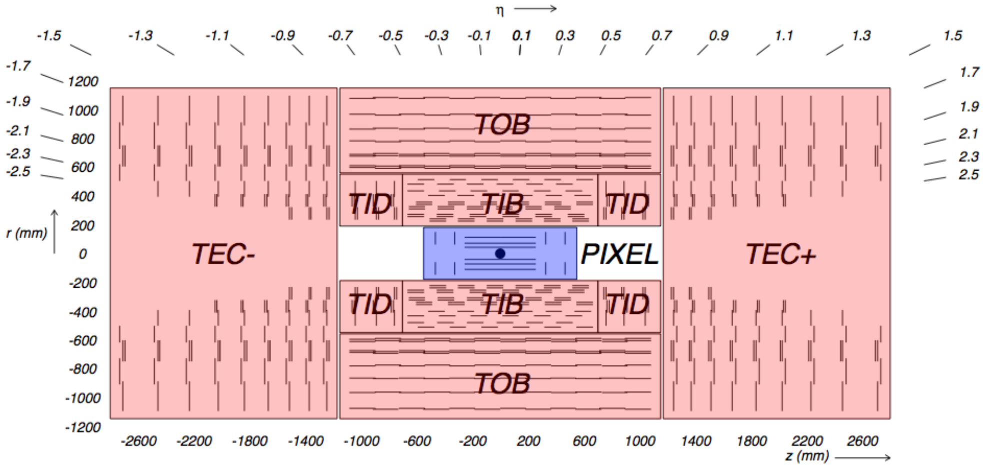


Figure 3.6: Cross section of the CMS tracker detector [59].

3.2.3 Calorimeters

The CMS calorimeter system consists of two different subsystems, the electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL). The calorimeter system is responsible to measure the energies of electromagnetically interacting particles (ECAL) and neutral or charged hadrons (HCAL).

Electromagnetic Calorimeter

The electromagnetic calorimeter is made up of very dense tungstate ($PbWO_4$) scintillating crystals. It is a homogeneous and nearly hermetic detector. The radiation length (X_0) of lead tungstate crystals is very short $X_0 = 0.89$ cm, so the ECAL can stop high energetic electrons while still remaining relatively compact. About 80% of light is emitted in 25 ns time window so it has a very fast response time. In the barrel region, this scintillation light is amplified by two silicon avalanche photodiods (APDs) and in the endcaps by vacuum phototriodes (VPTs) [60]. The general parametrization of the ECAL energy resolution is given by [?],

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 \oplus \left(\frac{N}{E}\right)^2 \oplus C \ , \tag{3.2}$$

where S stands for the stochastic term, N stands for the noise term and C is the constant. Using Equation 3.2, CMS ECAL is designed to provide excellent energy resolution of photons and electrons [61] by,

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 \oplus \left(\frac{12\%}{E}\right)^2 \oplus \left(0.3\%\right)^2 \ , \tag{3.3}$$

where the energy E is measured in GeV. The electromagnetic calorimeter is segmented into three regions, The ECAL Barrel (EB), The Endcaps (EE) and The Preshower (ES).

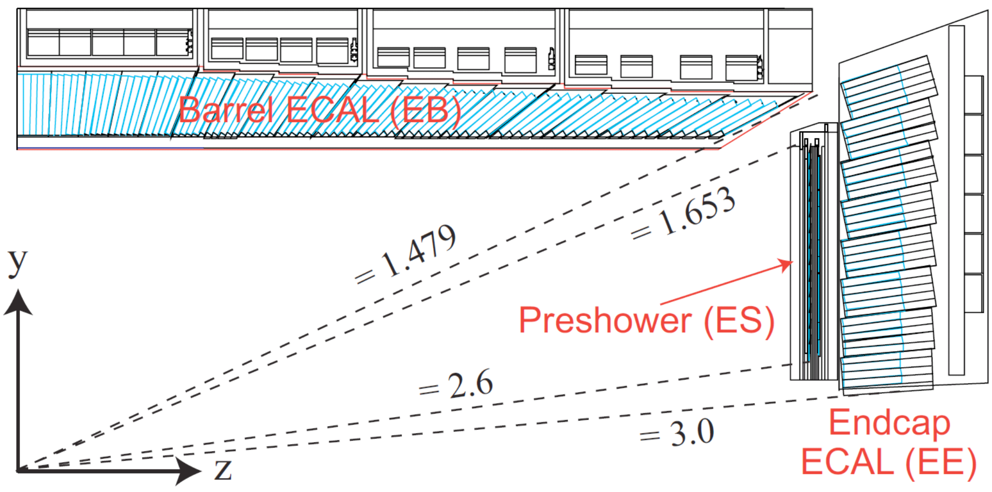


Figure 3.7: Cross section of a quarter of the ECAL [42].

The ECAL Barrel region ($|\eta| < 1.5$) is composed by 61200 scintillating crystals with inner radius of 129 cm. The length of barrel crystals is 230 mm and are mounted towards the nominal interaction point (IP) with a tilt of 3°. The crystals cover a front area of $22 \times 22 \ mm^2$ and thus cover 0.0174 in $\Delta\phi$ and in $\Delta\eta$ for $\eta \approx 0$.

The endcap calorimeter (EE) are placed 3.4 meters far from the interaction point (IP) in each side of the barrel. The endcap ECAL extends the pseudorapidity range to $1.479 < |\eta| < 3.0$. Each endcap contains 7324 crystals with surface size of 28.62×28.62 mm^2 which is slightly larger than the barrel ECAL crystals.The length of EE ceystals is 220 mm.

The last part of the ECAL is the preshower detectors (ES), installed infront of the endcaps. The coverage of ES is upto $1.653 < |\eta| < 2.6$. This cylinderical shape detector is 0.18 meters long with a total weight of 1850 Kg. The ES detector consists of two lead radiators to initiate electromagnetic cascades interleaved with silicon sensors of dimensions 6.3 cm $\times$ 6.3 cm. The silicon sensors are divided into 32 strips which form a circular shape covering the crystal endcap. The major purpose of the preshower detectors is to distinguish between photons and π° mesons. The preshower has very fine granularity than ECAL and can separate the photons in high η region.

Hadronic Calorimeter

The hadronic calorimeter is designed to provide good resolution for dijet events and it plays vital role in the measurement of the missing transverse momentum. Alternating layers of brass with a thickness of 5 cm are used as absorber material while plastic scintillator of thickness 0.37 cm is used active material. The scintillation light is converted by wavelength-shifting fibres and channeled to hybrid photodiodes (HPDs) via clear fibres. The HCAL can also be segmented into two parts, the Hadron Barrel (HB) and Hadron Endcap (HE) [63].

The barrel HCAL (HB) contains 15 brass plates and covers pseudorapidity range of $|\eta| < 1.4$. It is divided into 2304 segments ("towers"), each covering 0.087 in $\Delta\phi$ and $\Delta\eta$. As barrel HCAL is contained between ECAL and magnet, maximization of absorber material is crucial to catch the complete hadronic showers and leackage can occur. Therefore an extra scintillator layer is added just outside the magnet coil, called the hadron out calorimeter (HO) [64] which increases the thickness of the calorimeter system to about 10 hadronic interaction lengths and covers a pseudorapidity region of $-1.26 < |\eta| < 1.26$.

The endcap hadronic calorimeter (HE) on both sides of the barrel, cover the region $1.3 < |\eta| < 3.0$ partially over lapping the barrel detector and are composed of 2304 towers. The tower size in HE is the same as for HB upro $|\eta| < 1.74$. Beyond this region, the tower sizes are incrementally enlarged to a maximal $\Delta\eta \times \Delta\phi$ size of 0.350×0.174.

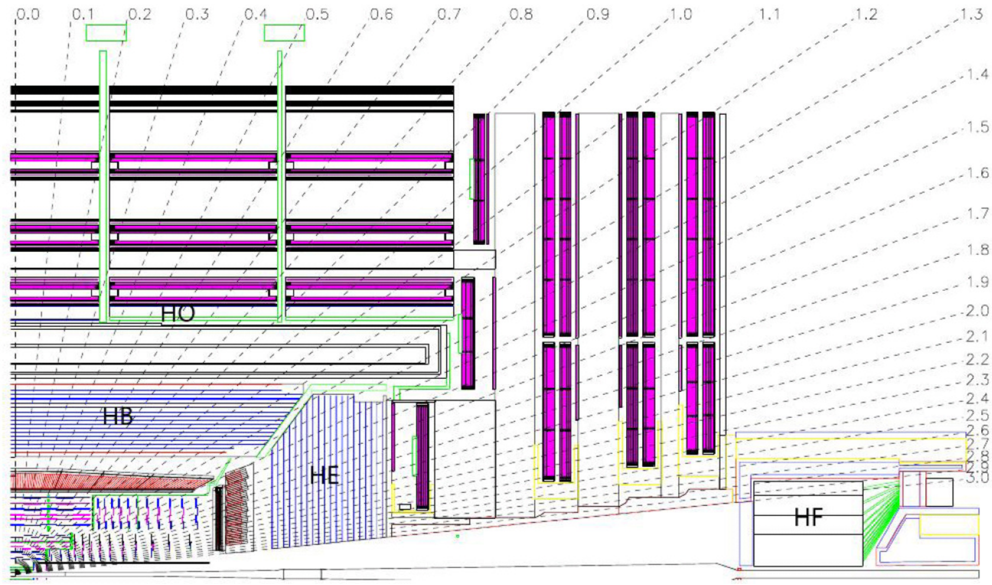


Figure 3.8: Detailed view of a quarter of the HCAL in $r\eta$-plane [42].

Finally the system is complemented with forward HCAL system (HF), installed at $3.0 < |\eta| < 5.0$ and at a distance of 11.2 m from the interaction point along the z-axis. Steel is used as an absorber and has a total depth of 1.65 m. Quartz fibres are inserted as active material. The particles emit Cerenkov light that is detected by HPDs. In total, the two HF modules contain 900 towers, having a segmentation in η of 0.1-0.3 and 10°-20° in ϕ. Each HF tower has two readout channels at different depths (8.2 and 9.5 hadronic interaction lengths) allowing for a separation of the electromagnetic and hadronic shower components.

$$\left(\frac{\sigma_E^2}{E}\right)^2 \approx \left(\frac{100\%.\sqrt{E}}{\sqrt{E}}\right)^2 \oplus (5\%)^2 \tag{3.4}$$

3.2.4 Superconducting magnet

One of the most prominent feature of CMS detector is the superconducting magnet which produces a magnetic field of 4T. The magnet is build of 50 km of superconducting coil of NbTi with a cross section of $64 \times 22mm^2$ in 2168 windings. It is 12.9 m long with internal diameter of 5.9 m. The tracker and the calorimeters are located within the coil while the muon detectors are located within the iron return yoke. The magnetic return flux in the iron yoke has a field strength of 2T, bending the path of muons into the opposite direction. The magnet produces a uniform magnetic field around the interaction point when it is operated with a current of 19.5 kA, storing a total energy of 2.7 GJ and momentum resolution [65]:

$$\frac{\delta p}{p} \approx 10\% \qquad at \quad p = 1\ TeV \tag{3.5}$$

3.2.5 Muon system

The muon system is the most outer part of CMS which provides a high muon identification efficiency and charge determination upto momenta of 1 TeV. Since muon is the least interacting and massive particle, it penetrates almost all layers of the CMS detector. Therefore, detection of muons is a key task of the CMS. The muon system is responsible for a precise reconstruction of the muons complementing the inner tracker and the calorimeter. Three different gas detectors are included in muon system, Drift tubes (DT), Cathode Strip Chambers (CSC), and Resistive Plate Chambers (RPC).

There are four DT layers, consisting of 172000 delicate wires and are placed 4.0, 4.9, 5.9 and 7.0 meters along the beam pipe in barrel region. 250 aluminum DT chambers are filled with 85% Ar and 15% CO_2 and covers the pseudorapidity range of $|\eta| < 1.2$. In this range, the neutron flux and muon rate is very low. The dimensions of drift tubes varies from 1990 $\times$ 290 $\times$ 2536 mm to 5966 $\times$ 290 $\times$ 2536 mm. Each DT station has a very good single-hit spatial resolution of around 100 μm and an angular resolution of 1 mrad.

The CSC chambers are located in the end cap region of CMS where the magnetic field is not uniform and muon flux is large. CSC's covers the pseudorapidity range ($|\eta|$) of 0.9 to 2.4. There are 486 cathode strip chambers filled with 30% Ar, 50% CO_2 and 20% CF_4. Cathode strip chambers have fast response time and high radiation resistance. The CSC's are optimized to large neutron-induced background. Spatial resolution of 486 CSC is 100-200 μm while angular resolution is of 10 mrad.

The resistive plate chambers are accompanied both by DT in barrel region and by CSC in end cap region. Each RPC has a double-gap backelite chamber with a gap width of 2 mm and operating in avalanche mode. These are filled with 95.5% of freon ($C_2H_2F_4$), 3.8% of isobutane (C_4H_10) and 0.3% of SF_6. These detectors provide very fast response and a good time resolution to assign a muon to the correct bunch crossing. The first two barrel layers of the DT system are sandwiched between two layers of RPC while the other DT layers have one accompanying RPC layer. The RPC's accopanying the CSC's in the endcap regions only cover the pseudorapidity range upto $|\eta| < 1.6$.

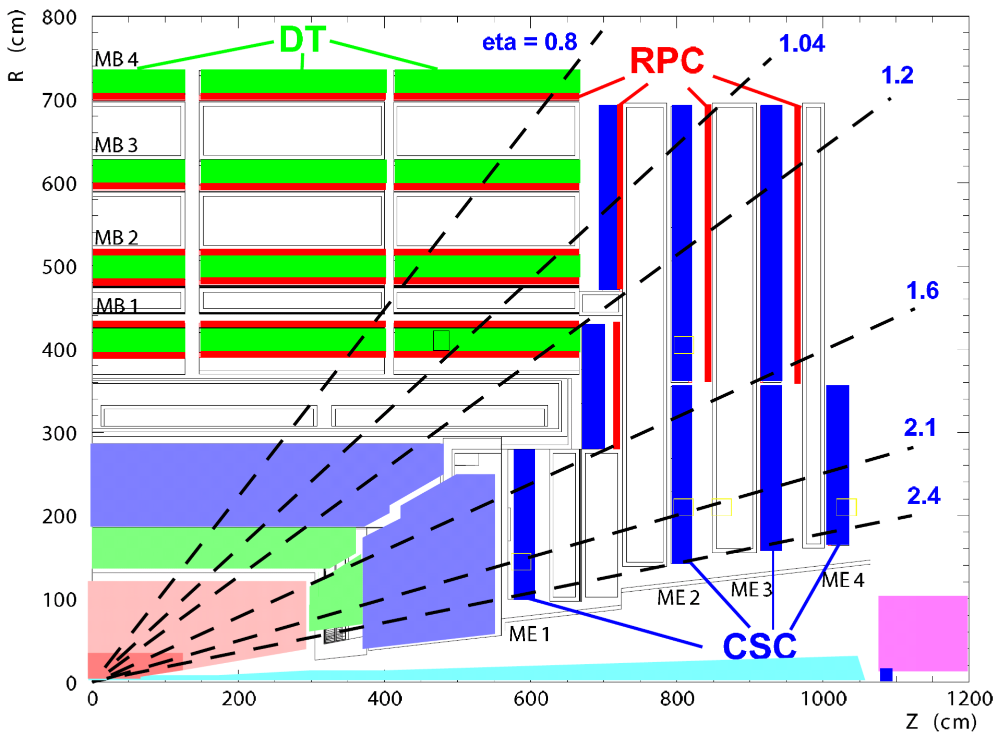


Figure 3.9: A slice of one quadrant of muon system in Compact Muon Solenoid experiment.

In total, the muon system covers an active area of 25 000 m^2 and has one million readout channels. The time resoluution for all muon subdetector systems is around 3 ns, providing excellent trigger capabilities. In combination with the inner tracker, the muon system improves the momentum resolution in the barrel for muons with $p_T > 200$ GeV.

3.2.6 Trigger

During nominall operation of LHC, at the design luminosity of $10^{34} cm^2 s^{-1}$ and 25 ns of bunch spacing, collisions rate reaches upto 40 MHz. It is not possible to store and access such high rate of data. Therefore a trigger system needs to be implemented that can select the potentially interesting events to be stored on tape. The CMS trigger system works in two steps, hardware triggers, also called Level-1 (L1) [66] and software triggers, called High Level Triggers (HLT) [67].

Level 1 trigger, a specialised electronics trigger, installed in the underground service cavern (USC). Figure 3.10 shows the structure of L1 trigger. As the LHC delivers collisions every 25 ns and this time frame is very short for full detector reconstruction, a buffer memory is installed next to the detector to host 128 collisions. L1 trigger is very fast and and decides to keep or reject an event in $\approx$ 1 μs. The L1 trigger takes crude information from the muon stations and the calorimeters. The L1 tigger reduces the data rate to 100 kHz.

The events that are passed L1 triggers are shipped to the High Level Trigger farm (HLT) which is running on large number of computers, restricting the output rate to no more than 100 Hz. The HLT uses a series of consecutive filters to reduce the data rate.

3.2.7 The LHC computing grid

In order to process and store the recorded or simulated data, LHC has designed a hierarchial structure of computing centers. The Worldwide Large Hadron Collider GRID (WLCG) makes all the collected data available to the students and scientists

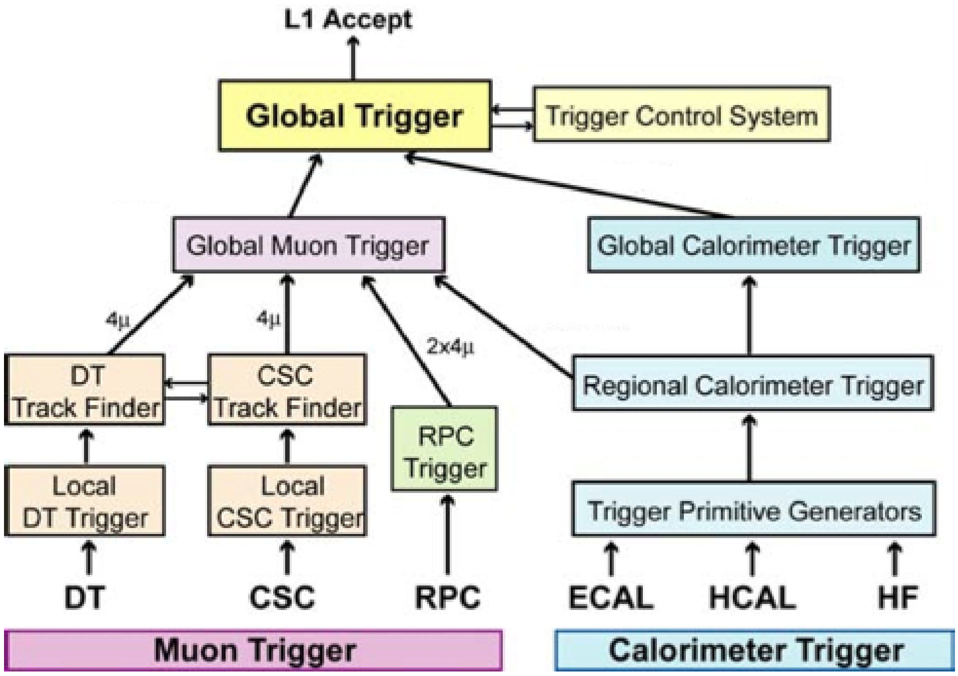


Figure 3.10: A schematic view of the L1 Trigger [68].

all over the world. The WLCG groups togather GRID systems from smaller geographical areas like the EGEE (Europe) and the OSG (United states). The WLCG consists of four types of layers or "Tiers" and allows for real-time data access. There are four Tiers in total ranging from 0 to 3. Consisting of several computer centers, each tier is designed to perform a specific set of services.

All data processed from HLT is passed through tier-0, which is located at the CERN Data Center at CERN Meyrin site. The raw data is reconstructed at tier-0. There are 13 Tier-1 centers located across the Europe, the USA and Asia. All these centers use the reconstructed data distributed by tier-0. Tier-1 centers provide round-the-clock support to grid. They are also responsible for staging the simulated samples, reprocessing the data and storage of the output. Tier-1 and Tier 0 are connected by optical fibers allowing for downloading speeds of 10 Gb/s.

Tier-2 clusters are usually universities and scientific institutes. Nearly 155 tier-2 sites are located around the world. Tier-3 are small scale, local resources. Tier-3 consists of local university clusters or individual computers dedicated to user activity and analysis tasks [69].

3.2.8 2015 data recorded by CMS

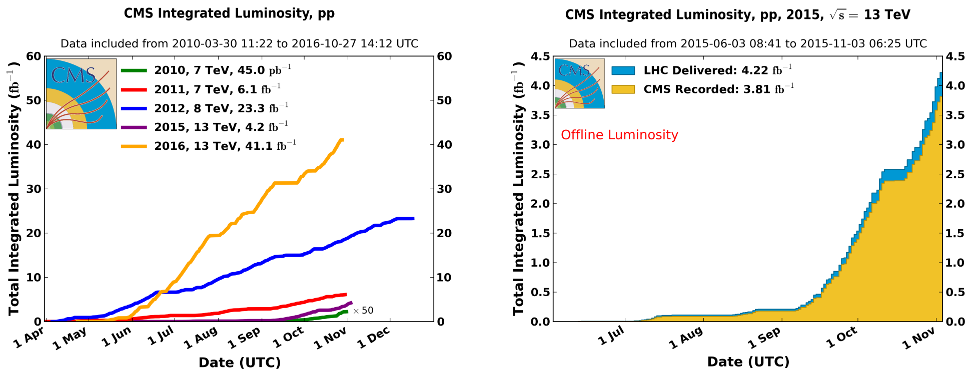


Figure 3.11: Data recorded by CMS as a function of time during the year 2015 (right) and the year 2016 (left) proton-proton collisions of the LHC at $\sqrt{s} = 13$TeV.

The experiments at LHC start recording data when the LHC machine is filled by proton beams and starts collisions. As discussed earlier, CMS is a complex machine, so recording data is not so easy task. There is a crew of experts continuously present at CMS control room and monitoring the data taking process. The team of experts include shift leader, incharge of Data Acquisition system (DAQ), trigger expert, Data Quality Monitoring (DQM) expert and Detector Control System (DCS) expert. They all monitor performance of their respective sub-detectors.

Sometimes run is stopped by any flaw of any subdetector of CMS. When the issue is fixed, all activities are resumed. During this period, LHC is providing collisions but CMS is not recording them. So the integrated luminosity recorded by CMS is always less than delivered by the LHC as shown in Figure 3.11 [70].

Chapter 4

Event generation and simulation

Event generation and detector simulation are the earliest steps in the event processing chain that leads to producing Monte Carlo (MC) samples suitable for physics analysis. MC event generation enables us to simulate the physics process of our interest. By comparing this simulated data with the real data of the LHC collisions, a deeper understading of the detector and the data cab be established. In this chapter, the event generation and simulation chain are briefly described with important steps. Section 4.1 briefly explains about the most popular event generators used at the CMS and section 4.2 offers a walk-through the simulation process.

4.1 Event generators

In CMS, a variety of event generators are available by the theory community. They include several general and some specific purpose generators. The complex procedure to generate $pp \to X$ events consists of several steps [71, 72]. As LHC is a pp collider, the protons as a whole do not enter in the physical process to be generated but rather the partons that make up the protons. A schematic view of a pp interaction down to the decays of long-lived particles in experiment is shown in Figure 4.1. The most commonly used event generators are briefly described in the next section.

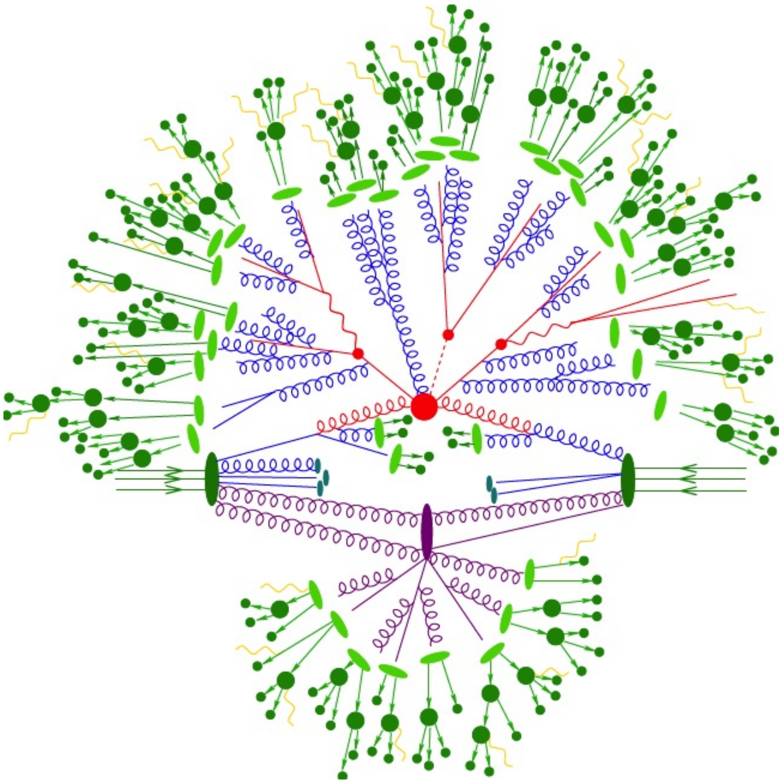


Figure 4.1: *pp* collision simulation. Matrix element generators simulate 'hard' part of scattering (red). Parton showers produce Bremsstrahlung (blue). Multiple interaction models simulate 'secondary' interactions (purple). Fragmentation models 'hadronize' QCD partons (green). Hadron decay packages simulate unstable hadron decay (dark green). Generators produce QED Bremsstrahlung (yellow). Figure shows how the partons of hadrons contribute to an interaction for some hard process in hadron-hadron collisions. [73].

4.1.1 MadGraph and aMC@NLO

MadGraph is a LO matrix element generator [74] which is able to generate the events for leading order Feynman diagram for any given process based on the SM or any other user defined model. MadGraph uses MadEvent package to generate events according to SM and BSM processes at the parton level. It automatically calculates the amplitudes for different processes and then mapping is performed for the integration over the phase space. The events generated by MadGraph are stored in Les Houches Event Files (LHEF) format. These files are then passed to pythia [75] or herwig ++ [76] for hadronization or showering. Recently MadGraph has been merged with an automated NLO generator aMC@NLO [77], which enables the production of events with NLO precision and by means of elaborated matching or merging techniques [78].

4.1.2 POWHEG and AMC@NLO

POWHEG (POsitive-Weight Hardest Emission Generator) [79] and AMC@NLO [80, 81] include a complete next-to-leading (NLO) order calculation including the virtual loop diagrams that are neglected in leading order approach like MadGraph. Since NLO calculations only calculate diagrams up to 1 additional parton [82–84], these generators have to be interfaced with a Parton Shower program to generate additional radiation. Where MadGraph produces real corrections to the LO calculation for any amount of additional partons, the NLO generators fall back to the Parton Shower approximation beyond 1 additional parton. The problem of double counting of partons when interfacing the NLO hard scattering process with leading order showering is efficiently solved by POWHEG method using the positive event technique.

4.1.3 COMPHEP

The main idea in COMPHEP [85] is to calculate cross sections and various distributions directly from Lagrangian. Unlike other generators, COMPHEP starts from the level of Feynman rules for a guage model Lagrangian and calculates symbolically the matrix element for any process defined by a user. So one can refer COMPHEP as "generator of generators".

4.1.4 PYTHIA

PYTHIA [75] is the most senior and established general purpose generator and most commonly used in high energy physics. It can be used as a standalone package both for the event generation of a physics process and also interfaced with other packages for the showering and hadronization. PYTHIA uses lund string model [86] for hadronization.

4.1.5 HERWIG ++

HERWIG ++ (Hadron Emission Reactions With Interfering Gluons) [87] is also a general purpose MC generator. In HERWIG ++, cluster fragmentation model is utilized for hadronization. The main difference between PYTHIA and HERWIG ++ is that in HERWIG ++ spin correlations between particles is conserved while in PYTHIA is destroyed.

4.1.6 SHERPA

SHERPA (Simulation of High-Energy Reactions of PArticles) [88] utilizes cluster model for hadronization. Parton showering is based on p_T -ordering and is improved to a dipole shower. It is a more recent tool included in matrix-element generators.

4.2 Event simulation

The events generated by the event generators are then passed through the detector simulation. The CMS detector simulation is based on an object-oriented framework GEANT 4 [89]. Detector geometry and the interaction of generated particles with the detector material, various magnetic field models and energy range of initial particlles can be controllled by the user. Multiple scattering, hadronic interactions, bremsstrahlung and electromagnetic interactions are taken into account during the process of detector simulation. The core of event simulation is the matrix-element calculation of the process under interest. For obtaining observable hadrons and passing them to a detector simulation, parton shower and hadronization models are needed.

4.2.1 Parton density functions

All the processes discussed in this thesis are induced by the interaction between high energetic proton beams. Each proton consists of three valence quarks (two up quarks and one down quark), which defines the quantum numbers of the proton. Moreover quarks and gluons can appear from quantum fluctuations inside the proton, violating four-momentum conservation. This can be defined by Heisenberg's uncertainty principle (but for very short interval of time) $\Delta t \Delta E \geq \hbar/2$ called partons. In LO scenario, the parton density function (pdf) can be defined as the probability to find a certain parton with the momentum fraction x. At low momentum transfer ($Q^2 = 10\ GeV^2$), three valence quarks become more and more dominant in the nucleon. At high Q^2, there are more and more quark and anit-quark pairs which carry a low momentum fraction. The PDFs are obtained from measurements on deep-inelastic scattering events using electron-proton collisions at SLAC, ZEUS and HERA, muon-proton collisions at NA4 (CERN) and neutrino-proton collisions at Fermilab.

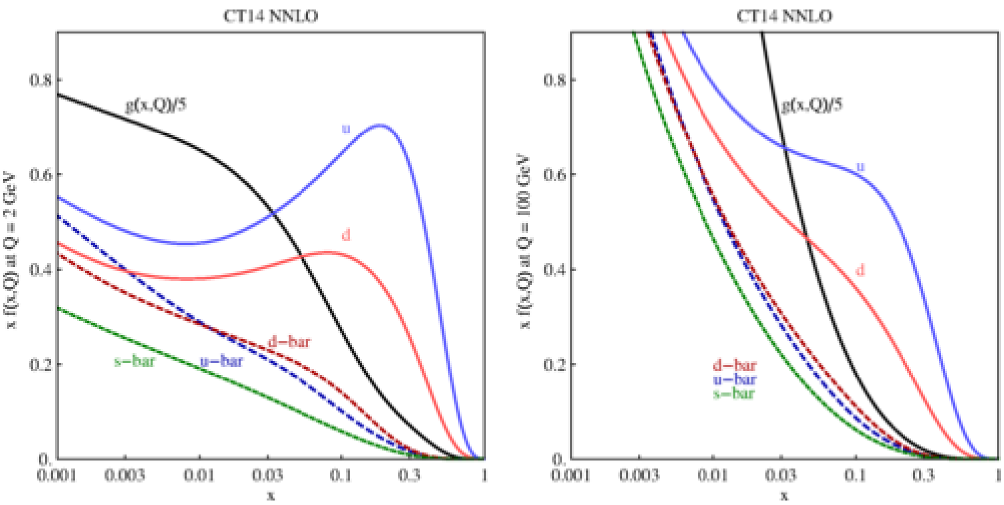


Figure 4.2: The CT14 parton distribution functions at $Q = 2$ GeV and $Q = 100$ GeV for $u, \bar{u}, d, \bar{d}, s = \bar{s}$ and g [90].

The DGLAP equations [91–93] are used for extrapolation between measurements at different Q^2, whcih are based on evolution kernels that contain the matrix elements for gluon radiation and splitting. In this thesis, the main parton distribution function (PDF) set used for PYTHIA MC is the CT14NNLO PDF set. The PDF for the particular Q^2 scale within CT14NNLO are shown in Figure 4.1.

4.2.2 Hard matrix element

In the CMS, as discussed earlier, different matrix element generators are available which generate events at the hard-process level with different orders of the perturbation theory. The user can specify the process to generate, the total number of events and the center-of-mass energy at which the process should take place. In perturbation theory, the differential cross section of any final state F is given by:

$$d\sigma_F = \sum_{k=0}^{legs} d\Phi_{F+k}|\sum_{l}^{loops} = 0M_{F+k}^{(l)}|^2, \tag{4.1}$$

where Φ_{F+k} is a specific momentum configuration, $M_{F+k}^{(l)}$ denotes the matrix element for the final state F, k is additional outgoing partons (legs) and l are the loops in the corresponding diagram. At LO simulation, no loop or additional quark or gluon is required. The general purpose generators like PYTHIA [75] and HERWIG [87] are usually implemented by tree level matrix element with $k = l = 0$. For $k > 0$, the matrix element contains an accurate description of additional quarks and gluons (or photons) in the final state. Currently, about $k \approx 5$ additional partons can be calculated. The highest-multiplicity matrix elements are implemented in ALPGEN with $k = 6$ additional partons. MadGraph [74] and SHERPA [88] can automatically calculate the matrix elements for $k = 3$ or $k = 4$ for $t\bar{t}$.

At NLO matrix elements, typically defined as $k + l = 1$, include one additional loop and one additional real emission. The KLN theorem [94, 95] states that due to unitarity the infrared (IR) divergences of the loop amplitudes cancel against those from the soft and collinear real emissions for constant $k + l$ as the singularity structure is similar. The NLO cross section can thus be rewritten with process-independent functions (Catany-Seymour dipole [96] or antenna [97] functions) that contain the singularities. These singularities are subtracted from the real emission but they are added to the loop integral calculation. The latest versions of MadGraph and SHERPA contain (interfaces to) loop amplitude generators [98–

100] so that matrix elements with multiple emissions can be evaluated at the 1-loop level. For more than one additional parton, the infrared divergences are not fully canceled, they are regulated by using kinematic cuts.

4.2.3 Parton Shower

The color charged particles are produced by the matrix-element generator involving large momentum transfers and have violently accelerated partons mostly. These color charged gluons and quarks, both in initial and final state, radiate further gluons and quarks leading to a parton shower. The former is called Initial State Radiation (ISR) and the latter is called Final State Radiation (FSR). This leads to the production of large number of particles.

Therefore, the simulation of this step has to rely on statistical models based on approximations. It is typically formulated by performing successive branches on the partons from high to low scales in which the partons are confined into hadrons ($\approx$ 1 GeV). There are three different possible branches, a quark radiating a gluon ($q \to qg$), a gluon emitting second gluon ($g \to gg$) and a gluon spliting into $q\bar{q}$ ($g \to q\bar{q}$).

In case of well-seperated and hard enough parton, matrix element calculation can be used to describe the radiations. However, when the radiated parton is at small angle with the mother parton (collinear divergence) or when the energy of one of the partons vanishes (soft or infrared divergence), then parton shower approach is applied. The evolution of the parton shower from the hard interaction energy scale Q^2 down to a lower scale, can be described by DGLAP equations [91–93]. If a parton a splits into two partons b and c, and the energy fraction $z = \dfrac{E_b}{E_a}$ is carried by parton b, the parton c is left with energy fraction *1-z*. The branching probability can be written as:

$$d\mathscr{P}_{a\to bc} = \frac{\alpha s}{2\pi}\mathscr{P}_{a\to bc}(z)dtdz = \frac{\alpha s}{2\pi}\frac{dQ^2}{Q^2}\mathscr{P}_{a\to bc}(z)dz, \tag{4.2}$$

where $\mathscr{P}_{a\to bc}(z)$ are called the splitting functions. The splitting functions for above three equations can be written as:

$$\begin{aligned} \mathscr{P}_{q\to qg}(z) &= C_F \frac{1+z^2}{1-z}, \\ \mathscr{P}_{g\to gg}(z) &= N_F \frac{(1-z(1-z))^2}{z(1-z)}, \\ \mathscr{P}_{g\to q\bar{q}}(z) &= \frac{n_f}{2}(z^2+(1-z)^2), \end{aligned} \tag{4.3}$$

where $C_F = \frac{4}{3}$ is the color factor, $N_C = 3$ is the number of colors and n_f is the number of quark flavors.

Above two equations indicate that in case of soft divergence ($z \to 0$) or collinear divergence ($Q^2 \to 0$), the branching probability diverges. This divergence is removed by applying a cutoff scale Q^2_{min} of the order of 1 GeV^2. The divergence problem is resolved but the branching probability still remains larger than 1. A Sudakov factor [101] is added in equation 4.2 to cancel this effect. The branching probability now becomes:

$$d\mathscr{P}_{a\to bc} = \frac{\alpha s}{2\pi}\frac{dQ^2}{Q^2}\mathscr{P}_{a\to bc}(z)dz\ \underbrace{exp\left(-\sum_{b,c}\int_{Q^2}^{Q^2_{max}}\frac{dQ'^2}{Q'^2}\int\frac{\alpha s}{2\pi}\mathscr{P}_{a\to bc}(z')(dz')\right)}_{Sudakov factor}, \tag{4.4}$$

here the sum runs over all possible branchings. The parton showers are implemented as iterative Markov chains. Sudakov factor for a random number between 0 and 1 is solved to find the radiation scale Q_2 while the starting scale of the hard process being Q_1. In the case of FSR, the process is iterated, starting from a hard interaction scale Q_2 down to a lower cutoff scale Q_0.

The ISR is more complex, as the incoming protons have a structure. The incoming parton momenta is fixed by the hard process, so that backwards evolution is necessary to generate the initial state shower. The Sudakov factor then contains the ratio of the PDFs at the different scales, the backwards evolution is suppressed when no matching “mother” partons are available in the PDF.

4.2.4 Matching matrix element and parton shower

Matrix element generators like MadGraph generate LO Feynman diagrams for any given process based on the SM or any user defined model. They produce hard and wide angle emissions. The parton shower programs, on the other hand, produce soft and collinear emissions. They are computationally cheap and has no constraint on parton multiplicities. Hence, both the approaches should not be treated as alternatives, rather they should be merged to give the optimal description of the data. This merging of matrix element and parton shower is faciliated in the Les Houches Accord [102]. The Les Houches Accord format is used to introduce a generic format to encode parton level information to be used as input to any general purpose generator. By merging parton showers with higher order matrix elements, the double counting of radiation might occur. For example, a $t\bar{t}$ event generated by MadGraph is accompanied by two jets. When this event is passed to pythia for parton showering and hadronzation steps, no further jets are emitted. In any other event, if MadGraph generates only one further jet and then pythia also provokes the emission of one further jet. Then these two events have the same emission multiplicity, and cause double counting. This problem is corrected at NLO. The MLM Matrix Element-Parton Shower (ME-PS) matching technique can be used to cancel any double counting. aMC@NLO and powheg have their own treatment to match NLO to the parton shower.

The three methods used for matching are outlined here. In aMC@NLO method [80, 103], full NLO matrix element computation is performed including *n+1* parton real and virtual corrections. Then it is calculated analytically how a first branching in the shower evolution of a *'n'* parton event would populate the *n+1* parton phase space. The result can be considered as a correction term to the *n+1* parton matrix element calculation. The remaining two event population, *n* and *n+1* partons, are passed to parton shower programme and are then added. In this approach, there is no guarantee the the ME is always above the PS in the phase space even though both converge in the collinear and soft region. So where the shower approximation is larger than the exact matrix element, this approach yields negative weights. Thus statistical accuracy of the generated sample is degraded by 20–60%.

In the POWHEG technique [83], after performing full NLO computation, the parton shower evolution is performed. This approach contains the hardest emission and positive virtual corrections so that no negative weights occur. The basic idea of this technique is to swap the hardest emission in the parton shower evolution with the NLO emission generated by POWHEG. The parton shower evolution is performed only starting from the p_T scale of the emission of the extra parton in the matrix element.

In MLM matchning approach [104,105], the basic idea is to veto the parton showers that lead to the same multi-parton final states as described by the matrix element itself. In this method, parton shower programme is run through the events from the matrix element computation that pass a number of acceptance cuts. The partons produced in the parton shower are clustered using a jet cone size R and are labeled as "jets". The matrix element parton is matched to the jet untill all partons have been matched.

4.2.5 Hadronization

After the evolution of parton shower down to a cut-off scale Λ_{QCD} in the order of 1 GeV, we enter the low-momentum transfer and long distance in which perturbative approach breaks down and the colored partons created in the parton shower and the hard interaction start to group into colour neutral hadrons. This is called *hadronization*.

The first step of hadronization process is fragmentation which is goverened by non-perturbative QCD. In PYTHIA, this process is implemented following the lund string model [86,106]. According to this model, a color-connected quark-antiquark pair ($q\bar{q}$) from the parton shower spans a color string with a tension $\kappa \approx 1$ GeV/fm. When these two color-connected partons seperate from each other, the potential energy stored in the field between them increases and the string can break at some point creating a new $q\bar{q}$ string, with a probability proportional to $exp(-\pi m^2/\kappa)$. If the invariant mass of either of these two strings is still large enough further string breaking might occur untill on-shell hadrons remain. The momentum transverse to the string direction is measured by a Gaussiann distribution with $\sigma \approx 0.3$ GeV. For light quark flavors, the symmetric lund fragmentation function is given by

$$f(z) \propto \frac{1}{z}(1-z)^a exp\left(\frac{-bm_\perp^2}{z}\right) \ , \tag{4.5}$$

where the Lund parameters have default values (but are tunable) of a $= 0.3$ and b $= 0.58$ GeV^{-2} and $m_\perp^2 = E^2 - p_z^2$ is the transverse mass of the new hadron. For heavy flavor quarkks, charm and beauty, lund fragmentation function is modified as [107]:

$$f(z) \propto \frac{1}{z^{1+r.bm_\perp^2}}(1-z)^a exp\left(\frac{-bm_\perp^2}{z}\right) \tag{4.6}$$

the parameter 'r' has a value between 0 and 1 and can be set for charm (r_c) and bottom (r_b) fragmentation seperately.

The cluster model, with different implementations in HERWIG and SHERPA, starts with forcing $g \to q\bar{q}$ branchings. After that, color singlet *clusters* are formed. If the cluster mass is above some threshold, it decays to two lighter clusters, untill all clusters are decayed to a pair of hadrons.

4.2.6 Underlying event

The underlying events (UE) at CMS is defined as the activity that is not part of the final-state activity originating from the most energetic parton scattering of collision [108]. Since the proton is composed of several valence quarks and sea quarks plus gluons and only one parton from each proton is resolved to produce the hard interaction while the other coloured remnants of the protons continue in almost the same direction as the initial proton. These coloured remnants interact with each other, hadronize and produce other particles. So the UE activity consists of hadrons coming from the initial and final state radiations (ISR,FSR), multiple parton interactions (MPI) possibly with their own initial and final radiation and proton remnants concentrated along the beam direction.

4.2.7 Color reconnection

Although color reconnection (CR) is well established from $B \to J/\psi$ but it is hard to imagine its prominent role in a high energy pp collision. Color reconnection models allow non-perturbative changes in the color configuration of the event before the hadronization stage, typically reducing the total potential energy between QCD color changes [109]. New CR models for pp collisions into PYTHIA 8 [110] are implemented to offer a broader spectrum of possibillites, partly to add further physics aspects, such as space-time and color structure to provide more realistic scenarios. Still, neither of the current models provide a satisfactory description of the individual p_T spectra for the identified hadrons. An example of color reconnection is explained in Figure 4.3.

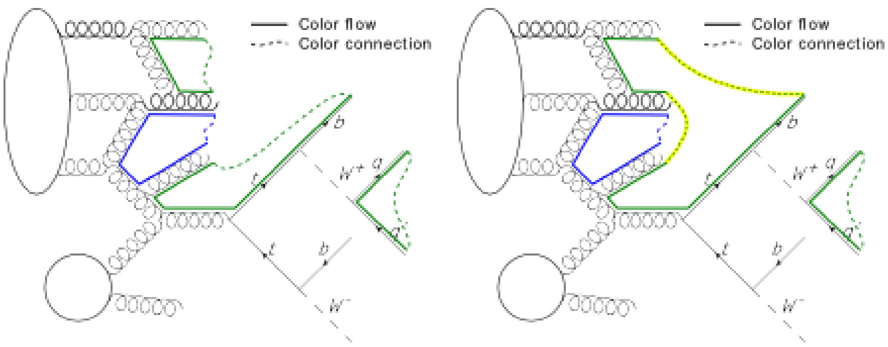


Figure 4.3: An example for a color reconnection that is likely to alter the kinematics of the bottom jet in a top-quark pair event. For simplicity, the color connections are attached to large-scale partons. In the real model, their ends are sitting at low-scale branchings further down the parton shower evolution.

Chapter 5

Event reconstruction

This chapter describes the offline reconstruction of physics objects. In order to analyze the data, it is important to translate the digital signals recorded by the CMS detector to the physical quantities. The information collected from all subdetectors [111] of the CMS is used to reconstruct these objects with an off-line software. The ParticleFlow algorithm (PF) [112, 113] is used to reconstruct the main physics objects like muons, electrons, taus, jets, missing transverse energy (MET) and photons.

5.1 Vertex Reconstruction

A sample of fully reconstructed charged particle tracks originating from the interaction point [114] are selected. The selection is based on the distance between track and beam spot at closest approach (impact parameter significance) below 5σ, p_T of the tracks and their fit quality. The Deterministic Algorithm [115] is used to cluster the tracks togather in vertex candidates based on the z-coordinate of their impact parameter. The primary interaction vertices are reconstructed by selecting high quality tracks with atleast five hits in the inner tracking system, with atleast two hits in the pixel tracker and a normalized $\chi^2 < 20$.

A vertex candidate is fitted using the Adaptive Vertex Fitter [114, 116]. In this iterative fit, the algorithm finds as many vertices as possible without splitting true vertices. It also assigns a weight w between 0 and 1 to each connection between a track and a vertex candidate, based on their positions on z-axis along with corresponding uncertainties. A minimal weight of 0.5 is required for each connection, and atleast two tracks of a candidate vertex must be incompatible ($w < 0.5$) with other vertices. The number of degrees of freedom for a fitted vertex is defined as [114]:

$$n_{dof} = -3 + 2 \sum_{i=1}^{tacks} wi, \tag{5.1}$$

where w_i is the weight of track i and is calculated as:

$$w_i(\chi_i^2) = \frac{exp(-\chi_i^2/2T)}{exp(-\chi_i^2/2T) + exp(-\chi_c^2/2T)}, \tag{5.2}$$

with χ_i^2 the compatibility of the track i with the vertex, χ_c^2 is the threshold where the weight equals 0.5 and T is a shape parameter. For each event, the vertex with the highest $\sum p_T^2$ of the associated tracks is regarded as the primary interaction vertex. It is required to be within 24 cm in longitudinal and 2 cm in transverse direction from the nominal interaction point, and to have $n_{dof} > 4$.

5.2 Particle flow reconstruction algorithim

The default reconstruction technique used in the CMS experiment is the PF algorithm [112]. It combines the information from all the CMS sub-detectors [111] to identify and reconstruct individual objects. The particles that can be reliably resolved in the detector are identified as charged hadron, neutral hadron, muon, electron and photon candidates. An overview of these particles and their signatures are illustrated in Figure 5.1.

By taking advantage of the high granularity of the ECAL, a highly efficient tracking algorithm and a good clustering algorithm have been developed for the PF event reconstruction [111]. The particle flow algorithm works in two steps. First, pairs of elements are grouped into blocks by the linking algorithm. Second, the linked blocks are fed to a particle reconstruction and identification step that generates the final list of particles in the event. Figure 5.1 shows that muons are traversing the whole detector, leaving signals in the inner tracker and the muon system. The linking is performed between the tracks and the muon chambers. The typical energy deposit in the ECAL is 0.5 GeV and in the HCAL is 3 GeV. The reconstructed tracks are known as PF muon candidates. If the global fit between the inner track and the muon track results in an acceptable χ^2, the result is considered as a global muon and if muon track can be fit to a multiple tracks, the global muon with the lowest χ^2 is retained [112].

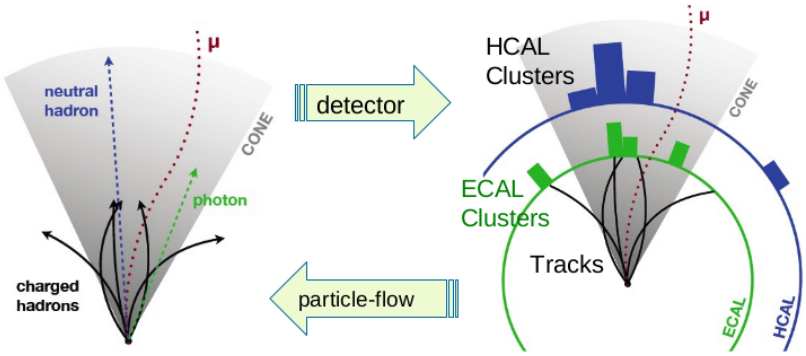


Figure 5.1: The concept of particle flow reconstruction [117].

Electrons are identified by their tracks in the tracker and an energy deposit in the ECAL. Whereas photons being neutral particles, don't leave tracks in the tracker and are identified by their energy deposits in the ECAL. The momentum of charged hadrons can accurately be measured in the tracker, and their energy is mainly deposited in the calorimeters. The neutral hadrons deposit their energy only in the calorimeters.

If the track momentum and the associated energy deposit in the calorimeters are compatible, the tracks are reconstructed as charged hadrons. If the energy of connected calorimeter cluster exceeds the track momentum significantly, the excess is attributed to an overlapping neutral hadron in the HCAL and to a photon in the ECAL.

5.3 Lepton Reconstruction

In this analysis, exactly one high energy electron or muon is considered. τ-lepton is not considered due to special constraints on its reconstruction.

Muons

Generally, muons are reconstructed using the track information from the silicon tracker and the muon systems, which are so-called global muons. Global muons are reconstructed by extrapolating the tracks [118] in both the systems with valid hits, at least 5 in tracker layers, one in pixel layer, and at least one in the muon systems. The track segments are reconstructed by a linear fit by using hits in each DT or CSC chamber seperately. Then RPC information is used to reconstruct muon trajectories using a Kalman Filter technique [119]. The reconstructed Particle Flow muon collection still contains significant amount of mis-identified charged hadrons and other particles punching through the calorimetric system. To purify the collection, the impact parameter |dB| with respect to the beam spot is required to be less than 0.02 cm and the distance to the primary vertex in z-direction needs to be smaller than 0.5 cm. In $t\bar{t}$ semi-leptonic decay, the muon from the W-boson decay is expected to be well-isolated from all other final state objects. To identify this type of muons, the particle flow isolation I with in a cone of radius R is defined as:

$$I(R) = \left(\sum_{ch(R)} p_T + max \left\{ \sum_{nh(R)} p_T + \sum_{\gamma(R)} p_T - \frac{1}{2} \sum_{chPU(R)} p_T, 0 \right\} \right) / p_T^{\ell}, \qquad (5.3)$$

where *ch(R)* represents all charged hadrons in a cone of radius R around the lepton direction, nh denotes all neutral hadrons, γ stands for the photons, $chPU$ for the charged hadrons coming from other vertices and p_T^{ℓ} represnts the transverse momentum of the lepton. The above expression reduces the effect of pileup events by assuming that the contribution from neutral pileup particles is approximately half of that from the charged pileup.

The resolution of muons is measured from the J/ψ and Z-boson resonances in the range $20 < p_T < 100$ GeV with a precision of 0.2%. The relative muon momentum resolution is in the range of 1.3–2% in the barrel and better than 5% in the endcap and has been validated in data. Further details of the muon reconstruction can be found in Ref. [120, 121].

Electrons

The procedure for the electron reconstruction is generally much more complicated than that of muon reconstruction. Whenever electrons travel through the tracker material in a strong magnetic field, they radiate a significant amount of their energy by radiating bremsstrahlung photons even before reaching the calorimeters. About 35% of all electrons looses upto 70% of their energy in the tracker and roughly 10% looses even upto 95% of their energy in the tracker material [122].

In the barrel, a supercluster is formed because the bremsstrahlung yields an additional energy deposit in ϕ direction. Whereas in the endcaps, the superluster is built from 5×5 clusters around the seed crystal in a range of ±0.07 rad in η and ±0.3 rad in ϕ. The energy deposits from the preshower are added in a range of up to ±0.45 rad in ϕ and a range of ±0.15 rad in η.

There are two procedures for electron reconstruction; tracker-driven seeding [123] and ECAL-driven seeding [124, 125]. The tracker-driven seeding provides good seeding efficiency for low p_T electrons and electrons within the jets [126]. Electron tracks require some additional reconstruction requirements due to non-Gaussian energy losses. The tracks which are not pre-identified as electrons, are rejected. The energy loses due to bremsstrahlung lead to a change in curvature as the electron travels into the outer-tracker region and the Kalman Filter track (KF) [118, 119] can not be fitted very well. The Gaussian Sum Filter (GSF) [127] is a non-linear generalization of the KF. The benefit of using GSF track is that it allows for a better extrapolation of the electron track to the ECAL and is able to consider the change in curvature. The GSF is also used to improve poorly fitted KF tracks.

The ECAL-driven seeding works best for electrons with p_T larger than 5 GeV/c. The energy-weighted average position of the calorimeter cluster is propagated through the magnetic field to the most inner layer of the pixel detector to identify the first hit. If this layer does not yield any hit, the next layer is looked to over-come pixel seeding in-efficiencies. This pre-identification is then performed by Boosted Decision Tree (BDT) [128] that uses information of the quality of the track-to-cluster matching, the shower shape, additional HCAL deposits, and the difference between the KF and GSF-fitted tracks. Further details on electron reconstruction can be found in Ref. [125].

5.4 Jet Reconstruction

The quarks and gluons produced in an event carry a color charge. QCD confinement dictates these fragments and hadronise them. A spray of hadrons, leptons and photons is observed in the detector, which is referred as a jet. Jets are experimental signatures of the partonic quarks and gluons. They are composed of charged hadrons, neutral hadrons, isolated and non-isolated particles. Their energies are corrected by applying different jet energy corrections for additional particles from pileup and detector effects. The jets stemming from the bottom quarks are identified by applying b-tag algorithms.

Jet Clustering

This study is based on semi-leptonic decay channel of $t\bar{t}$ pairs, at the leading order. In this decay mode, at least four jets are produced, two light jets from W-boson decay and two b-jets from the $t\bar{t}$ decay. To fully reconstruct the $t\bar{t}$ event topology, jet clustring algorithms are the most impotant tools. Several jet clustring algorithms are developed but at the CMS experiment the default algorithm used is the anti-k_T algorithm [129, 130] with $R = 0.5$. The main advantage of the anti-k_T is that it belongs to a class of infrared and collinear safe (IRC) jet algorithims. An infrared-safe clustering algorithm means that the resulting jets are unchanged by any soft gluon emission and it is collinear-safe when an input particle splits into two particles with nearly the same direction without changing the result. For the clustering algorithms following distance parameters can be used:

$$d_{ij} = min(p_{T,i}^{2k}, p_{T,j}^{2k})\frac{\Delta_{ij}^2}{R^2}, \tag{5.4}$$

$$d_{iB} = p_{T,i}^{2k}, \tag{5.5}$$

where d_{ij} is the distance between two particles i and j and d_{iB} is the distance between particle i and the beam. $\Delta_{ij}^2 = (y_i + y_j)^2 + (\phi_i + \phi_j)^2$ is the distance in

$\eta - \phi$ space, k steers the behaviour of the algorithm and R is the radius parameter of the algorithm. Two particles i and j are clustered together when $d_{ij} < d_{iB}$ and when $d_{iB} < d_{ij}$ particle i is considered as a jet and removed from further clustering. The clustering stops when no particles are left.

When parameter k = 1, the Durham k_T algorithm [131] is reproduced and when k = 0, the Equation 5.4 turns into the Cambridge-Aachen (CA) algorithim [132,133]. If k = -1, the Equation 5.4 yields the anti-kT algorithm. Different jet shapes produced by these algorithims are shown in Figure 5.2.

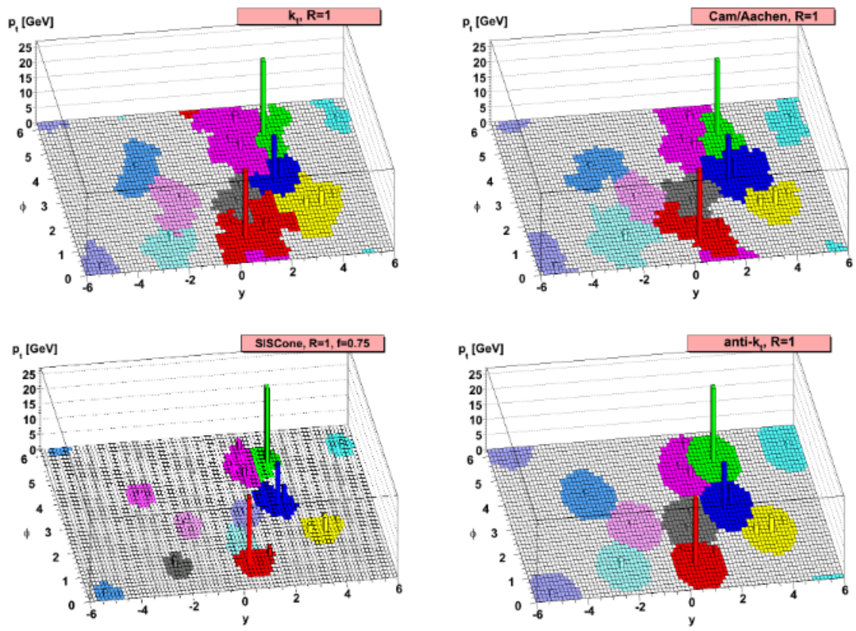


Figure 5.2: The graphical representation of the sensitivity of different jet cluster algorithms to soft radiation for the same event [130].

Charged particles originating from pile-up interactions and isolated leptons are excluded from clustering. Fake jets are suppressed by applying additional requirements on their energy. Only those jets are retained whose energy is purely attributed to either neutral hadrons or photons. Jet candidates are required to have an energy fraction between 0.0 and 0.99 from charged hadrons, an energy fraction smaller than 0.99 from both photons and neutral hadrons. Each jet is required to consist of at least two particles, where at least one is required to be a charged hadron, and a muon or an electron.

Jet Energy Corrections

There can be large discrepancies between the energy (or rather the transverse momentum) of the reconstructed particle-flow jets and true-particle jets [134]. A true-particle jet is defined as GenJet at the hadron level, accessible in simulated events, excluding neutrinos and pileup contributions. The Jet Energy Scale Corrections (JES or JEC) are calibrated to relate the energy of reconstructed jet to the energy of a corresponding GenJet. The corrections are factorized as shown in Figure 5.3.

The first correction (so-called *L1* correction) is applied to remove the electronic noise or pileup induced responses in the detector [134]. The pileup correction consists of rejecting tracks that stem from other vertices than the selected primary one. Therefore an absolute offset correction factor, C_{Offset} as [135, 136] is introduced,

$$p_T^{Corr} = p_T^{Uncorr} - C_{Offset} \tag{5.6}$$

The *L1* pileup offset correction factor depends on the jet transverse momentum p_T , the jet pseudorapidity η, the offset momentum density ρ calculated from k_T jets with $R = 0.6$ [134] and the effective jet area A. It can be found using an extension of the *hybrid jet area* method [129, 137].

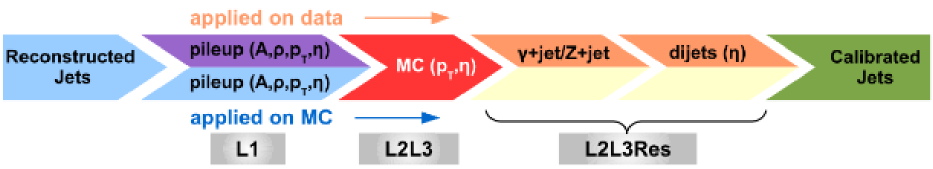


Figure 5.3: The steps involved in the CMS jet energy correction scheme [134].

After applying the *L1* offset correction, the response is independent of the primary vertex multiplicity but due to non-uniform and non-linear behaviour of the ECAL, we still need to apply p_T and η dependent corrections. In these corrections, named as *L2L3*, the energy of the reconstructed jet is calibrated to the true-particle jet using simulation. These corrections are verified in the data using Z+jet, γ+jet and dijet events.

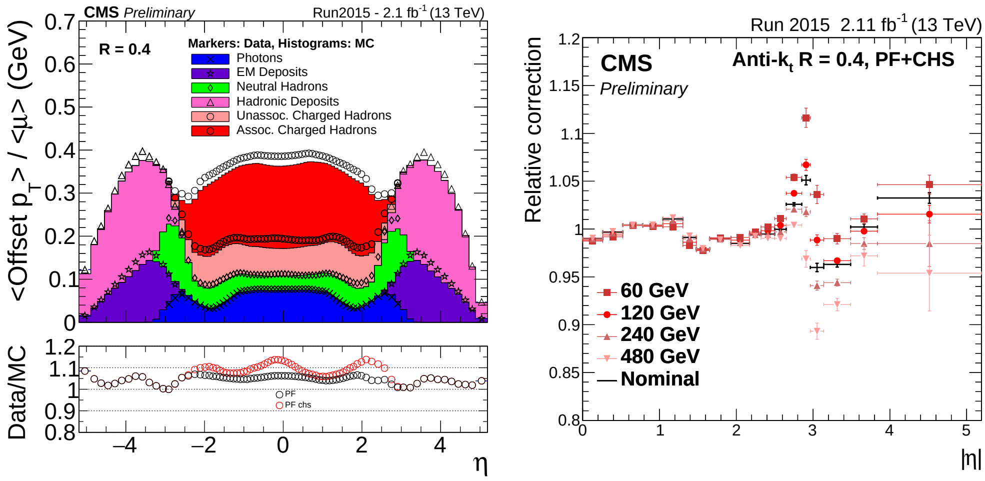


Figure 5.4: L1 pileup correction in dependence of the number of interaction vertices and the jet η (left), L2L3 MC truth correction in dependence of the jet p_T and η (right) [140].

In order to correct the data and monte carlo simulation differences, the residual/absolute energy corrections (*L2L3Res*) are applied to the data only. The relative residual correction (*L2 Residual*) is determined from dijet events using the Missing E_T Proection Fraction (MPF) method [138]. It corrects for the η-dependance ($|\eta| < 1.3$), binned in p_T . The absolute residual correction (*L3 Residual*) is determined by the p_T balance [139] and MPF methods in $30 < p_T < 700$ GeV from γ+jets and Z+ets events.

There are several other corrections available to calibratethe reconstructed jet energy. These corrections are not the default corrections for each physics analysis but are used to calibrate the reconstructed jet energy back to the parton level. The *L4* correction improves the jet resolution and ensures that the jet response is uniform with respect to the jet resolution. The *L5* corrections are determined relative to the default flavour-mixing and are therefore dependent on the sample i.e. QCD multijet or top quark pair events.

Jet Energy Resolution

After applying jet energy corrections, there is still slight difference in the width of the corrected jet p_T distributions between the data and simulation. As compared to the simulation, the resolution in the data is found to be worse by about 7–20% on average, and upto 40% atmost [134], depending on the pseudorapidity of the jet as shown in Figure 5.5. An η-dependent scale factor is applied to transverse momentum of the simulated jet [141] to obtain a jet energy resolution that improves agreement with the data.

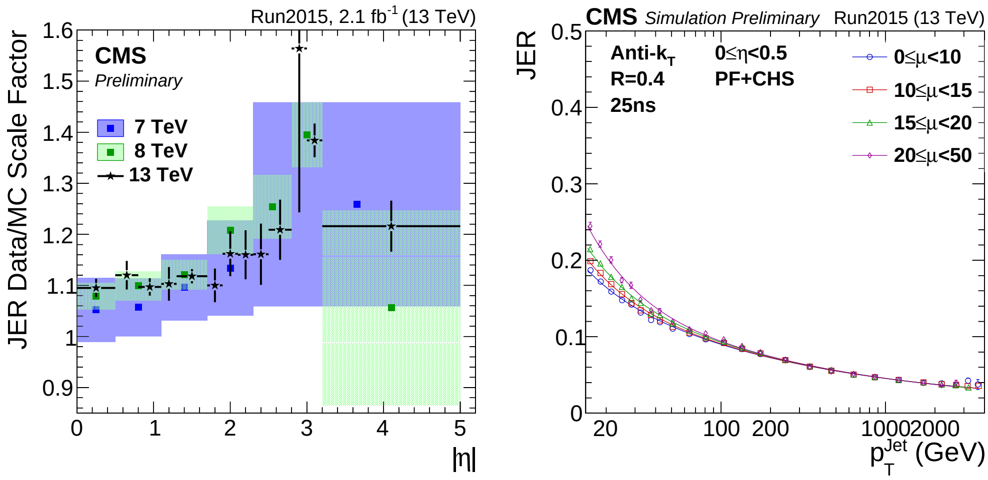


Figure 5.5: Corrections for the jet energy resolution in dependency of pseudorapidity η (left) and of transverse momentum p_T (right) [140].

The Jet Energy Resolution (JER) uncertainty sources include the p_T imbalance [142] caused by particle-level effects such as fragmentation, and an overall bias between the measurement of simulated and the true jet resolutions. The JER uncertainty varies from 3–5% in the tracker region.

B-tagging

The jets that are produced from the hadronization of b-quarks are called *b-jets*. The *b-jets* are produced in many physics processes like $t\bar{t}$ decay, the Higgs boson and various supersymmetric models [143]. Identification of *b-jets* is very critical feature for seperating signal from the background processes. In $t\bar{t}$ decay, two light jets are expected from a W-boson decay while two *b-jets* are expected from the two top quark decays. This is a unique characteristic of $t\bar{t}$ decay topology.

The CMS detector is able to resolve the secondary vertices of *B-hadron* decays using the pixel tracker. Several algorithms are developed to identify jets originating from a b-quark. The algorithm itself is called a tagger. These algorithms exploit the unique properties of *B-hadrons* i.e. a mass of B-hadron ($m_B \approx 5$ GeV),

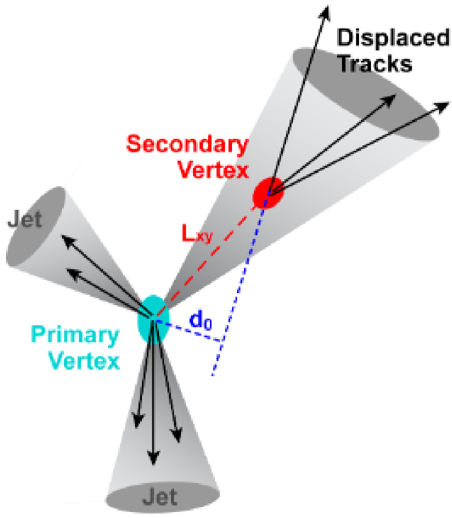


Figure 5.6: Secondary vertex displaced with a distance of L_{xy} in x-y plane with respect to the primary vertex [144].

lifetime $\tau_B \approx 10^{-12}$ sec, the flight distance ($\gamma_B \beta_B \tau_B \approx 7$ mm) with the relativistic Lorentz factor ($\beta_B \gamma_B \approx 0.4\frac{m_t}{m_B}$) to identify quarks coming from the decay of B-hadron [145]. Typical observables used for *b-tagging* are the flight distance (or its significance) to the secondary vertices, or reconstructed invariant mass of all charged particles associated to the secondary vertex, which is of order $mB/2$ [146].

The fraction of reconstructed jets that are identified as *b-jets* is called *b-tagging efficiency*, while the fraction of wrongly identified light jets as *b-jets*, is referred as *mistag probability*. There are usually three different working points for these algorithms that require different minimal discriminator values. The loose working point has 10% mistag probability and 80% of *b-tagging efficiency*, medium working point having 1% mistag uncertainty and 65% *b-tagging efficiency*, and the tight working point has 0.1% mistag uncertainty with 50% b-tagging efficiency [145]. Thus a tighter working point corresponds to a lower b-tagging efficiency, but also a reduced probability of mistakenly tagging a light jet as a *b-jet*. The Track Counting High Purity (TCHP), the Track Counting High Efficiency (TCHE), Jet Probability (JP) and Combined Secondary Vertex (CSV) are available algorithms for $\sqrt{s} = 13$ TeV analyses in the CMS collboration. A complete description of these algorithims can be found in [143].

5.5 Missing Transverse Momentum

In the semi-leptonic $t\bar{t}$ decay, one of the W-bosons decays into a lepton (electron or muon) and its associated neutrinos and other decays into two light jets. The neutrinos interact only weakly and leave the detector unnoticed. Using conservation of momentum of all other objects in the transverse plane, the neutrino energy can be estimated [147]. The imbalance

$$E_T^{\vec{}miss} = - \sum_{i\epsilon particles} p_T^{\vec{}}, i \tag{5.7}$$

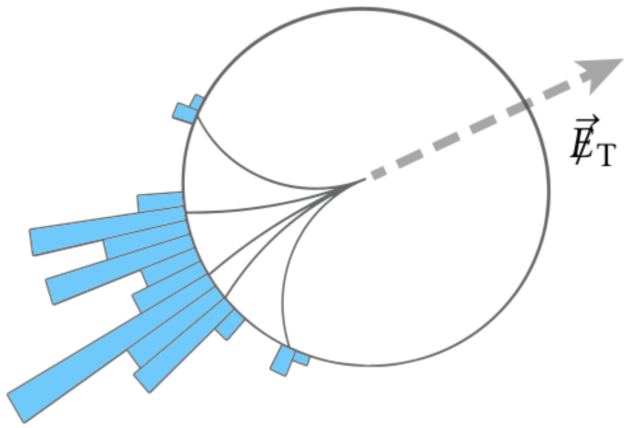


Figure 5.7: Schematic diagram of missing transverse energy [148].

is called the *missing transverse momentum* or *energy* ($\not{E}_T$) [149]. The moments of the uncalibrated jets in the above expression are replaced with the momenta of the corresponding calibrated jets if their p_T is larger than 10 GeV. Since jets are used to calculate the $\not{E}_T$, the corrections to the jet energy scale (so-called Type-I corrections) play an important role in its determination.

As $\not{E}_T$ is one of the most important signature in many searches beyond the standard model, such as supersymmetry and extradimensions, therefore, the $\not{E}_T$ performance of the CMS has been carefully validated [150]. In this analysis, no selection requirements are imposed on $\not{E}_T$ since other selection criteria already yield a very clean sample of $t\bar{t}$ events.

Chapter 6

Event selection

In previous chapter, the object reconstruction and identification criteria used in the CMS experiment were discussed. The purpose of this identification criteria is to separate a real reconstructed object like a muon or an electron from charged hadrons that are misidentified. In this chapter, the selection criteria to select only those events that are likely to contain $pp \rightarrow t\bar{t}$ will be discussed. An additional criteria is also used to obtain a very pure $t\bar{t}$ sample (electron+jets or muon+jets) with a large enough signal efficiency.

Section 6.1 describes the datasets and corresponding simulations used in this study. Section 6.2 is dedicated to event selection and selection efficiencies while in section 6.3, stability of the selection criteria during the whole data taking period is discussed. Finally, the results of event selection are shown in section 6.5.

6.1 Datasets

The 2015 CMS data analyzed in this study corresponds to an integrated luminosity of 2.2 fb^{-1} [151] for both the electron and muon channels. The datasets are listed in Table 6.1.

Monte Carlo Simulations

The data is compared to Monte Carlo (MC) simulated events. The list of simulated

Table 6.1: The first version (v1) of primary datasets corresponding to 2015 datataking periods C and D, for pp collisions at 13 TeV with a proton bunch spacing of 25 ns, for muon and electron channels.

Primary dataset	Integrated Luminosity
SingleMuon/Run2015C_25ns-16Dec2015-v1	2.2 fb^{-1}
SingleElectron/Run2015C_25ns-16Dec2015-v1	
SingleMuon/Run2015D-16Dec2015-v1	
SingleElectron/Run2015D-16Dec2015-v1	

events along with cross section [152–154] can be found in Table 6.2. At next-to-next-to-leading-order (NNLO) the expected $t\bar{t}$ cross section is 832^{+20}_{-29} $(scale)$ $\pm$ 35 $(PDF{+}\alpha_S)$ pb [155]. In this study, the NNLO is used as reference to normalize all the $t\bar{t}$ samples.

As MC samples are also partially used to predict the main backgrounds, therefore, they are well-tuned with the data [156]. The single top processes are normalised with respect to approximate NNLO prediction [157], W+jets and Drell-Yan (DY) to the NNLO prediction computed with FEWZ [158] and di-bosons to the NLO prediction predicted by MCFM [159, 160]. The $t\bar{t}$ signal is modeled with POWHEG v2 [83, 84] and matched to PYTHIA v8.205 [161, 162]. The NNPDF3.0 NLO [163] parton distribution functions (PDF) and the CUETP8M1 [77,164] underlying event tune are used in the MC samples. To evaluate systematic uncertainties, the simulations with the QCD scale choice at parton shower level is varied by a factor of 2 or 1/2. The two matrix element generators, MG5_aMC@NLO v5_2.2.2 [165] and MadSpin [166], are compared for both the PYTHIA 8 and HERWIG++ v2.7.1 [167] hadronizers. In addition, the samples used for systematic uncertainties are listed in Table 6.3. The $t\bar{t}$ cross section is measured in this analysis, and the theoratical value for m_t = 172.5 GeV, computed at NNLO+NNLL (next-to-next leading log): $\sigma_{t\bar{t}}^{\text{NNLO+NNLL}} = 832^{+20}_{-29}\,(\text{scale}) \pm 35\,(\text{PDF} + \alpha_s)\,\text{pb}$ computed with TOP++ [168] is used as a reference. The W+jets process is simulated at NLO with MG5_aMC@NLO. A higher statistics Born-level MadGraph v5.1.3.30 [74] including up to four extra partons emitted at matrix element level is used for the derivation of the shapes.

All the simulations include an emulation of response for the CMS detector using GEANT4 v.9.4p03 [169]. The real data effects induced by multiple pp collisions are considered for the same and neighbouring beam crossings (within 25 ns) and added to the generated hard interaction for pile-up effects.

6.1.1 Corrections applied to the simulation

Different types of corrections are applied to the simulated events, which are briefly described below.

Pileup re-weighting

All generated samples are passed through a full simulation of the CMS detector based on GEANT 4 [89] and include additional minimum bias interactions superimposed in order to account the effect of in-time and out-of-time pileup in the events. At generator level, the number of extra interactions are reweighted to match the one estimated in the data. A minimum bias cross section of 69 mb is used to estimate the pileup distribution following the standard recommendations [170]. A 5% uncertainty assigned to the minimum bias cross section is assumed. Figure 6.1 shows the distribution of the data compared to expectations in the number of primary vertices reconstructed in events, showing a fair agreement.

Trigger efficiency

The differences in performance for the trigger paths used in the data and simulation are corrected by applying recommended p_T and η-dependent scale factors [171, 172] which are also cross-checked for this study (see section 6.2.2).

Lepton identification and isolation efficiency

The difference in performance for the lepton identification and isolation between the data and simulation is also corrected by applying p_T and η-dependent scale factors. The scale factors are derived using the tag-and-probe method, (section 6.2.2) and also cross-checked with the recommended values [171,172]. The leptons in this study are identified on the basis of tight muons [173] and tight electrons [174] criterion.

Table 6.2: The simulation samples along with cross sections (σ) [152–154], corresponding to $t\bar{t}$ signal and different background processes, used in this study.

Process	Dataset	$\sigma[pb]$
Signal		
$t\bar{t}$	TT_TuneCUETP8M1_13TeV-powheg-pythia8	832
Background		
$t\bar{t}$ +W	TTW$\rightarrow$ Lν_L	0.2
	TTW$\rightarrow$ QQ	0.4
$t\bar{t}$ +Z	TTZ$\rightarrow$ QQ	0.5
	TTZ$\rightarrow$ LL$\nu_L\nu_L$	0.2
WZ	-	47.1
WW	WW$\rightarrow$ Lν_LQQ	50.0
	WW$\rightarrow$ LL$\nu_L\nu_L$	12.2
W+jets	W$\rightarrow$ Lν_L	61527.0
tW	ST-tW $\rightarrow$ anti-top	36.0
	ST-tW $\rightarrow$ top	36.0
t-ch	ST-t $\rightarrow$ top	44.1
	ST-t $\rightarrow$ anti-top	26.2
DY	DYJets $\rightarrow$ LL M-10to50	18610.0
	DYJets $\rightarrow$ LL M-50	6025.0

Table 6.3: Simulation samples for different processes and different dataset parameters, such as TT (top pair) and ST_tW (single top, tW-channel) used for systematics.

Systematic uncertainty	Dataset parameters
Parton shower scale	powheg-scaledown (TT)
	powheg-scaleup (TT)
m_t	$m_t = 169.5$ GeV (TT)
	$m_t = 169.5$ GeV (ST_tW_top)
	$m_t = 169.5$ GeV (ST_tW_anti-top)
	$m_t = 178.5$ GeV (TT)
	$m_t = 178.5$ GeV (ST_tW_top)
	$m_t = 178.5$ GeV (ST_tW_anti-top)
Generator	amcatnloFXFX-pythia8 (TT)
Hadronization	powheg-herwigpp (TT)
tW QCD scale	scaledown (ST_tW_top)
	scaledown (ST_tW_anti-top)
	scaleup (ST_tW_top)
	scaleup (ST_tW_anti-top)
tW/$t\bar{t}$ interference	ST_tW_top
	ST_tW_anti-top
W+jets generator	W$\rightarrow$ Lν_L

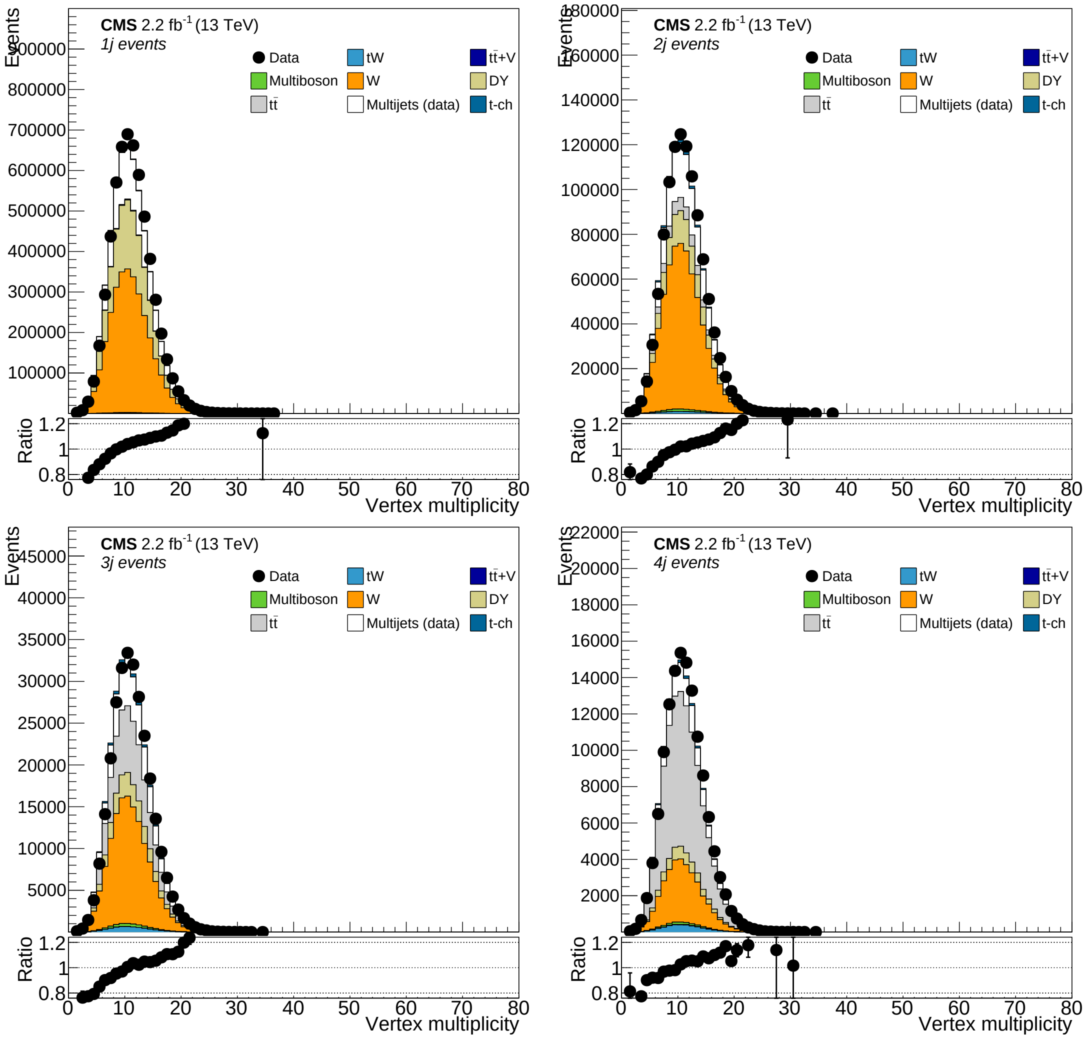


Figure 6.1: Distributions of the number of primary vertices in the combined ℓ+jets events. From left to right events with 1, 2, 3 or at least 4 jets are shown.

Jet energy scale and resolution

Recommended jet energy scale corrections are applied [175]. In simulation the nominal jet energy resolution is smeared using a p_T and η-dependent parameterization [176]. In both cases alternative scenarios generated by shifting the corrections according to their uncertainties are considered and shall be discussed in detail in section 7.1.

b-tagging efficiency

The difference in performance of the b-tagging algorithm used in the analysis is accounted for by applying a p_T dependent scale factor. The scale factors are used to correct a-posteriori the b-tagging decision in the simulation [177]. The medium working point of the Combined Secondary Vertex (v2) algorithm is used (section 6.2.1).

Generator level weights

The simulated processes are scaled according to their effective integrated luminosity which can be computed from the sum of the per-event weights. The general formula applied to determine the number of events expected for a given process ($\hat{N}$), is:

$$\hat{N} = \mathscr{L} \cdot \sigma \cdot \frac{\sum_{i=1}^{\mathrm{N}_{sel}} w_i}{\sum_{i=1}^{\mathrm{N}_{gen}} w_i} \quad , \tag{6.1}$$

where $\mathscr{L}$ is the integrated luminosity, σ is a theory predicted inclusive cross section and w_i are the per-event generator level weights.

6.2 Event selection

The event selection is based on the decay topology of $t\bar{t}$, where one of the W bosons decays to either muon or electron while the other decays in jets. The particle flow PF algorithm is used for reconstruction of final state objects [113]. This algorithm combines signals from all sub-detectors to enhance the reconstruction performance and allows to identify muons, electrons, photons, charged hadrons and neutral hadrons produced in a *pp* collision.

6.2.1 Object selection

Trigger and Vertex

As a first step, events are required to pass a trigger selection criteria. The trigger path stores events during the online event reconstruction; atleast one isolated muon (electron) with a transverse momentum of 22 GeV (30 GeV) within pseudo-rapidity of $|\eta| < 2.1$ for this analysis. The data selected by this trigger is the part of SingleMuon (SingleElectron) primary dataset having events with at least one muon (electron). At least one primary vertex (PV) is required to be reconstructed from at least four tracks, with longitudinal (radial) distance of less than 24 cm from the center of the detector. Among all, the primary vertices in the event with the largest scalar sum of associated particle transverse momenta are selected.

Muons

Muons are reconstructed using the track information from the silicon tracker and the muon system. In this analysis, the reconstruction of a track in the muon system that is compatible with a silicon tracker track is considered. In order to increase the purity of muons originating from the primary interaction and to suppress misidentified muons or muons from the decay-in-flight process, additional selection criteria are applied [173]. The first candidate muon passing the tight identification (tight ID) criteria is additionally required to have $p_T > 30$ GeV and $|\eta| < 2.1$. Then no additional muon, passing the medium ID with $p_T > 15$ GeV and $|\eta| < 2.5$ must be found in the event. Muon isolation is based on the sum of the p_T of the charged and neutral hadrons as well as photons reconstructed in a cone $R = 0.4$ around the muon momentum. The sum of the p_T of the charged hadrons associated to other vertices than the PV, is used to correct the pileup contamination in the total flux of neutrals, found in the muon isolation cone. A factor of $\beta = 0.5$ is used to scale this contribution as,

$$I_{\mu} = \frac{1}{p_{\mathrm{T}}} \cdot \sum_{R<0.4} \left[p_{\mathrm{T}}^{ch} + \max(p_{\mathrm{T}}^{\gamma} + p_{\mathrm{T}}^{nh} - 0.5 \cdot p_{\mathrm{T}}^{pu\ ch}, 0) \right] \tag{6.2}$$

where p_{T}^{ch}, stands for the transverse energy deposition of charged hadrons in a cone size of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ across the direction of muon. Similarly, p_{T}^{γ} and p_{T}^{nh} are the energy deposits of photons and neutral hadrons in the same cone size. $p_{\mathrm{T}}^{pu\ ch}$ is the transverse momentum of tracks which are associated with non leading vertices. In order to estimate the contribution of neutral particles from pile-up events, a multiplicative factor of 0.5 is applied to $p_{\mathrm{T}}^{pu\ ch}$. A tight muon must verify $I_{\mu} < 0.15$ while a medium muon must verify $I_{\mu} < 0.25$. For tight muons, the significance of the 3D impact parameter (SIP3d), rejecting events for which the muon has SIP3d>4, is used. This effectively reduces the contamination from heavy-flavour leptonic decays while keeping a high efficiency for the signal. A similar cut is used in $t\bar{t}$ +V analyses [178, 179].

Electrons

The electron candidates are reconstructed from a combination of the track momentum at the main interaction vertex and the corresponding energy deposition in the ECAL. This analysis uses the electron selection criteria specified in Ref. [174]. The candidate electron must pass the tight identification and isolation requirements in addition to $p_T > 30$ GeV and $|\eta| < 2.1$. Electrons reconstructed in the ECAL barrel to endcap transition ($1.4442 < |\eta| < 1.5660$) are not considered in the analysis. Similar to the muon case, the first leading electron is also required to pass tight ID and SIP3d>4. No additional electron, passing the loose identification and isolation requirements with $p_T > 15$ GeV and $|\eta| < 2.5$ must be found in the event.

Jets

In this analysis, jets are reconstructed by clustering the particleflow candidates using the anti-k_T clustring algorithm [129, 130] with a distance parameter of 0.4. The charged particle candidates closer along the z-axis to any vertex other than the main primary vertex, are ignored. Pile-up interaction energy in the jet cone, is extrapolated by energy flow in the detector which is neither clustered in jets nor reconstructed as leptons and resulting jet energy is updated. Jets that originate from the hadronization of b-quark are identified by a combined secondary-vertex algorithm (CSV). This b-tagging discriminant combines information from secondary vertices with information from track-based lifetime estimations. The medium working point is used [180].

6.2.2 Lepton selection efficiency

The lepton selection efficiency is measured in the data and compared to the expectations from simulation, using $Z \to \ell\ell$ events. The results are used to derive a data-to-simulation scale factor, which is then applied on an event-by-event basis to correct the MC expectations, as described in section 6.1.1. These scale factor measurements are further used to evaluate possible alternative selections which may increase the purity of the sample. In addition, it is interesting to test these measurements in a high jet multiplicity environment such as the one expected in a $t\bar{t}$ event.

Muon efficiencies

By using tag-and-probe method for muons, the tagged muon is required to be matched with the trigger object and it should be reconstructed offline with $p_T > 25$ GeV, $|\eta| < 2.5$ and $I_{rel}(\Delta\beta, R = 0.4) < 0.2$. The fits are perfomed using a Voigtian function [181] for signal and Chebyshev polynomials [182] for the background. The results obtained with these functional form have also been compared to those using a sum of Crystal-ball functions for the signal and the same background model. The results are found to be fully compatible with each other. Other alternatives have also been tried, but the the two above methods describe better the Z-line shape in the data and MC, in both the pass and fail categories used for the extraction of the efficiency.

Appendix-A shows the fit results for trigger efficiency for different jet multiplicies in the data and simulation correspondingly.

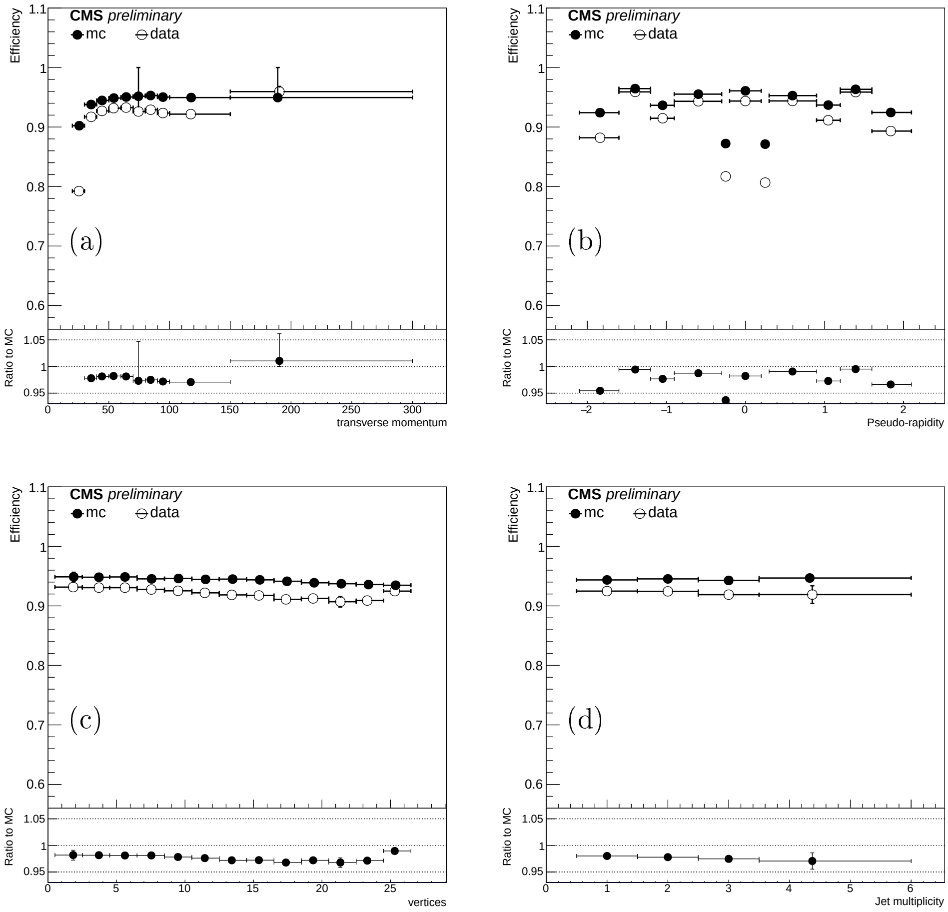


Figure 6.2: Muon trigger efficiencies as function of (a) muon p_T , (b) muon η, (c) vertex multiplicity and (d) jet multiplicity. The top panels display the measured (expected efficiencies) while the bottom panels display the data-to-simulation scale factors. The error bars represent the statistical uncertainty.

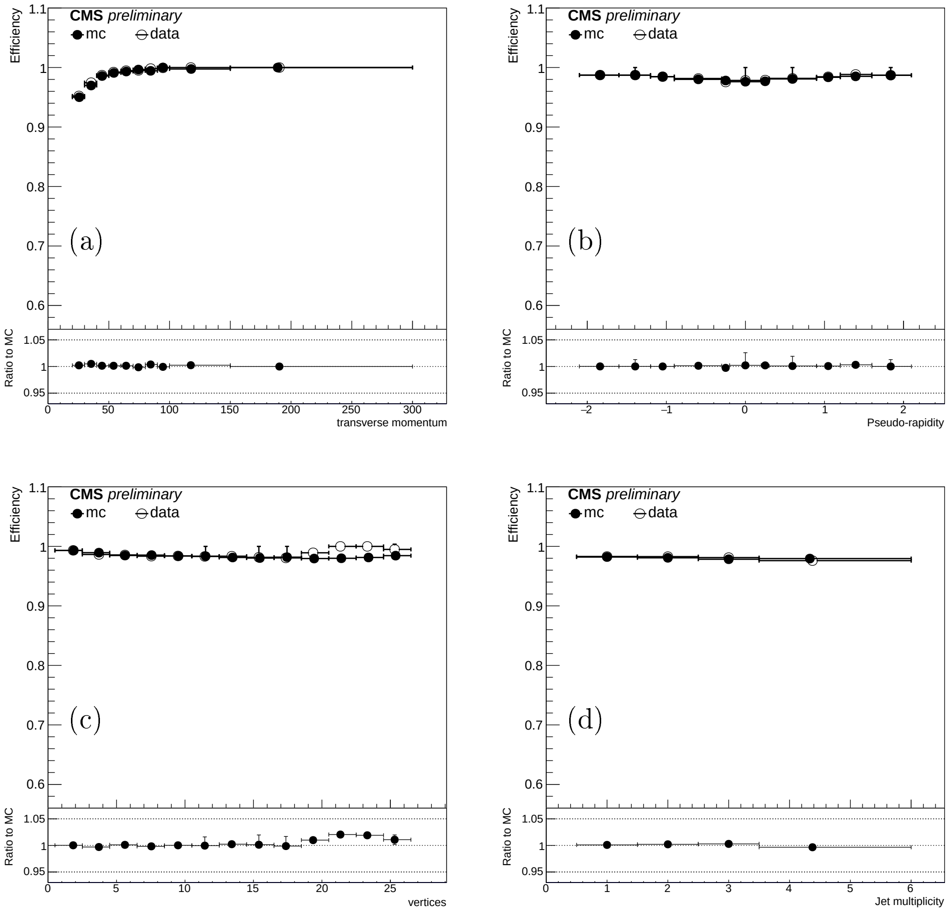


Figure 6.3: Muon identification and isolation efficiencies as function of (a) muon p_T , (b) muon η, (c) vertex multiplicity and (d) jet multiplicity. The top panels display the measured (expected efficiencies) while the bottom panels display the data-to-simulation scale factors. The error bars represent the statistical uncertainty.

Figure 6.2 shows the results obtained for the muon trigger efficiency as a function of different variables. A flat dependency is observed for the efficiency scale factors as a function of p_T and the vertex multiplicity, and the main variations are observed in efficiency as a function of η. The behavior as a function of the vertex multiplicity is found to be stable, within 2%, which is assigned as an uncertainty. The behavior of the identification and isolation criteria is significantly more stable, as shown in Figure 6.3. A cross check has been performed for a tighter isolation working point $I_\mu < 0.06$, which could potentially benefit the analysis in terms of multijets rejection. However this tighter selection induces a dependency on the jet multiplicity, which is avoided.

Electron efficiencies

Similar to muons, the tagged electron is required to be matched to the trigger object and the offline medium working point criteria, having $p_T > 25$ GeV, $|\eta(\mathrm{S}C)| < 2.5$ (excluding the barrel to endcap transition). The fits are perfomed using a Voigtian function for the signal and a Chebychev polynomial for the background.

Figure 6.4 reports the results obtained for the electron trigger efficiency as function of different variables. A flat dependency of the efficiency scale factor is observed. The behavior of the identification and isolation criteria is significantly more stable as shown in Figure 6.5.

6.3 Event Rate Cross Check

The first step towards measuring the cross section is verifying how stable the event rate is throughout the data taking period. This will ultimately determine which data runs should be considered to attain maximum precision. The rate of selected events (trigger+offline) is measured on a run-by-run level and divided by the estimated integrated luminosity of that run. The median of the rate observed after each jet multiplicity requirement is reported in Fig. 6.6 (left). The asymmetry

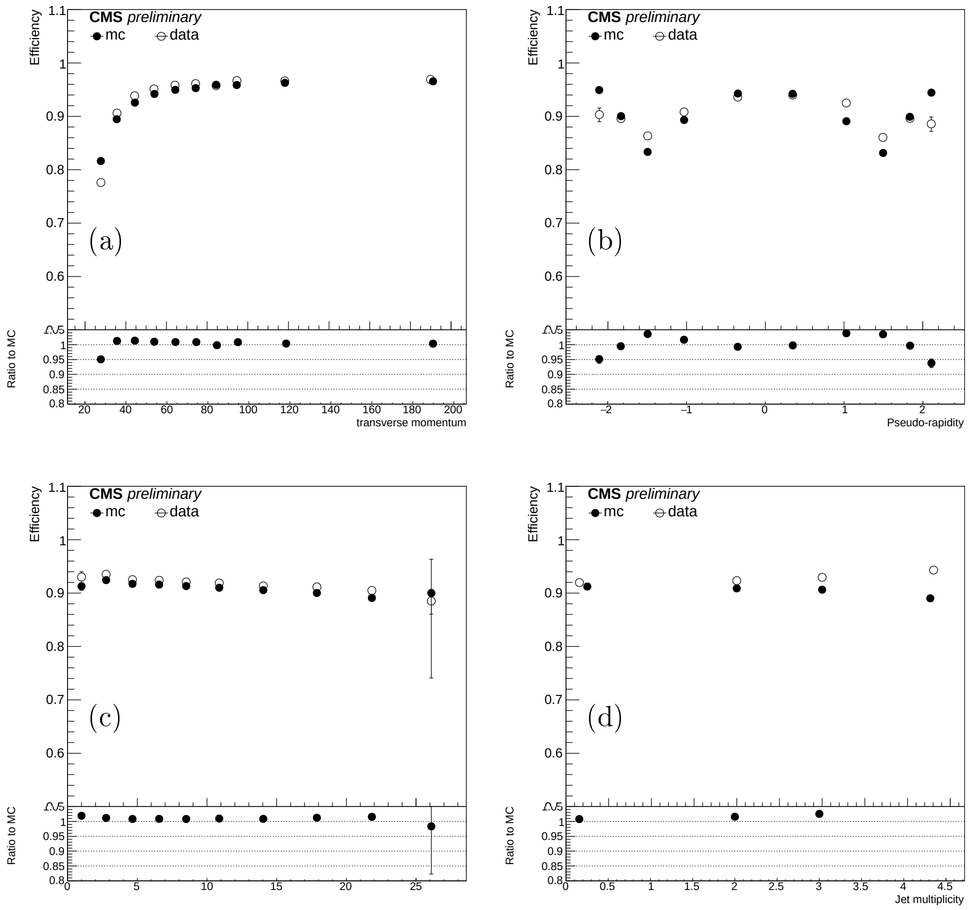


Figure 6.4: Electron trigger efficiencies as function of (a) electron p_T , (b) electron η, (c) vertex multiplicity and (d) jet multiplicity. The top panels display the measured (expected efficiencies) while the bottom panels display the data-to-simulation scale factors. The inner error bars represent the statistical uncertainty. The outer error bars report the differences obtained with respect to the central value by using a different fit model.

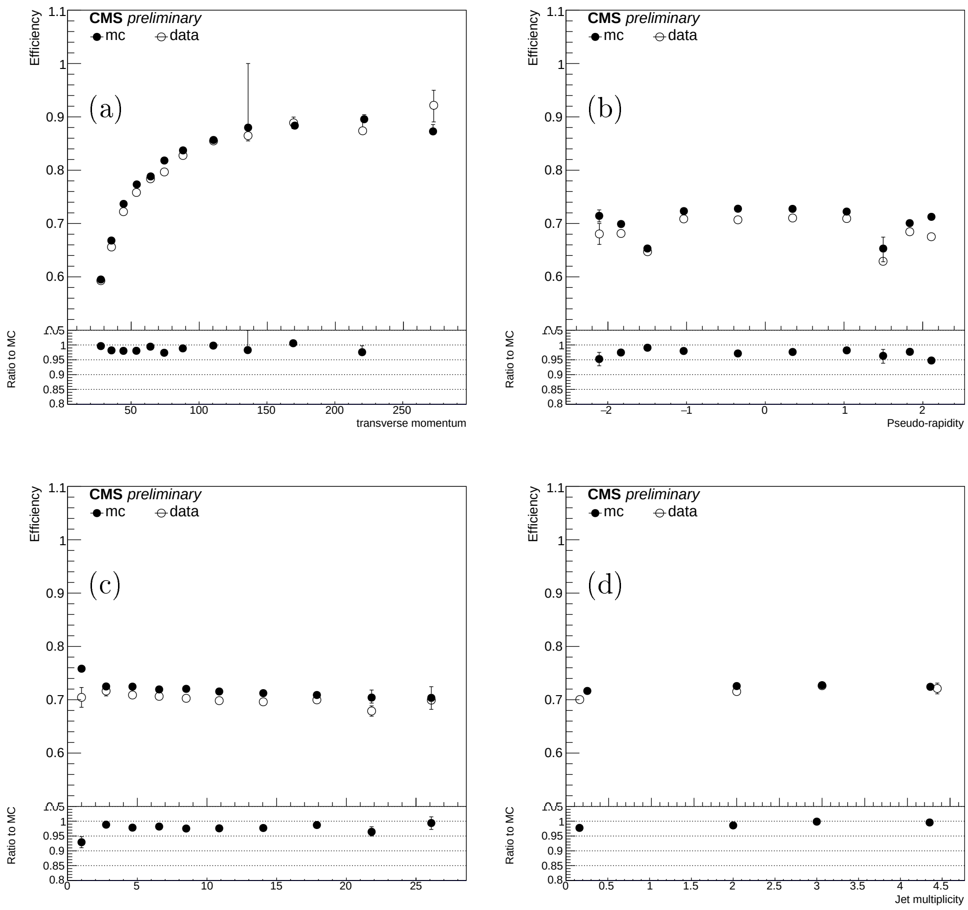


Figure 6.5: Electron identification and isolation efficiencies as function of (a) electron p_T , (b) electron η, (c) vertex multiplicity and (d) jet multiplicity. The top panels display the measured (expected efficiencies) while the bottom panels display the data-to-simulation scale factors. The error bars represent the statistical uncertainty.

in yields in the +/- channel tends to be diluted with more jets in the event as expected. The rates between electrons and muons differ slightly due to different background composition and tighter selection applied on electrons. Fig. 6.6 (right) shows the relative rate variation at 68% and 90% confidence levels, defined from intervals corresponding to the 68% and 90% quantiles of the rates, divided by the median of the rate. For events with more than 2 jets, the dispersion starts being too large: the cross section starts competing with the integrated luminosity of the runs leading to a very large dispersion, in particular for shorter runs. However, the relative rate variation for events with one or two jets is very instructive. At 68% CL, the dispersion is estimated to be at the level of 3% in both the electron and muon channels, positively or negatively charged. This indicates the ball-park of the limit reachable in precision in this dataset with the single lepton final states, due to the uncertainty in the integrated luminosity, trigger and offline selection efficiencies.

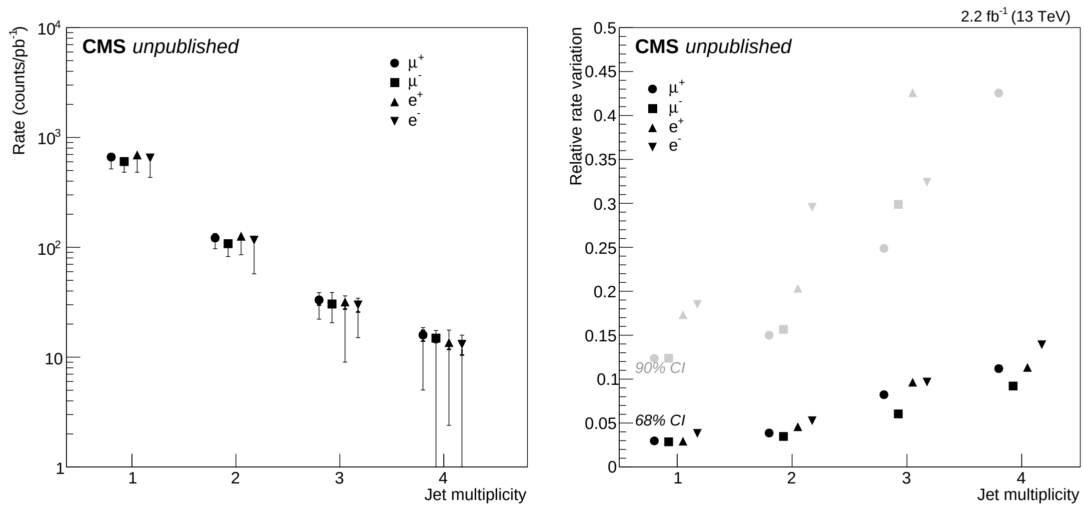


Figure 6.6: Left: Median rate per run in the 2015 dataset for different lepton flavours and charges and as function of the number of jets required in the event. The error bars indicate the 68% and 90% confidence intervals derived from the quantiles of the distribution of rates in each sample. Right: Relative variaton of the rate per run at 68% and 90% confidence levels as function of the number of jets required in the event.

Figures 6.7 and (6.8) show the rate of events in the μ^+ and μ^- (e^+ and e^-) samples correspondingly.

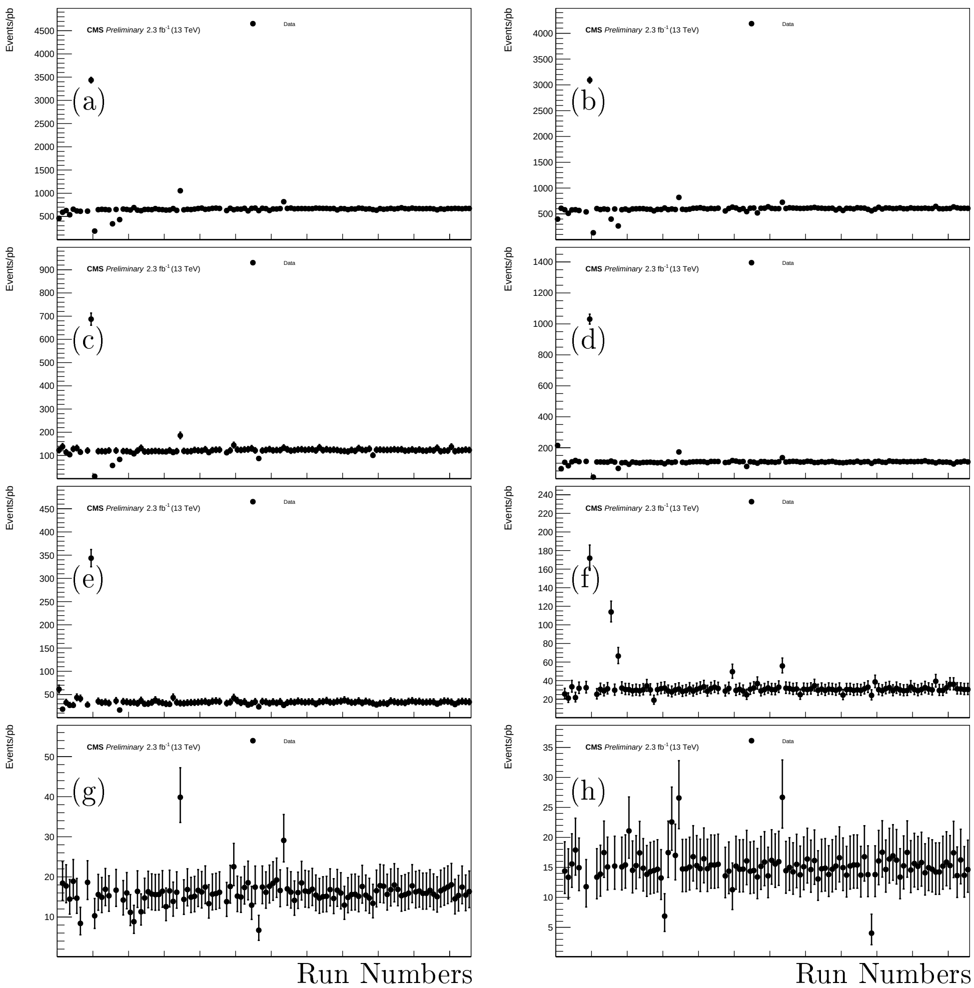


Figure 6.7: Observed rate per run after the selection of one positively (left) or negatively (right) charged muon and 1 (a/b), 2 (c/d), 3 (e/f) or at least 4 (g/h) jets. Run numbers are ploted along x-axis.

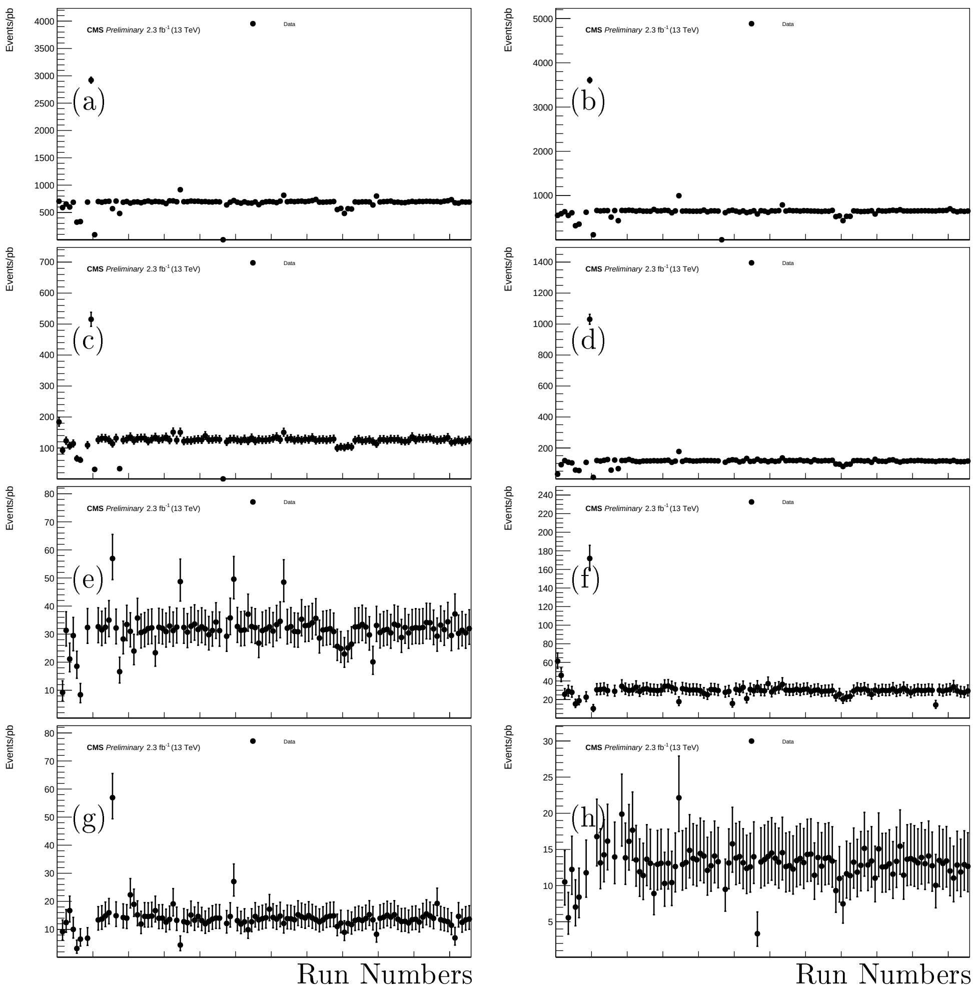


Figure 6.8: Observed rate per run after the selection of one positively (left) or negatively (right) charged electron and 1 (a/b), 2 (c/d), 3 (e/f) or at least 4 (g/h) jets. Run numbers are ploted along x-axis.

6.4 Signal discriminating variables

The main characteristics expected of the signal are the following:

High jet and b jet multiplicity

Two jets are expected from a W-boson decay while two b jets are expected from the two top quark decays, which is a unique characteristic of $t\bar{t}$. Low jet multiplicities (less than 4 jets) and low b-jet multiplcities (less than 2 b-tag jets) are expected for all the backgrounds, making the categorization of the analysis in these variables relevant to constrain the main backgrounds while isolating the signal.

Production of a neutrino in the final state through a $W \rightarrow \ell\nu_{\ell}$ decay

The emitted neutrino will generate genuine $\not{E}_T$ in the event, in constrast to QCD multijet background. Therefore, this distribution can be to constrain the normalization of these backgrounds in events with low b jet multiplicity.

Resonant production of a b and isolated lepton from a $t \rightarrow Wb$ decay

If correctly paired to the top quark decay, the $b + \ell$ system is expected to have a characteristic distribution with an endpoint which is related to the W and t masses, as $\max M_{\ell,b} = \sqrt{m_t^2 - m_W^2}$. This distribution can be used to further isolate the signal contribution while exploring the sensitivity to the top quark mass at the same time [183].

The prefit distributions for the $M(\ell, b)$ and $\min M(\ell, b)$ are shown in Figures 6.10 and 6.11. The $M(\ell, b)$ variable is computed from events having at least one b-tagged jet. In case of more than one b-tagged jet, the two leading tagged jets are considered and the $\ell + b$ system with lowest invariant mass, which is defined as $\min M(\ell, b)$, is used. A fair agreement is observed for all categories of jets. In some case the QCD estimate tends to over predict pairs at very low mass, this is however mostly covered by the normalization of the QCD multijets.

6.5 Pre-fit distributions

Appendix C contains kinematic distributions of the selected objects. Figures 6.9 and 6.10 show the inclusive distributions prior to the fit. A moderate excess is visible for events with 3 jets and 2 b tags. As it will become clear later (section 8.3), this slight excess is expected to be within 1σ of the total uncertainty.

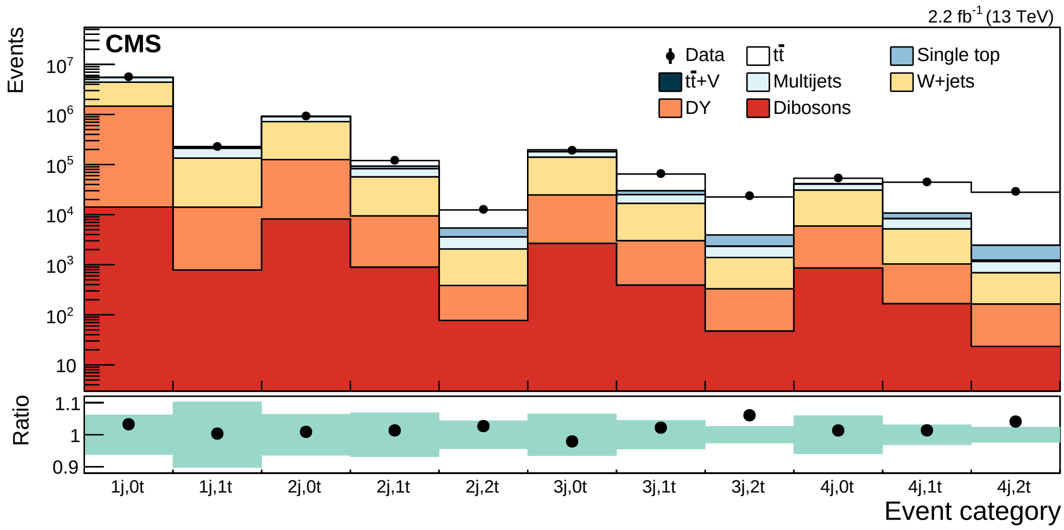


Figure 6.9: Event yields from data and the expected $t\bar{t}$ signal and backgrounds for each of the 11 independent categories, depending on combination of light and b-jets. Distributions are combined for the two lepton charges and flavors. The bins represent the measured number of jets (j) and b-tagged jets (b). The bottom panel shows the ratio between the data and the expectations. The relative uncertainty owing to the statistical uncertainty in the simulations, the systematic uncertainty in the total integrated luminosity and in the normalization of the multijet background is represented as a shaded band.

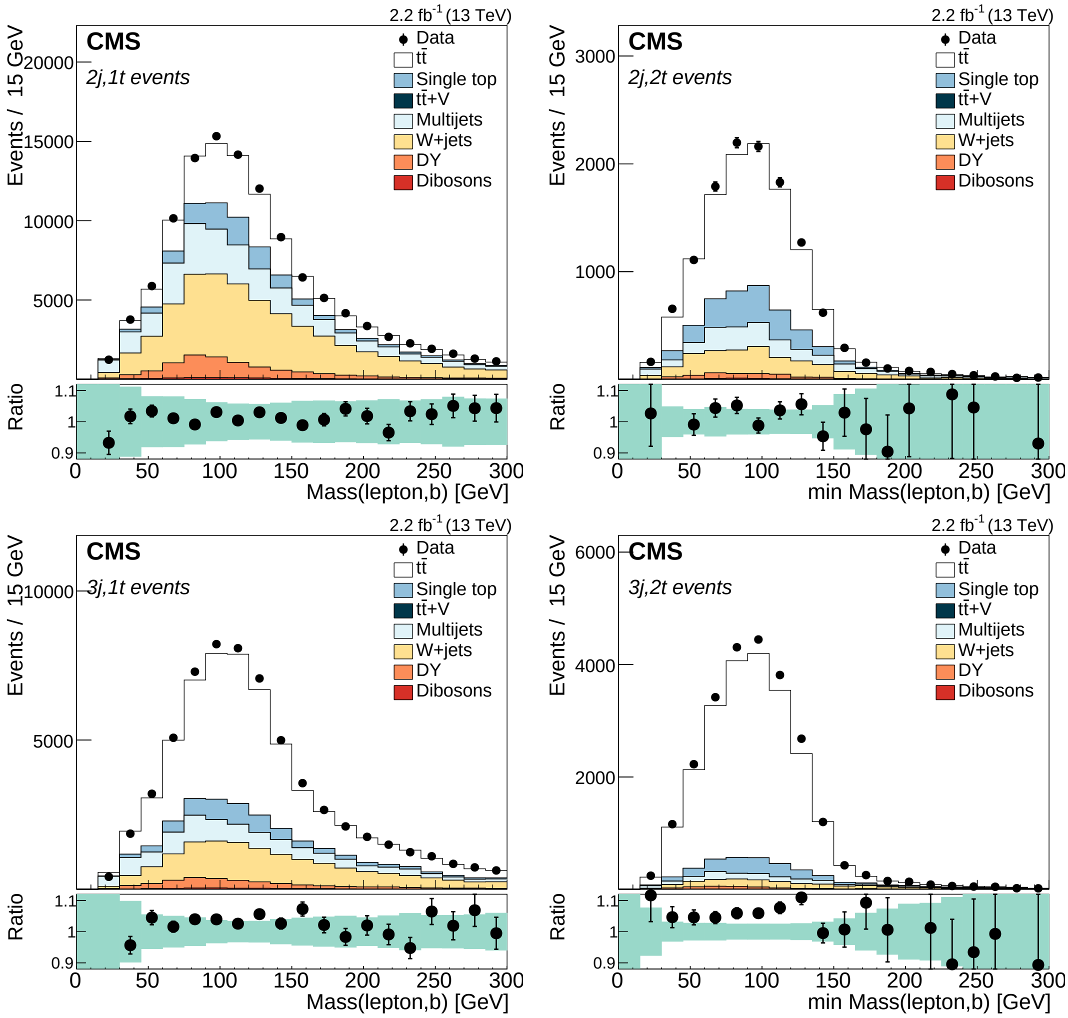


Figure 6.10: Distributions in the observables used to fit the data with the contributions from all leptons and charges combined. Panels on the left show the distributions in $M(\ell, b)$, and on the right in $\min M(\ell, b)$, for events with one and two b-tagged jets, respectively. From top to bottom, the events correspond to those with 2 and 3 jets. The lower plot in each panel shows the ratio between the data and expectations. The relative uncertainty owing to the statistical uncertainty in the simulations, the systematic uncertainty in the total integrated luminosity and in the normalization of the multijet background is represented as a shaded band.

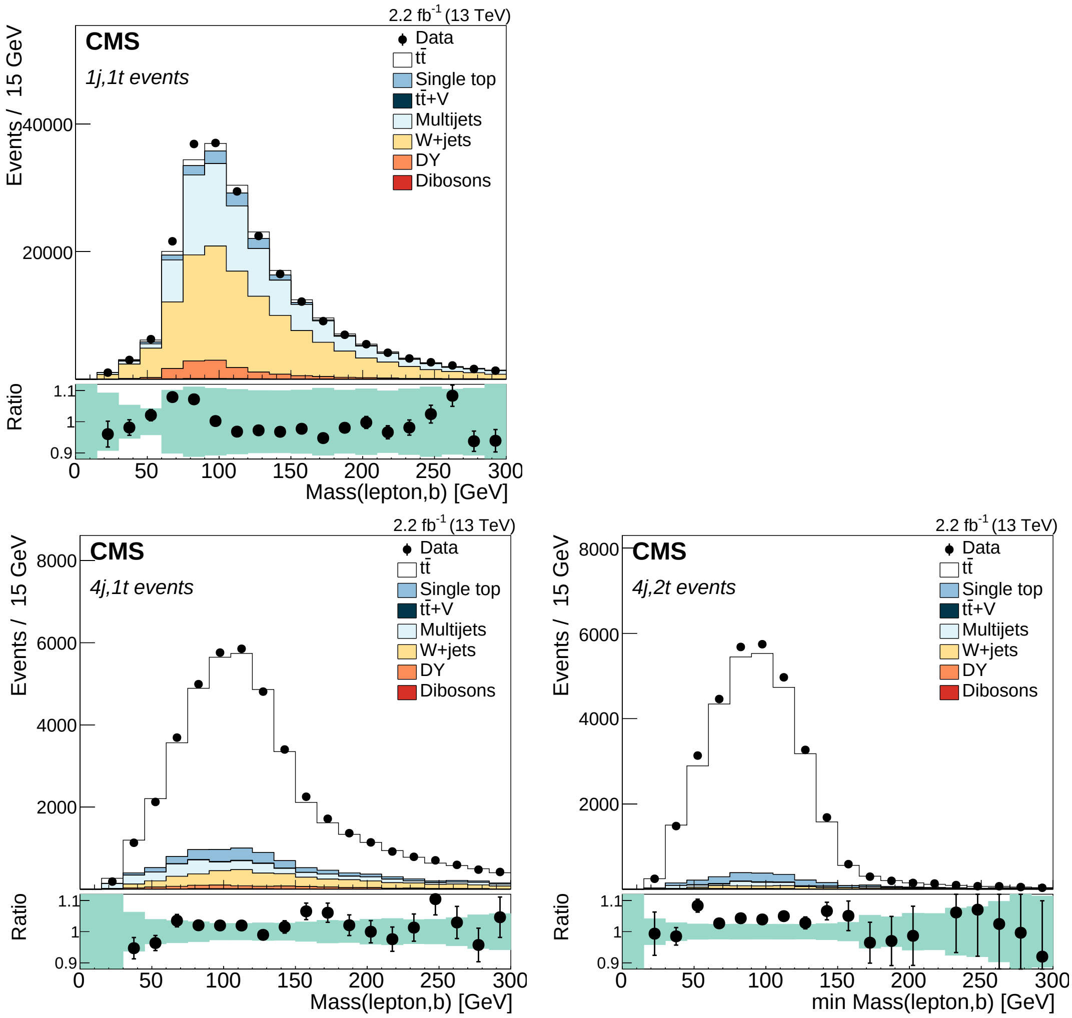


Figure 6.11: Distributions in the observables used to fit the data with the contributions from all leptons and charges combined. Panels on the left show the distributions in $M(\ell, b)$, and on the right in min $M(\ell, b)$, for events with one and two b-tagged jets, respectively. From top to bottom, the events correspond to those with 1 and at least 4 jets. The lower plot in each panel shows the ratio between the data and expectations. The relative uncertainty owing to the statistical uncertainty in the simulations, the systematic uncertainty in the total integrated luminosity and in the normalization of the multijet background is represented as a shaded band.

Chapter 7

Systematic uncertainties

This chapter is dedicated to the systematic uncertainties that are relevant for this study and the methods used to evaluate them. Section 7.1 explains the experimental uncertainties while theoratical uncertainties are described in section 7.2.

7.1 Experimental uncertainties

The Profile likelihood ratio fit (PLR) is used (section 8.3) to extract $t\bar{t}$ production cross section. The uncertainties involved in this analysis are used as the input to fit to extract the best-fit value. The experimental uncertainties including, pileup, trigger and selection efficiency, lepton and jet energy scales, jet energy resolution, unclustered $\not{E}_T$ and QCD multijet normalization are constrained by log-normal distributions. Also any variation in the lepton energy scale, jet energy scale and jet energy resolution would result into recalculation of the $\not{E}_T$. The following experimental uncertainties are involved in this study.

Pileup

Although pileup is included in the simulation, but also there is an intrinsic uncertainty involved on how to model it appropriately. To estimate the effect of mismodelling in the pileup, the average pileup scenario is varried through the choice of the minimum bias cross section parameter by 5% with respect to its initial estimate.

Trigger and selection efficiency

The uncertainties on trigger efficiency, lepton identification and isolation efficiency scale factors are propagated by re-weighting the simulation after shifting the nominal values 1σ up and down. The scale factors are determined by using the tag-and-probe method, as discussed in section 6.2.2.

Lepton energy scale

A variation of 0.5% of the muon p_T scale is used to estimate this uncertainty. For electrons, the energy scale is varried using a p_T and η dependent function [184]. The electron energy scale is typically varied between 1% (at low p_T in the endcap) and 0.1% (at high p_T in the barrel).

Jet energy scale

A p_T and η dependent parameterization of the jet energy scale is used to vary the calibration of the jets in the simulation [185]. The full breakup of the 29 jet energy scale uncertainty sources is considered in this analysis.

Jet energy resolution

According to jet energy resolution uncertainties [176], each jet is further smeared up or down depending on its p_T and η with respect to the central value measured in data. The JER is expected to impact mostly the $\not{E}_T$ shapes and to have a negligible impact on the $M(\ell, b)$ distributions.

b-tagging and misidentification efficiencies

The nominal efficiency expected in the simulation is corrected by the p_T -dependent scale factors [186]. Depending on the flavour of each jet the b-tagging decision is updated according to the measured scale factor. The scale factor is also varied according to its uncertainty. In the fit, the efficiency for tagging b, c and other jets (light quarks, gluons or unmatched) are separately quantified. Both nuisances are constrained by log-normal distributions. The expected effect on the signal shapes is depicted in Figure 7.1. The b-tagging efficiency is expected to have the largest impact on the signal shapes.

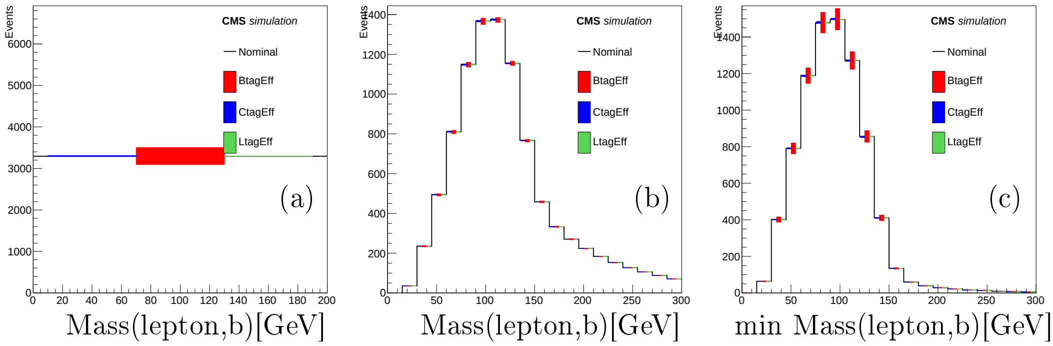


Figure 7.1: Distributions for signal shapes with the b-tagging related uncertainties superimposed in μ^{+}+4 jet events with 0 (a), 1 (b) or at least 2 (c) b-tags.

Unclustered $\not{E}_T$

The scale for the contribution to the energy flow in the detector is neither clustered in jets, nor reconstructed as leptons, is varied by 10% up and down and added back to the clustered component of the $\not{E}_T$. This estimates the residual uncertainty on the $\not{E}_T$ energy scale, which is not covered by the jet energy scale/resolution or lepton energy scale uncertainties. This has an effect only if the $\not{E}_T$ distribution is used in the fit.

QCD multijet normalization

A lepton channel and jet multiplicity dependent uncertainty on the initial normalization of the QCD multijet background is considered in the fitting model (see Table 8.1). In each jet multiplicity category, the nuisance parameter affecting the initial normalization is assumed to be distributed as a log-uniform distribution.

QCD multijet shape

The uncertainty in the amount of non-QCD processes subtracted in the side band is propagated to the analysis by deriving two alternative QCD shapes. A 30% variation on the expected yields for the non-QCD processes is used to generate the shape uncertainty envelope.

Integrated luminosity

The integrated luminosity has an uncertainty of 2.3% [187]. This uncertainty is included as a nuisance parameter in the fit, as it affects the predition for the yields of MC-based processes. In the extrapolation to the full phase space, the uncertainty is externalized again, and there is an ambiguity in the extrapolation due to the estimate of the integrated luminosity.

7.2 Theory uncertainties

Similar to experimental uncertainties, the theory uncertainties like QCD scale choices, matrix element (ME) generator, hadronizer choice and parton shower scale are also assumed to be log-normal distributed. In simulation, for hadronizer choice and parton shower scale, the b-tagging efficiency in the HERWIG ++ is corrected to match the expected one for the PYTHIA 8 simulation for which the data/MC scale factors are derived. The following theory uncertainties are considered in this analysis:

Top quark mass

The variation of the shape induced by the top quark mass in simulation is parameterized by using two alternative samples with masses 166.5 GeV and 178.5 GeV. The MC top mass parameter is thus constrained in the fit. The variations expected are depicted in Figure 7.2. The extrapolation to the full phase space results into variation of acceptance values, and hence the top mass, which will be described in section 8.4.

QCD scale choices

The independent variations of the factorization and renormalization scales (μ_R/μ_F) in the $t\bar{t}$ and W+jets, by factors of 0.5 and 2, are considered. These variations are saved in the simulated events as alternative sets of weights and used in the evaluation of this systematic. In the fit these uncertainties are not correlated for different processes. The effect on the signal shapes is very small for the signal, as depicted in Figure 7.2. The effect on the normalization is externalized and assigned separately as an extrapolation uncertainty. This shall be discussed in more detail in section 8.4. Figure 7.3 shows the impact of the QCD scale choice for the W process. In this case the impact is large. For tW, the considered variation is simultaneous and obtained from alternative POWHEG (v1) samples where the scale has been varied. See Figure 7.5 for the effect induced on the tW sample.

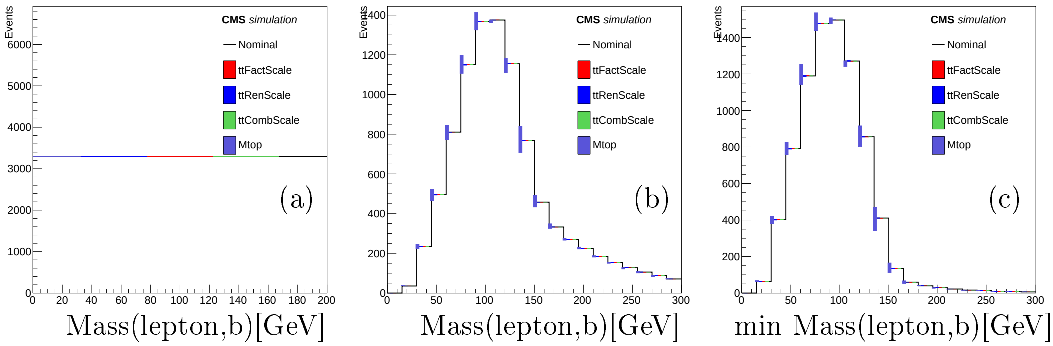


Figure 7.2: Distributions for the signal shapes with the QCD-scale choice related uncertainties and a $\Delta m_t \pm$ 6 GeV variation superimposed in μ^++4 jet events with 0 (a), 1 (b) or at least 2 (c) b-tags.

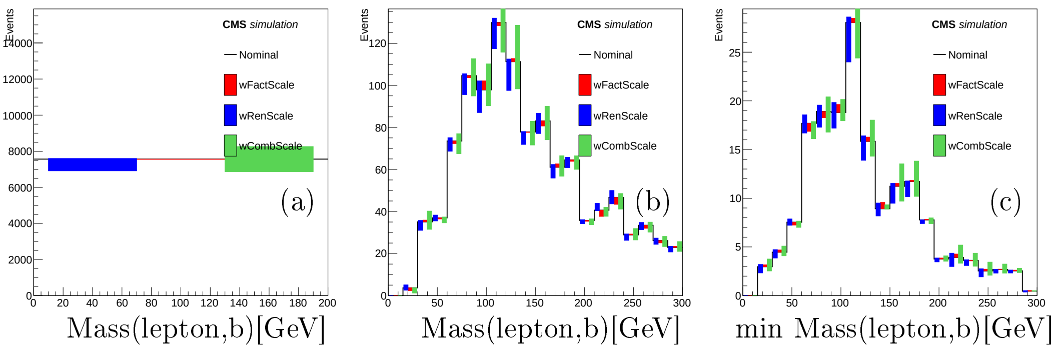


Figure 7.3: Distributions for the W+jets shapes with the QCD-scale choice related uncertainties superimposed in μ^++4 jet events with 0 (a), 1 (b) or at least 2 (c) b-tags.

Top p_T

The uncertainty due to the modelling of the top quark p_T distribution in simulation is evaluated by applying a re-weighting at the generator level [188]. The variations induced by this re-weighting are depicted in Figure 7.4.

ME generator

The default simulation is based on POWHEG, which is compared with MG5_aMC@NLO+ PYTHIA 8-based simulation and an uncertainty is assigned to it. The variations found in the shapes are mirrored. The variations expected are depicted in Figure 7.4.

Hadronizer choice

Alternative predictions obtained with POWHEG +HERWIG ++ are used to estimate the uncertainty in modelling the fragmentation and hadronization of jets, as well as the modelling of the underlying event. The variation is symmetrized at each bin. The variations expected are depicted in Figure 7.4.

Parton shower scale

Alternative POWHEG +PYTHIA 8 samples, where the parton shower scale choice is varied by a factor of 0.5 and 2, are used in the analysis. This affects the fragmentation and hadronization of the jets initiated by the matrix element calculation as well as the emmission of extra jets by the hadronizer. The effect on the signal shapes is depicted in Figure 7.4.

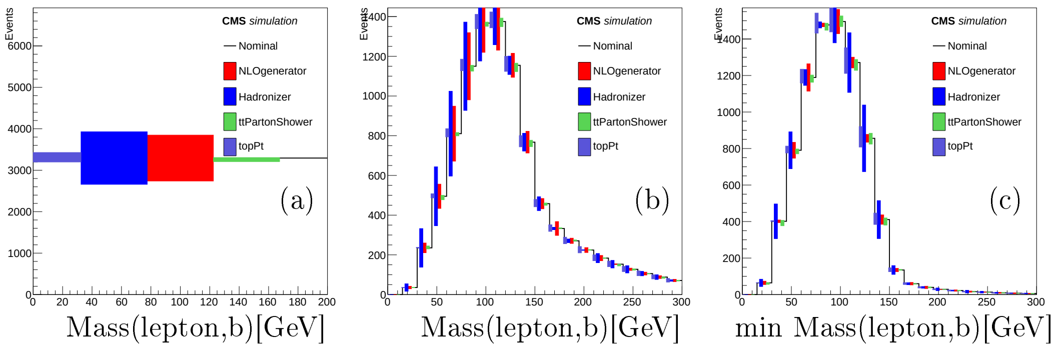


Figure 7.4: Distributions for the signal shapes with the parton shower scale, hadronization, NLO generator choice, and top p_T uncertainties superimposed in μ^++4 jet events with 0 (a), 1 (b) or at least 2 (c) b-tags.

PDFs

The impact on the shapes is expected to be small and thus this uncertainty is externalized. The RMS of 100 replicas of the NNPDF3.0 set [163] is computed and in added in quadrature to 2 extra variations corresponding to different choices of α_S.

tW/$t\bar{t}$ interference

As at NLO QCD, tW production is expected to interfere with $t\bar{t}$ production, given the similar initial and final states of some diagrams [32, 189, 190], two schemes for defining the tW signal, distinguishing it from $t\bar{t}$ production have been compared in this analysis: the diagram removal (DR) [32], in which all doubly resonant NLO diagrams are removed, and the diagram subtraction (DS) [32, 33], where a gauge-invariant subtraction term modifies the NLO tW cross section to locally cancel the contribution from $t\bar{t}$. Figure 7.5 illustrates the expected effect on the shape and normalization of tW.

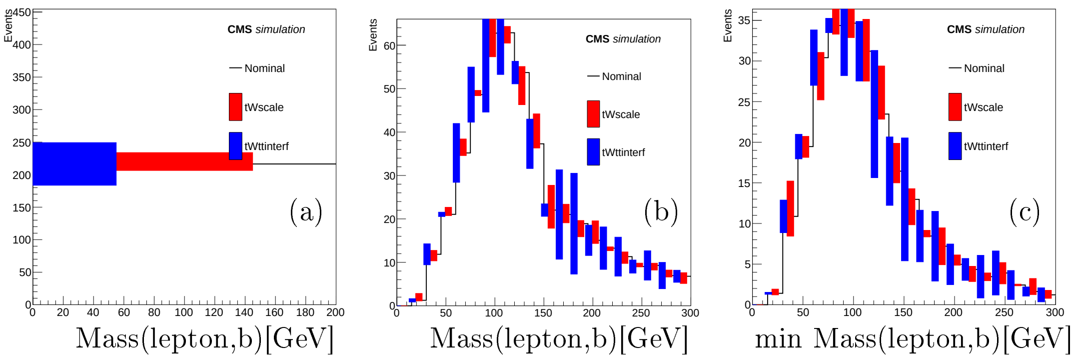


Figure 7.5: Distributions for tW shapes with the QCD-scale choice and tW/$t\bar{t}$ -related uncertainties superimposed in μ^{+}+4 jet events with 0 (a), 1 (b) or at least 2 (c) b-tags.

Background normalization (non multijet QCD)

The theory uncertainty on the cross sections for the backgrounds are included as nuisance parameters.

Chapter 8

Measurement of the Top-Quark Pair Production Cross Section

This chapter describes the measurement of the $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$ at 13 TeV center-of-mass energy of pp collisions. The precise $\sigma_{t\bar{t}}$ measurement is performed by using semileptonic decay mode along with further corrections in addition to pileup re-weighting, jet energy scale and resolution, trigger and lepton efficiencies etc. section 6.1.1.

In this chapter, section 8.1 describes the strategy adopted in this thesis, section 8.2 is dedicated to explain how major backgrounds in this study are estimated. Section 8.3 is about the statistical methods used in the analysis. Section 8.4 describes the acceptance measurement and finally section 8.5 describes the extration of cross section in fiducial region and extrapolation to full phase space.

8.1 Strategy

The *fiducial phase space* is defined as the kinematic region of the decay products of any system that can be measured by the detector. The $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$ in *fiducial phase space* is defined as:

$$\sigma(t\bar{t})_{fid} = \frac{N^{obs} - N^{bkg}}{\mathscr{L}\epsilon^{tot}} \ , \tag{8.1}$$

where N^{obs} are the total number of observed events, N^{bkg} are the background events, $\mathscr{L}$ is the integrated luminosity and ϵ^{tot} is the total efficiency of the reconstructed and selected $t\bar{t}$ events.

In the semileptonic (lepton+jets) final state, the strategy is to explore the heavy flavor content of the events. Different categories of events are defined on the basis of number of jets and number of b-jets. In semileptonic final state, two heavy b-jets are produced from the $t\bar{t}$ decay and two light jets are produced from the decay of W-boson. The categories with number of light jets less than four and b-jets less than two, are termed as low jet multiplicities in this study. On the other hand, the category with at least four jets two of which are b-jets, is termed as high jet or b-jet multiplicity. This high jet multiplicity is defined as the signal region of $t\bar{t}$ decay in this analysis. Low jet multiplicity categories are used to constrain the backgrounds while high jet multiplicity category is used to fit the signal. A profile likelihood ratio fit is used (section 8.4) to establish the signal strength in the *fiducial phase space* and the result is extrapolated to the *full phase space* through a generator-based acceptance model.

As W+jets is one of the major background in this analysis and due to charge asymmetric nature of this process [191], the analysis is also divided according to lepton charge. Since $t\bar{t}$ is assumed to be charge symmetric [192], this division is used to estimate the W+jets contamination. The analysis is started from one jet with 0 b-jet category (1j0t) and proceeded to 4 jets with two b-tagged jets (4j2t) category, which is the signal region. W+jets, which is a dominant background in 1j0t category is well modelled in 4j2t category as shown in Figure 8.1.

8.2 Background estimations

QCD multijet and W+jets are the main backgrounds in the semi-leptonic decay channel. In the next sections two strategies are discussed to model and normalize these two processes. All the background processes are estimated from simulation, with the exception of QCD. The usage of different categories in the analysis will allow to derive in-situ corrections to the simulated predictions of these background

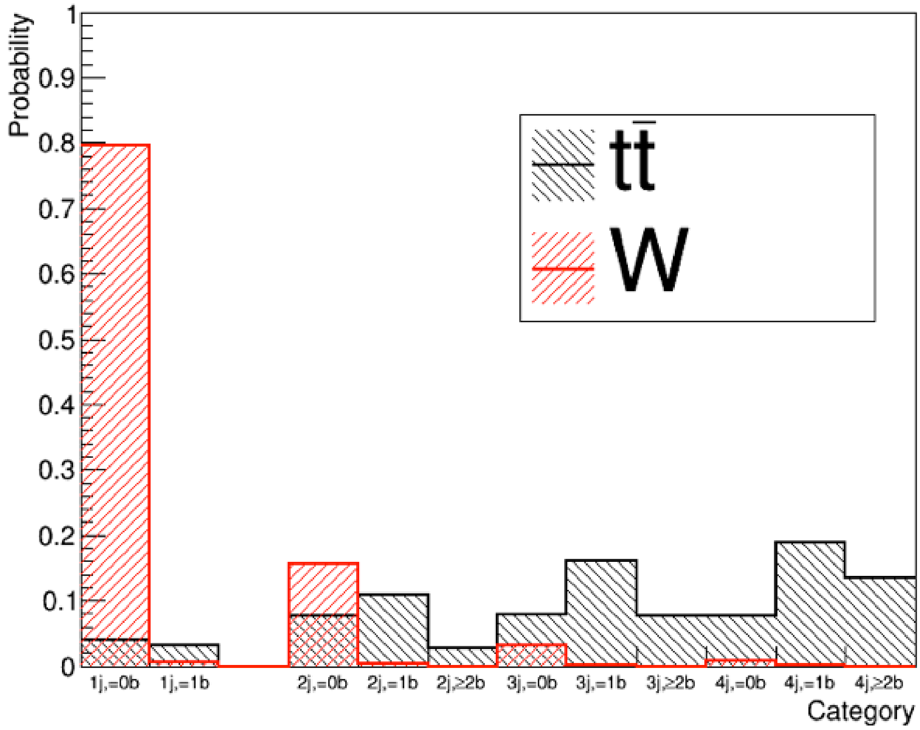


Figure 8.1: The probability of MC simulations of W+jets and $t\bar{t}$ for different categories having different combinations of light and b-tagged jets. For example, in 1j,=0b represents one light jet and no b-tagged jet.

processes. Nevertheless, a particular attention is given to the W+jets background, which will be described later.

8.2.1 QCD multijet estimation

Multijet QCD background can mimic the semi-leptonic $t\bar{t}$ process through semileptonic decay of heavy flavour jets and fake leptons in the final state.

Semi-leptonic decay of heavy flavour jets

Long-lived charmed and beauty hadrons give rise to non-prompt and less isolated leptons that are contained within a jet. These are the true leptons which can acquire a significant p_T given that charm or beauty hadrons tend to carry most of the energy of a jet due to the fragmentation and hadronization of the initial c or b quark.

Fake leptons

Fake leptons either from punch-through hadrons or jets with a high electromagnetic fraction are not absorbed in the hadronic calorimeter and can generate tracks in the muon system.

The significance of impact parameter in 3 dimensions (SIP3d) with respect to the primary vertex variable can be used to flag the first source of leptons as long-lived heavy flavoured hadrons yielding displaced vertices. Figure 8.2 shows the distribution of the SIP3d variable for events with different jet multiplicities in the non-isolated (the region in which isolation criteria of the lepton is inverted due to which leptons are not well-isolated) and isolated (the region in which leptons are well-isolated) muon and electron regions. In the analysis, the criteria required for SIP3d is 4σ above background, given there is some mismodelling of the alignment. This does not impact significantly the uncertainty in the lepton selection efficiency. The variable removes a considerable amount of non-prompt leptons from the QCD background. For electrons the gain from this cut is not as pronounced as for the muons, as the electron track reconstruction has poorer resolution than the muon case. Nevertheless the same cut in both channels is applied.

The simulated MC samples can not reproduce the shape and size of the QCD multijet background because of the well-isolated and hard leptons requirements. Therefore, this background is estimated from data-driven technique. Assuming that the kinematic properties of the multijet events in selected phase space are independent of the lepton isolation in the event, a control region is defined where the isolation criteria is reversed. This control region (CR) is enriched with multijet background events and has low contamination from other processes ($t\bar{t}$ included). The CR is chosen by inverting the isolation requirement used for the lepton while all other cuts remain unchanged. The relative isolation, $I_\ell(\Delta\beta) > 0.4$, is required in both cases. The $I_\ell(\Delta\beta)$ value is chosen as a tradeoff between the statistics and the expected contamination of the CR by signal and W+jets events. Such impurities are still there, and potentially biase the obtained shape of the multijet background, especially in high jet multiplicity. The expected contribution from

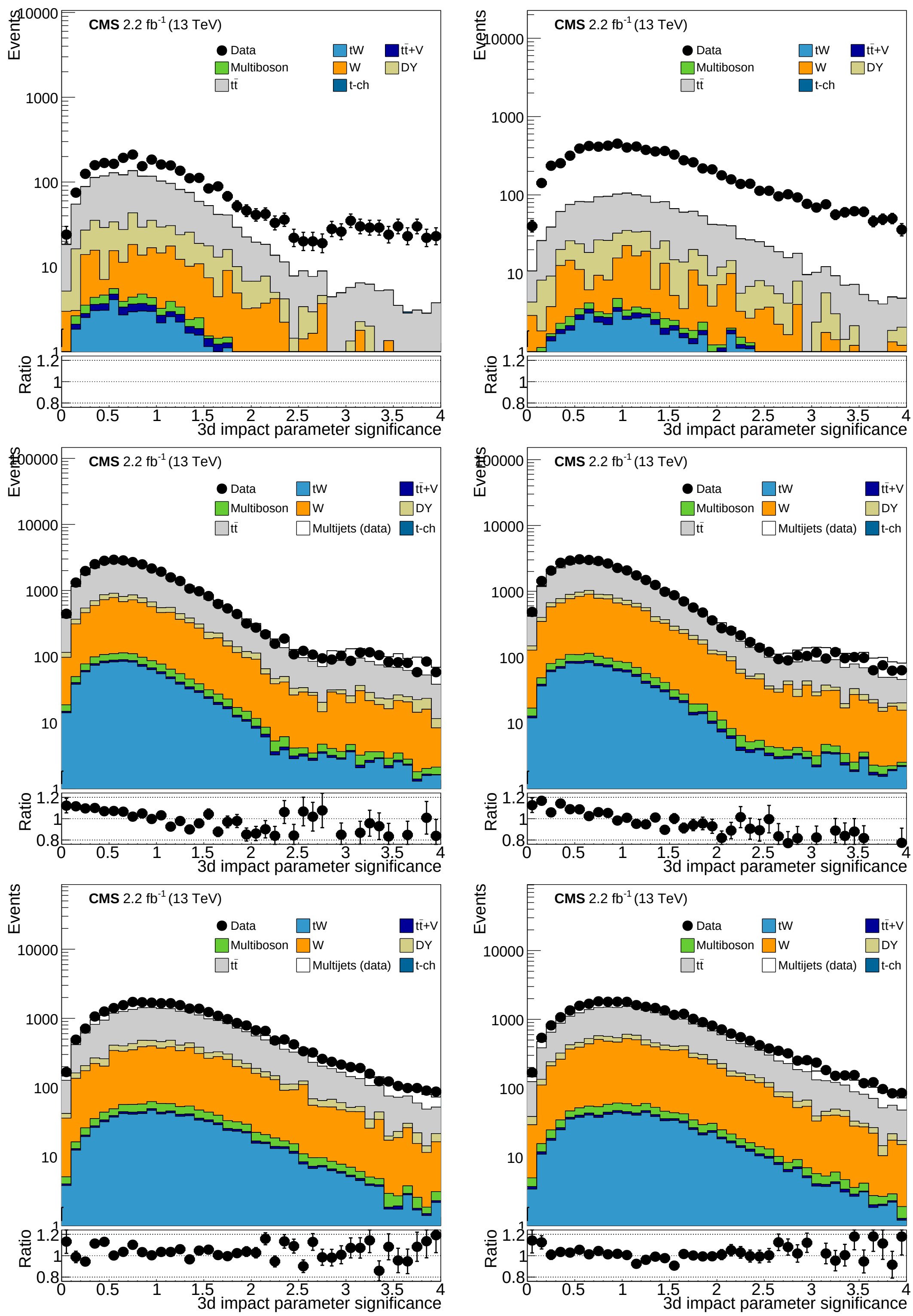


Figure 8.2: Distributions for the SIP3d variable for at least 4 jets events in the non-isolated region (top), isolated μ (center) and isolated e (bottom). From left to right events with minus and plus lepton charge. The bottom pannel shows ratio of the data to MC simulations (shown in top pannel).

non-QCD processes is subtracted from the data in the CR by assigning a 30% uncertainty to it, as shape uncertainty (explained in section 7.1).

In order to normalize the QCD in the signal region (SR), the $\not{E}_T$ distribution is used. Low $\not{E}_T$ events are expected to be enriched by processes with no neutrinos where reconstruction effects and pileup dominate the $\not{E}_T$ resolution. The number of events with $\not{E}_T$ <20 GeV in the signal and control regions are counted. After subtracting the expected contribution from non-QCD processes, the ratio of the number of events observed in the CR with respect to the number of events observed in the SR is used as the transfer factor to be used to re-scale the number of QCD events, i.e.

$$N_{\rm SR}(QCD) = [N_{\rm CR}(QCD)] \cdot \frac{N_{\rm SR}^{E_T^{miss}<20}(obs) - N_{\rm SR}^{E_T^{miss}<20}(non-QCD)}{N_{\rm CR}^{E_T^{miss}<20}(obs) - N_{\rm CR}^{E_T^{miss}<20}(non-QCD)} \,, \quad (8.2)$$

where $N_{\rm CR}(QCD) = [N_{\rm CR}(obs) - N_{\rm CR}(non-QCD)]$, SR(CR) refers to signal (control) region, obs refers to the observed data and non-QCD refers to the estimate of non-QCD processes. An uncertainty is assigned to the initial estimate of QCD events in the SR, as provided by Equation 8.2. This is done by applying a variation of 30% on the non-QCD subtracted events in the SR and CR and taking into account the statistic uncertainty associated to the counting of the data in the SR and CR. The envelope of the variations is used to assign the initial uncertainty on the normalization for QCD.

The determination of the scale factor is performed using different requirements. A large fraction of QCD events is expected from the $b\bar{b}$. Thus for events with one or two jets, the scale factors are estimated from events with one b-tag. However, for events with at least three jets, $t\bar{t}$ starts contaminating more effectively this region, therefore, no b-tag is required to avoid signal lost.

Table 8.1 summarizes the scale factors to be applied to the CR distributions for QCD to normalize them in the SR. The uncertainty assigned to the normalization is also mentioned. Overall the scaling factors are similar accross the different jet multiplicity categories and tend to increase for higher jet multiplicities. It is noticed that the uncertainty is higher in the positively charged lepton sample as a consequence of charge asymmetry in W+jets, i.e. more prompt positively charged leptons have to be subtracted in the sideband. All scale factors are close to two, which is related to the fact that data is triggered by isolated lepton triggers, in practice only a portion of non-isolated leptons is captured offline. A fitting procedure will be used, to improve the intial estimate of this background (see section 8.3).

Table 8.1: Scale factors applied to the data-based estimate of the QCD multijet backgrounds in the non-isolated lepton control region. The uncertainty on the scale factor (or final normalization) is shown in percentage.

Category	=1 jet	=2 jets	=3 jets	≥4 jets
μ^+	1.89 ± 0.09	1.63 ± 0.14	2.77 ± 0.24	2.17 ± 0.23
μ^-	1.59 ± 0.05	1.19 ± 0.05	2.64 ± 0.22	2.01 ± 0.15
e^+	2.71 ± 0.09	2.29 ± 0.10	2.60 ± 0.13	2.14 ± 0.30
e^-	2.50 ± 0.12	2.06 ± 0.05	2.56 ± 0.13	2.03 ± 0.10

Figure 8.3 shows the $\not{E}_T$ distributions in the muon and electron channels in the CR and SR (including the re-scaled shape obtained from the CR for QCD for an indicative case of four jets). No b-tagging requirement is made, and a fair agreement is observed for all categories.

8.2.2 Controlling W+jets contamination

The expected contribution from W+jets is modelled from simulation. In each lepton channel, the analysis is subdivided according to the lepton charge. While $t\bar{t}$ production is expected to be symmetric with respect to the charge, W production is expected to be asymmetric due to the PDF composition of the protons [191]. The charge asymmetry is expected to depend on the jet multiplicity, as well as on the flavour of the jets [193] in the event. The asymmetries predicted by simulation are assumed, and the corresponding uncertainties from the choice of the QCD scales and PDFs are assigned. The fit to be performed (see section 8.3) will fine tune the initial predictions.

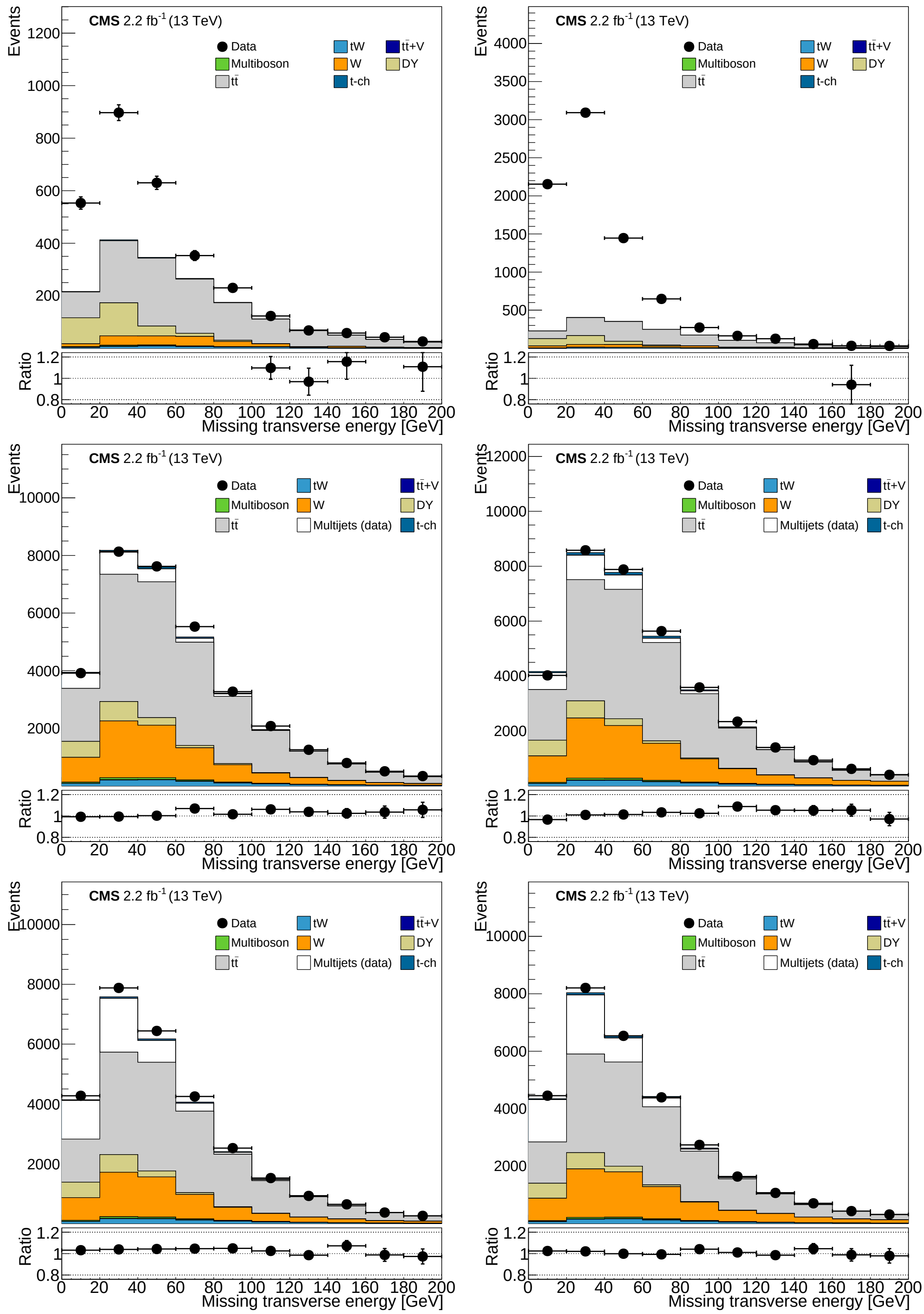


Figure 8.3: Distributions for the $\not{E}_T$ variable for at least 4 jets events in the non-isolated region (top), isolated μ (center) and isolated e (bottom). From left to right events with minus and plus lepton charge. The bottom pannel shows ratio of the data to MC simulations (shown in top pannel).

The probability distribution functions for different observables are currently derived from the W+jets MG5_aMC@NLO [165] sample, despite the large fraction of events with negative weights, the sample has sufficient statistics to provide a fairly smooth template. However this sample is compared with the MadGraph [74] one. A difference in normalization between the two, in the different categories, is observed. But the shape of the distributions is fairly similar. The normalization difference attributed to the NLO corrections included in the MG5_aMC@NLO sample. Figures 8.4 and 8.5 summarize the results obtained for each category and distribution to be used in the analysis, for $\not\!\!E_T$, $M(\ell, b)$ and min $M(\ell, b)$ three and four jets.

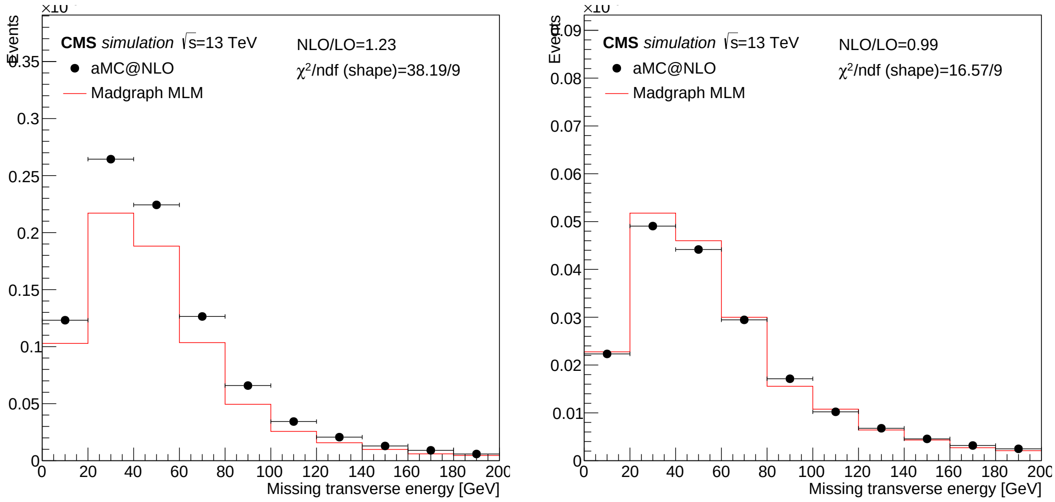


Figure 8.4: Distributions for the $\not\!\!E_T$ variable with 3j0b-tagged jets (left) and 4j0b-tagged jets (right).

Table 8.2 shows the ratio of the events predicted by MG5_aMC@NLO with respect to the MadGraph simulation. Overall, more events are expected in the NLO simulation, except when at least four jets are present in the event. In that case, the tendency is to obtain a lower normalization by ≈3-4%.

As a consequence of the previous observation, the expected charge asymmetry defined as $A_c = (N_+ - N_-)/(N_+ + N_-)$ is different. The QCD scale variations are also assigned an uncertainty. These are expected to cover the differences in yields (charge asymmetry) between the MadGraph and MG5_aMC@NLO predictions, although the latter are expected to be more accurate as higher order QCD corrections are included in the computation.

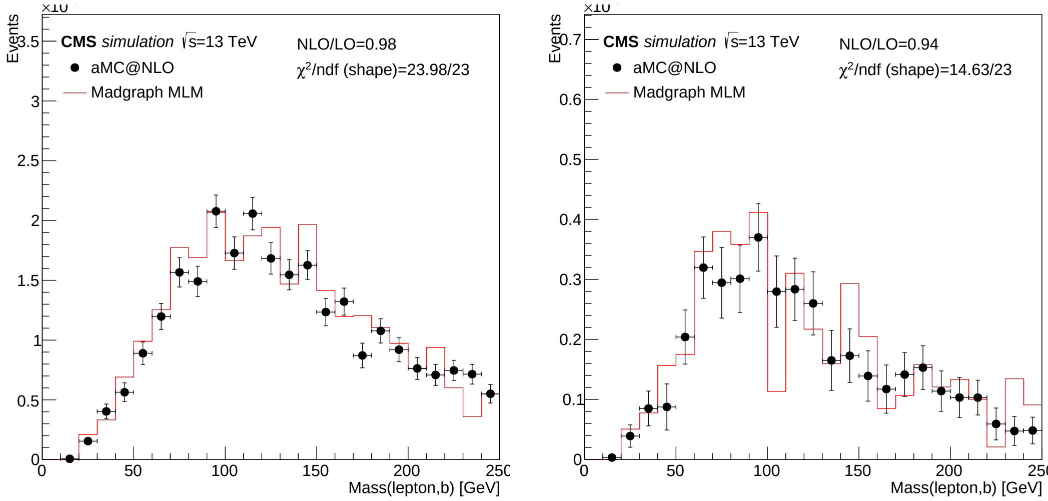


Figure 8.5: Distributions for the $M(\ell, b)$ varible events with 4j1b-tagged jets (left) and min $M(\ell, b)$ for 4j2b-tagged jets (right).

Table 8.2: NLO/LO scale factors for W+jets derived from MadGraph and MG5_aMC@NLO samples for the different categories used in the analysis.

Category	=1 jet	=2 jets	=3 jets	≥4 jets
μ^+	1.112±0.002	1.184±0.006	1.20±0.01	0.96±0.02
μ^-	1.095±0.003	1.170±0.006	1.19±0.01	0.97±0.02
e^+	1.102±0.003	1.196±0.007	1.21±0.02	0.96±0.02
e^-	1.288±0.004	1.38±0.008	1.40±0.02	1.09±0.03

8.3 Method used for the statistical analysis

The cross section for $t\bar{t}$ production is extracted after performing a profile likelihood ratio fit (PLR) to the number of events counted in the different categories. In this case, a distribution is used in a given category and the fit is naturally extended to include the different bins of the variable of interest. The PLR fit takes into account the expectations for the different background processes as well as the signal. The expectations for the signal and a background process in a category can be estimated as depending on the simulation or from a data-driven method ($\hat{s}$ or $\hat{b}$ for signal and background correspondingly), and nuisance parameters (θ_i) that can parametrize the effect of the systematic uncertainties associated to the estimate of these expectations.

The nuisance parameters can be expressed as multiplicative factors, which affect the nominal prediction of the event yields of each process i.e. for a given expected rate-x, it is affected by $x \rightarrow x(1+\sigma_i\theta_i)$ where the θ_i is distributed according to a pre-defined probability distribution function, and the σ_i are the relative variation in the yields induced by the i-th source of uncertainty. The nature and values of the nuisances are discussed in the next sections, as well as the probability distributions assigned in the fit. The constraints allow the nuisances to float in the neighbourhood of its expected value enabling a fine adjustment of their initial values using the observed data. For the signal, further a multiplicative factor is considered which is the ratio of the cross section to be measured in the data to expected theoretical value, i.e. the signal strength $\boldsymbol{\mu} = \sigma/\sigma_{th}$. For each category (or observation bin), the total number of expected events are written as:

$$\hat{N}_k(\boldsymbol{\mu},\theta_i) = \hat{N}^s_k(\mu,\theta_i) + \hat{N}^b_k(\theta_i) = \boldsymbol{\mu}\cdot\hat{s}_k\cdot\prod_i(1+\sigma^s_i\theta_i) + \sum_{b\in bkg}\hat{b}_k\cdot\prod_i(1+\sigma^b_i\theta_i) \quad (8.3)$$

Using Equation 8.3, the likelihood is written as a product of Poisson distributions and nuisance constraints:

$$\mathscr{L}(\boldsymbol{\mu},\theta_i) = \prod_k \mathscr{P}_{oisson}\left[N_k|\hat{N}_k(\mu,\theta_i)\right]\cdot\prod_i pdf(\bar{\theta}_i = 0,\theta_i,1) \quad (8.4)$$

In Equation 8.4, N_k is the number of events observed in the k-th category. Based on the likelihood expressed in Equation 8.4 the PLR-test statistics are defined as:

$$\lambda(\boldsymbol{\mu}) = \frac{\mathscr{L}(\boldsymbol{\mu},\hat{\hat{\theta}}_i)}{\mathscr{L}(\hat{\boldsymbol{\mu}},\hat{\theta}_i)} \quad (8.5)$$

The quantities $\hat{\hat{\theta}}$ corresponds to the value of θ which maximise the likelihood for the specified signal strength (also known as conditional likelihood). The denominator $\mathscr{L}(\hat{\boldsymbol{\mu}},\hat{\theta}_i)$, is called the un-conditional likelihood function depending on estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\theta}$. This approach allows to cover the effect of the systematic uncertainties in the fit; in the presence of the nuisance parameters the likelihood as function of $\boldsymbol{\mu}$ tends to be broader with respect to the one obtained for fixed values. This reflects the loss of information in $\boldsymbol{\mu}$ due to systematic effects. More details can be found in [194].

For the computation of the PLR described in Equation 8.5, RooStats [195] framework is used through the Higgs combination tool [196].

8.3.1 Fit setup

In order to improve the precision of the measurement, a good decorrelation of the postfit uncertainties is required. This section describes the techniques to handle nuisances for achieving a better decorrelation.

Factorize rate and shape uncertainties

For each source of uncertainty the contribution to rate and shape is factorized separately. Rate uncertainties are assigned a log-normal PDF, while shape uncertainties are assigned a gaussian PDF and its effect on the shape is parametrized linearly (quadratically) inside (outside) the 1-σ range.

Finite statistics in the MC

Most of simulations employed have enough statistics and this affects to the backgrounds in the highest jet or b-jet multiplicities, such as Drell-Yan and tW. For each multiplicity, more than 30% relative variation is found, an alternative shape is instantiated where the bin is allowed to vary independent of the others.

QCD multijet shape

The uncertainty on the QCD multijet shape is derived by changing the amount of non QCD multijet, processes which is subtracted in the isolation sideband. In this analysis bin-by-bin variations are used.

QCD multijets normalization

In some cases (most notably for the electron channel), it is identified that the prescription to assign at least 30% uncertainty on the initial normalization of this background is being followed exactly.

Lepton efficiencies

Besides applying the rate and shape factorization (duplicating the number) of nuisances associated to trigger and offline selection efficiencies the prescribed recommendation to add 0.5% in quadrature to the muon trigger efficiency [197] is also considered.

W QCD scales

W QCD scales have been uncorrelated among categories. This is maximally conservative as it limits the capability of the fit to constrain these variations by correlating different categories.

DY background

DY contribution has been uncorrelated among categories.

Integrated luminosity

The integrated luminosity and associated uncertainty are updated according to [198]. The total integrated luminosity decreases by 2.7% and the total uncertainty assigned is 2.3%.

Beam energy uncertainty -

The uncertainty on the beam energy is used as 0.1% [199]. This translates to an almost negligible uncertainty on the theory prediction for the $t\bar{t}$ cross section of 0.23%.

The so-called MIGRAD algorithm with strategy 2, as implemented in MINUIT2 [200] is adopted to compute the impacts of the nuisances. This algorithm is strongly dependent on derivatives, but if the computation of the derivatives is numerically stable, it is expected to have the best performance comparatively to other methods. In order to separate the rate and shape systematics, it is expected that overall performance of MIGRAD is stable. The datacards used in this study are available in [201].

8.3.2 Expected performance

The fit is performed in two different strategies, cut-in-categories or count analysis and shape or distribution analysis. The results obtained applying the fit procedure in these analysis strategies are compared in this section.

Cut-in-categories or Count analysis

The data events are counted in different categories containing different levels of $t\bar{t}$ purity. Since the categories are defined by the jet and b-tagged jet multiplicities and the lepton charge, so a total of $2\times2\times11 = 44$ categories are used in the analysis. The expected evolution of the counts in the categories is correlated and anti-correlated according to the expected effect of each systematic. With this approach the fit is expected to constrain some of the experimental systematic uncertainties and also some of the theory uncertainties pertaining the modelling in the fiducial volume. A simplified version using only events with 4 jets (a total of $2\times2\times3 = 12$ categories) was also considered.

In Table 8.3 the expected and observed results of the fits in the individual channels and after different combinations are reported. The results are in agreement between the different channels and the final combination. The combination of the electron and muon channels and of the +/- lepton charge categories does help in improving the final uncertainty at both the expectation and observation levels. The handling of the QCD multijet background in the electron channel is significantly good and in agreement with the muon results.

Figure 8.6 compares the data with the post-fit distributions for an indicative case of four jets used in the cut-in-categories analysis. For each plot, the distortion induced by the fit in the shapes of the signal and background is shown and compare the data to the prefit and postfit models. The agreement with the postfit model is significantly improved with respect to the prefit one. The changes in the signal shape are mild, and evolve with b-tagging multiplicity. The changes in the background shape are most notorious in the low b-tagging categories in the electron channel (due to adjustment of the QCD contribution) and in the

Table 8.3: Expected and observed fit results for different categories and combinations in the cut-in-categories analysis. The uncertainties are the total uncertainties (statistical+systematic) in the fiducial phase space.

Category	Expected	Observed
μ^+	$1.000^{+0.048}_{-0.046}$	$1.081^{+0.051}_{-0.052}$
μ^-	$1.000^{+0.049}_{-0.048}$	$1.068^{+0.052}_{-0.051}$
μ	$1.000^{+0.043}_{-0.043}$	$1.074^{+0.047}_{-0.045}$
e^+	$1.000^{+0.062}_{-0.058}$	$1.033^{+0.062}_{-0.059}$
e^-	$1.000^{+0.058}_{-0.055}$	$1.022^{+0.060}_{-0.057}$
e	$1.000^{+0.056}_{-0.053}$	$1.030^{+0.057}_{-0.053}$
+	$1.000^{+0.042}_{-0.041}$	$1.067^{+0.045}_{-0.043}$
-	$1.000^{+0.041}_{-0.040}$	$1.039^{+0.045}_{-0.043}$
≥2j	$1.000^{+0.039}_{-0.038}$	$1.055^{+0.047}_{-0.041}$
≥3j	$1.000^{+0.040}_{-0.039}$	$1.049^{+0.043}_{-0.041}$
≥4j	$1.000^{+0.063}_{-0.059}$	$1.016^{+0.064}_{-0.060}$
combined	$1.000^{+0.038}_{-0.038}$	$1.054^{+0.043}_{-0.041}$

high jet/b-tagging multiplicity bins. In both cases (signal/background) shapes are changed by 10%.

The change of the shapes (rates) reflects changes in the post-fit values of the shape (rate) nuisances. To assess the impact of each nuisance in the fits, the fits are repeated by fixing one nuisance at the time at its $\pm 1\sigma$ post-fit uncertainty. The deviation induced in the signal strength is called the impact. The impact of the different sources of uncertainties in the fit is shown in Figures 8.7 and 8.8 for the inclusive fit. Table 8.4 summarizes the estimated impact and correlations of different nuisances with the signal strength. In the fits, no nuisance is pulled by $> 1\sigma$ and the post-fit uncertainties are very similar to the ones expected from the Asimov datasets. The uncertainties due to luminosity, trigger and selection efficiencies are neither constrained in the fit, nor pulled significantly from their initial estimate. The larger impacts are due to the W modelling (renormalization scale), DY and QCD multijet through their contribution to the 3 jets category. The uncertainties in the integrated luminosity, trigger and selection efficiency are also among the ones with highest impact. Their apparent impact is mitigated by correlations with the other nuisances when the fits are repeated. All other

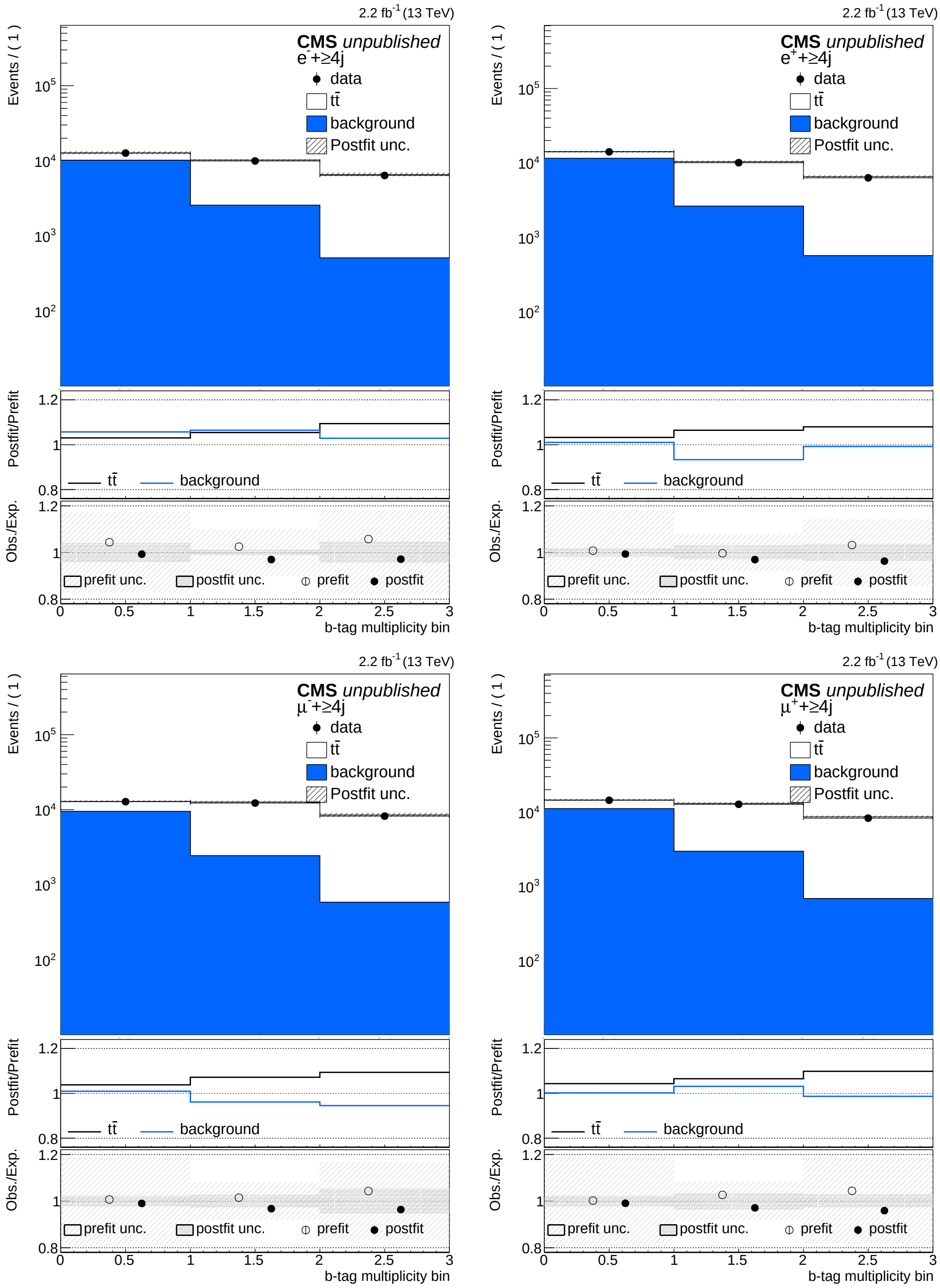


Figure 8.6: Comparison of the post-fit distributions for at least 4 jets category with the data for e channel (top) and for μ channel (bottom). Negatively (positively) charged distributions are shown on left (right). The top panels show the stacked distributions of signal and background processes and the observed data. A shaded band represents the post-fit uncertainty. The middle panels show the ratio of the post-fit to pre-fit distributions for the signal and background sum. The bottom panel shows the data to signal plus background expectations, prior and after the fit. The error bar in the points is the statistics in data. A shaded band represents the relative postfit and prefit uncertainties.

nuisances are estimated to have an impact less than 1% even if at pre-fit level they may be larger in magnitude.

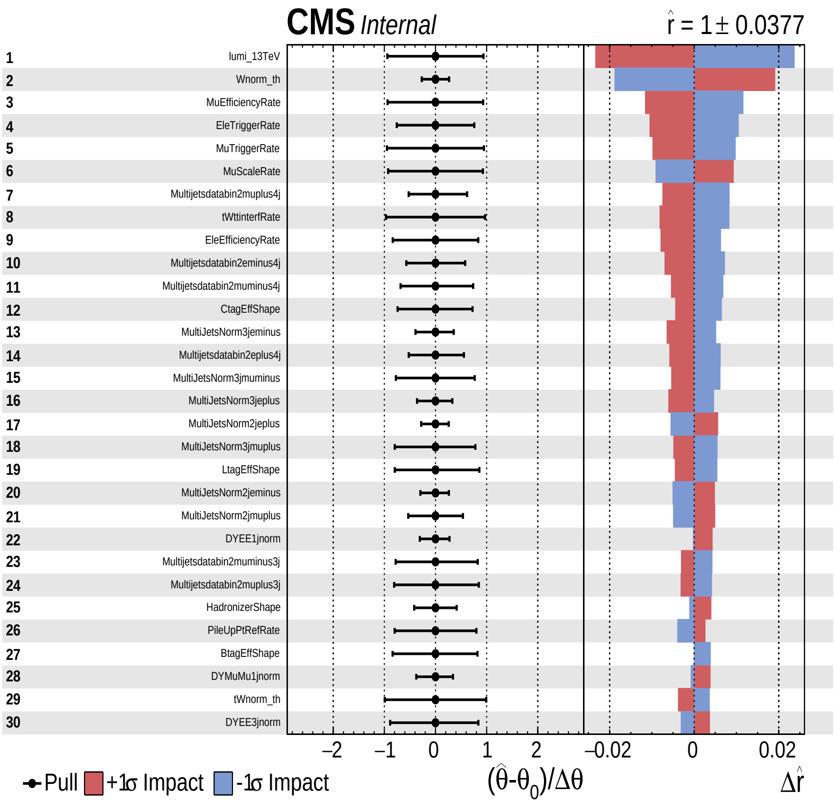


Figure 8.7: Summary of the fit results to the Asimov dataset using the cut-in-categories analysis. The left panel shows the post-fit pull (value and uncertainty) of each nuisance, while the right panel displays the estimated impact on the fit for the $t\bar{t}$ signal strength. Only the first 30 nuisances are displayed and their name is shown at each row of the plots.

The distribution of the post-fit correlations of the nuisances with higher impact and the signal strength with each other are shown in Figure 8.9. Although both distribution peak at 0, they have a ≈10% RMS. The correlation matrix for this set of nuisances is also shown in Figure 8.10. Given the non-negligible correlations between nuisances, the quadratic sum of the impacts is neither expected to yield back the total uncertainty of the fit, nor expected to reflect the direct impact of the nuisance if it had been externalised in the fit. In addition, smaller impact in the fit can't be interpreted as a reduction of the uncertainty associated to a nuisance.

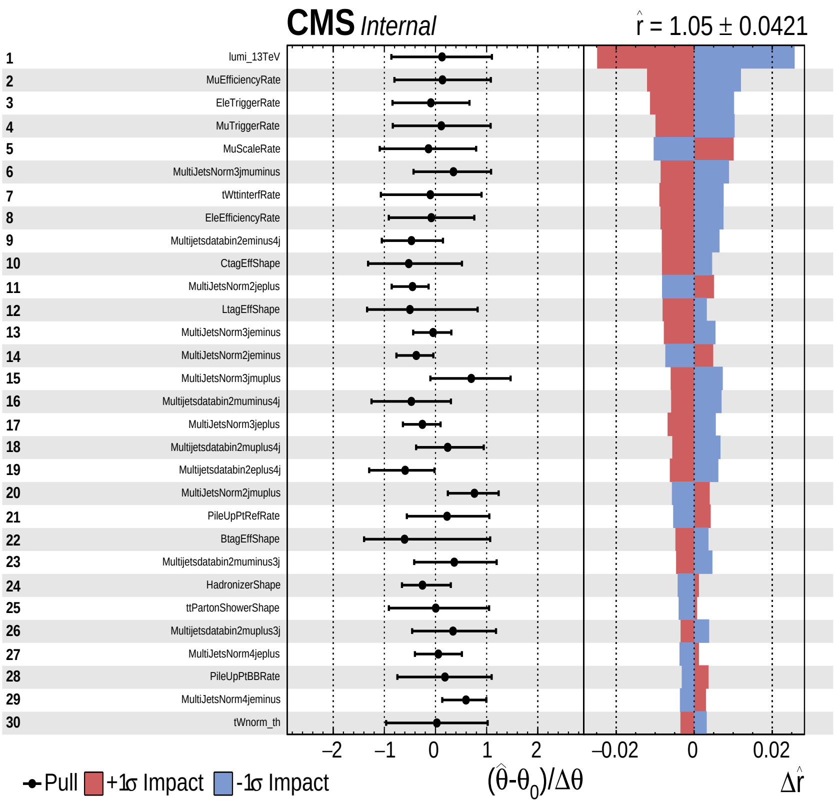


Figure 8.8: Summary of the fit results for data using the cut-in-categories analysis. The left panel shows the post-fit pull (value and uncertainty) of each nuisance, while the right panel displays the estimated impact on the fit for the $t\bar{t}$ signal strength. Only the first 30 nuisances are displayed and their name is shown at each row of the plots.

Table 8.4: List of nuisances with their impact on the fit in the cut-in-categories analysis. The middle column shows the estimated impact, while the right-most column shows the correlation with the signal strength in the fit.

Source	$\Delta\mu$	$\rho(\mu)$
Integrated luminosity	0.026	-0.05
Trigger/selection efficiency		
e trigger	0.011	-0.02
μ trigger	0.010	-0.05
μ eff.	0.012	-0.05
e eff.	0.009	-0.03
QCD modeling		
$\mu^- + 3j$ normalization	0.009	-0.09
$\mu^+ + 3j$ normalization	0.007	-0.09
$\mu^+ + 2j$ normalization	0.006	-0.04
$\mu^- + 4j$ 2b tag	0.007	-0.16
$\mu^+ + 4j$ 1b tag	0.007	-0.15
$\mu^- + 3j$ 1b tag	0.005	-0.13
$e^- + 4j$ 1b tag	0.008	-0.19
$e^+ + 4j$ 1b tag	0.006	-0.21
$e^+ + 2j$ normalization	0.008	0.03
$e^- + 3j$ normalization	0.008	-0.06
$e^- + 2j$ normalization	0.007	-0.16
$e^+ + 3j$ normalization	0.007	-0.06
$t\bar{t}$ /tW modeling		
$t\bar{t}$ /tW interference	0.009	-0.32
Hadronizer (HERWIG++ VS PYTHIA 8)	0.004	0.14
Experimental		
Misidentification probability (shape)	0.008	-0.11
c identification probability (shape)	0.008	-0.00
b identification probability (shape)	0.005	-0.10
μ scale (rate)	0.010	0.04
JES Pileup (rate)	0.005	0.03

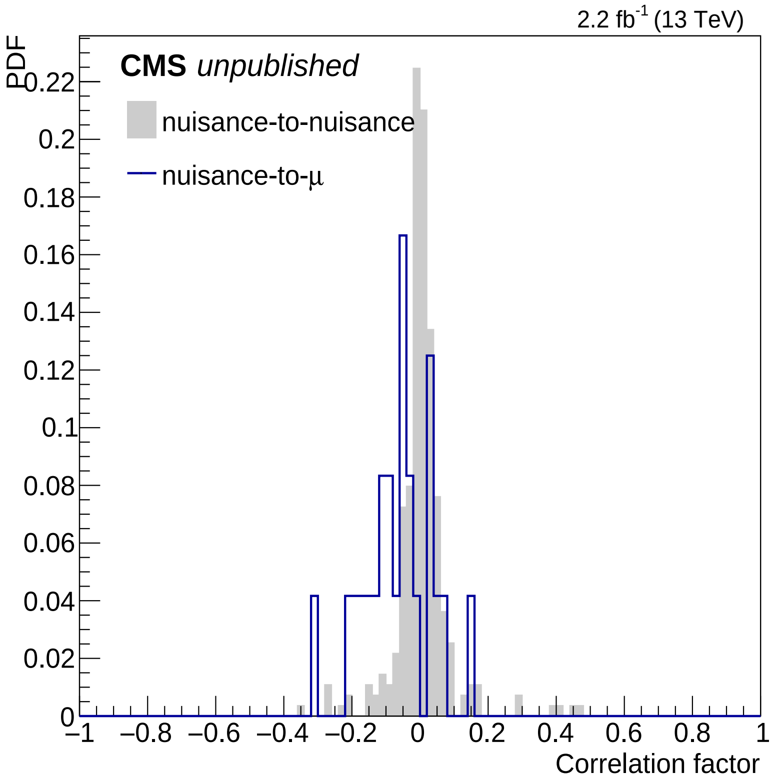


Figure 8.9: Distribution of the inter-nuisances correlation and correlation of the signal strength and nuisances. Only nuisances with an impact > 0.005 are shown.

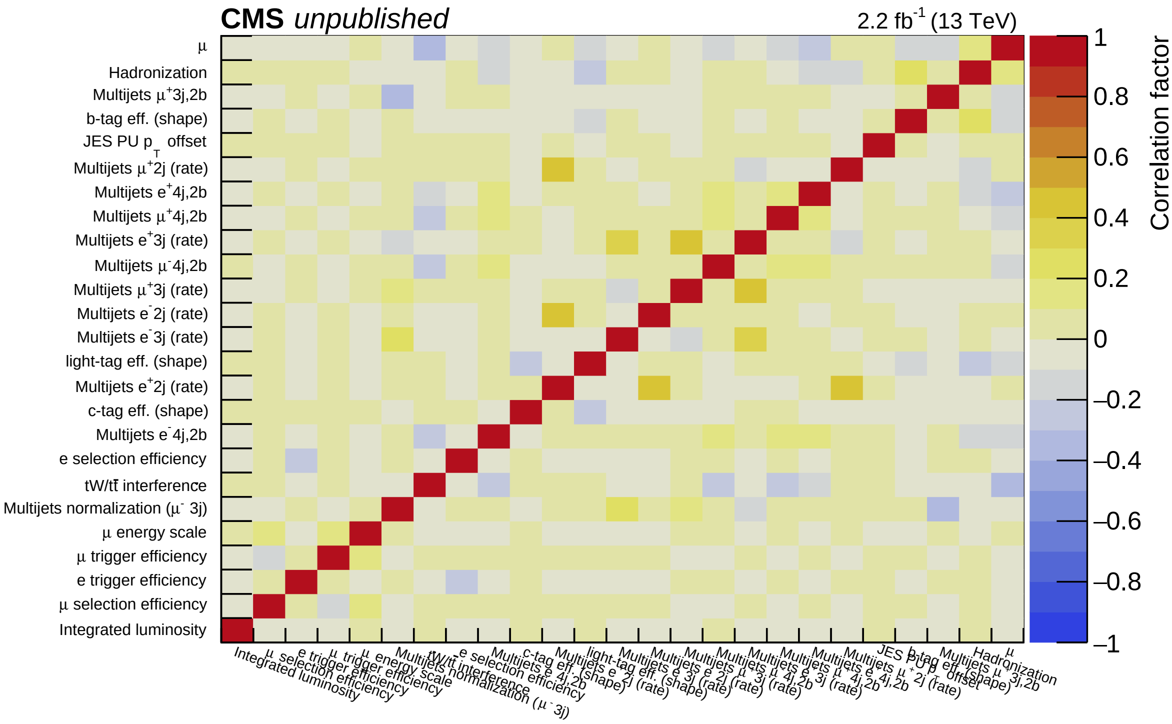


Figure 8.10: Postfit correlation matrix. The top row shows the correlation of the nuisances with the signal strength μ. A diamond signals rate-only uncertainties. Only nuisances with an impact > 0.005 are shown.

Shape or Distribution analysis

The other strategy of this analysis is shape analysis. In this strategy, the events are looked into finer detail by using a differential distribution which allows to seperate further the signal from the background. For events without any b-tagged jet the missing transverse energy distribution is used as it provides a handle on seperating the non-resonant QCD multijet background from the W+jets background, while constraining at the same time the normalization of the latter. For the categories with at least $1b$-tagged jet, the $M(\ell, b)$ while in case of more than $1b$-tagged jets, $\min M(\ell, b)$ is examined.

The results of the fits are reported in Table 8.5. Figure 8.11 explains the comparison of the post-fit distributions for 4 jets and at least 2 b-tagged jets, both for electron and muon channels. Figures 8.12 and 8.13 show the impact of the different sources of uncertainties in the fit for shape analysis. Finally, Figure 8.14 explains the post-fit correlations of the nuisances with higher impact and signal strength and Figure 8.15 shows the correlation of matrix of the nuisances.

Table 8.5: Expected and observed fit results for different categories and combinations in the shape analysis. The uncertainties are the total uncertainties (statistical+systematic) in the fiducial phase space.

Category	Expected	Observed
μ^+	$1.000^{+0.043}_{-0.043}$	$1.054^{+0.044}_{-0.043}$
μ^-	$1.000^{+0.042}_{-0.041}$	$1.061^{+0.044}_{-0.042}$
μ	$1.000^{+0.039}_{-0.037}$	$1.054^{+0.040}_{-0.038}$
e^+	$1.000^{+0.060}_{-0.057}$	$1.036^{+0.061}_{-0.059}$
e^-	$1.000^{+0.058}_{-0.055}$	$1.059^{+0.065}_{-0.056}$
e	$1.000^{+0.054}_{-0.052}$	$1.062^{+0.056}_{-0.052}$
+	$1.000^{+0.039}_{-0.039}$	$1.053^{+0.041}_{-0.040}$
-	$1.000^{+0.038}_{-0.037}$	$1.067^{+0.039}_{-0.038}$
≥2j	$1.000^{+0.036}_{-0.035}$	$1.072^{+0.038}_{-0.037}$
≥3j	$1.000^{+0.038}_{-0.037}$	$1.060^{+0.039}_{-0.038}$
≥4j	$1.000^{+0.058}_{-0.056}$	$1.024^{+0.061}_{-0.063}$
combined	$1.000^{+0.035}_{-0.034}$	$1.067^{+0.037}_{-0.035}$

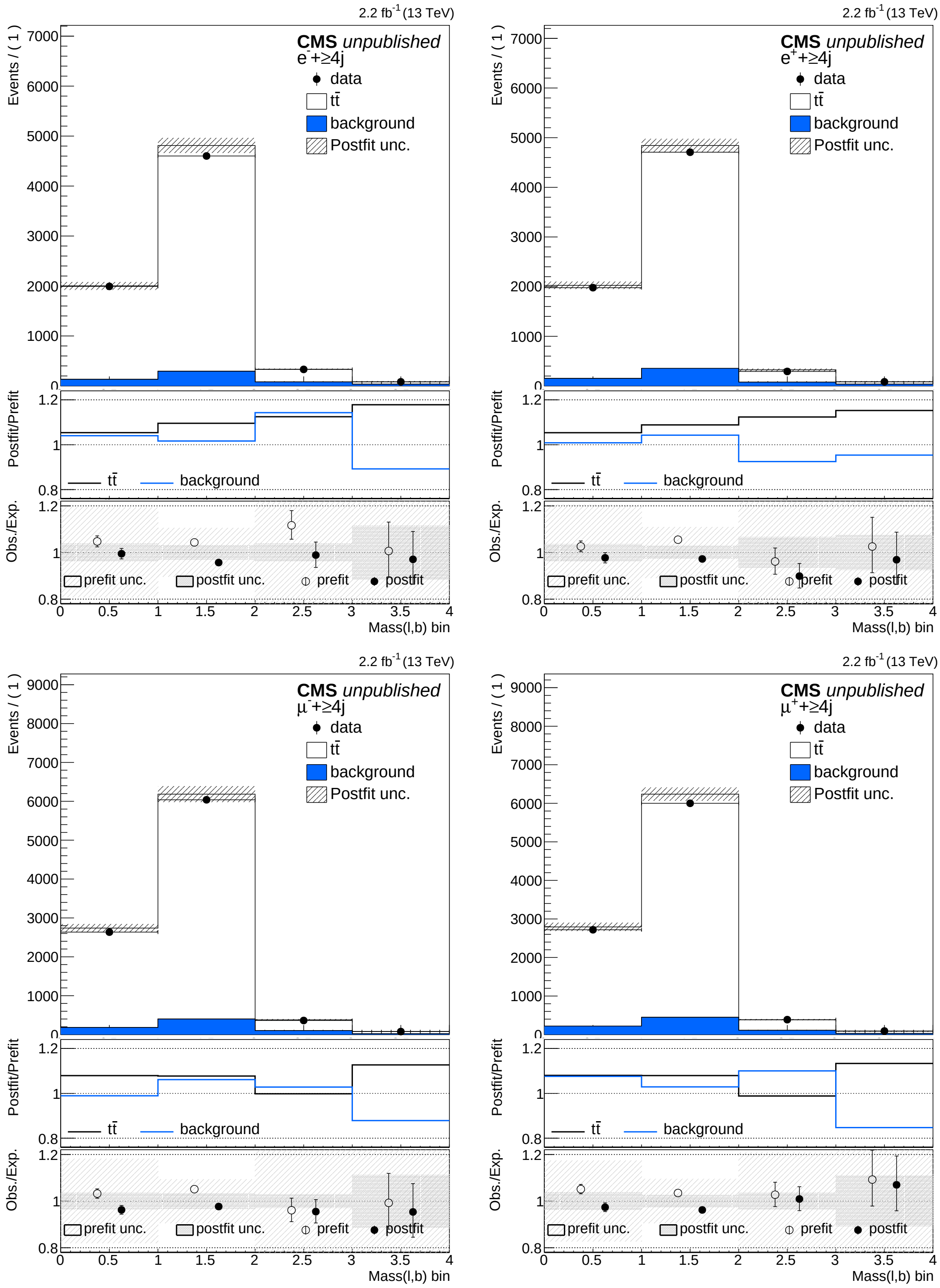


Figure 8.11: Comparison of the post-fit distributions with the data with events 4jets and at least 2 b-tagged jets for the e-channel (top) and for μ-channel (bottom). Negatively (positively) charged distributions are shown on left (right). The top panels show the stacked distributions of signal and background processes and the observed data. A shaded band represents the post-fit uncertainty. The middle panels show the ratio of the post-fit to pre-fit distributions for the signal and background sum. The bottom panel shows the data to signal plus background expectations, prior and after the fit. The error bar in the points is the statistics in data. A shaded band represents the relative postfit and prefit uncertainties.

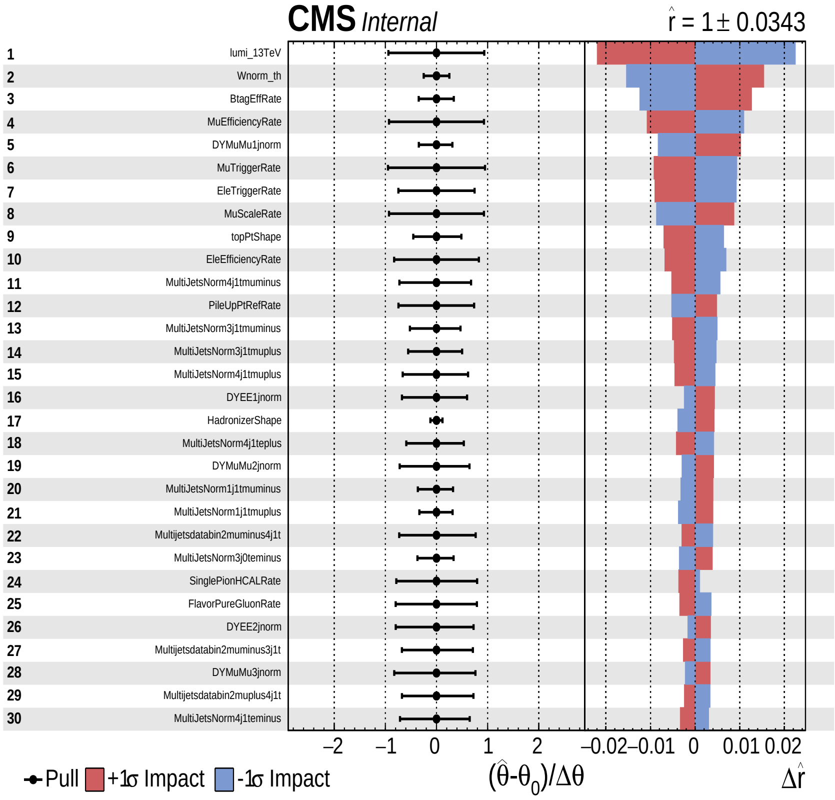


Figure 8.12: Summary of the fit results to the Asimov dataset using the shape analysis. The left panel shows the post-fit pull (value and uncertainty) of each nuisance, while the right panel displays the estimated impact on the fit for the $t\bar{t}$ signal strength. Only the first 30 nuisances are displayed and their name is shown at each row of the plots.

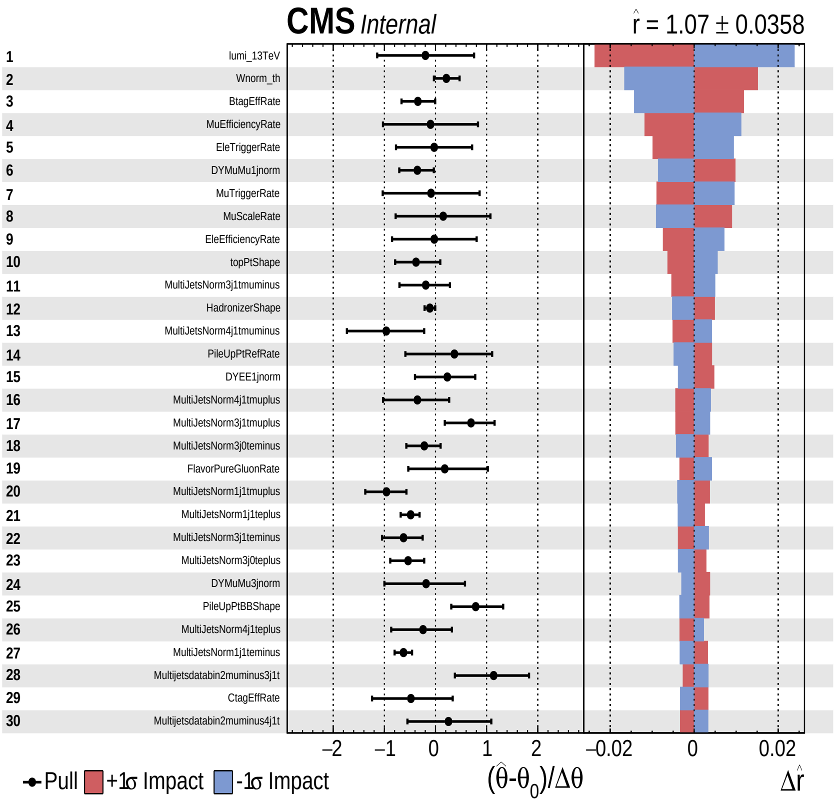


Figure 8.13: Summary of the fit results for data using the shape analysis. The left panel shows the post-fit pull (value and uncertainty) of each nuisance, while the right panel displays the estimated impact on the fit for the $t\bar{t}$ signal strength. Only the first 30 nuisances are displayed and their name is shown at each row of the plots.

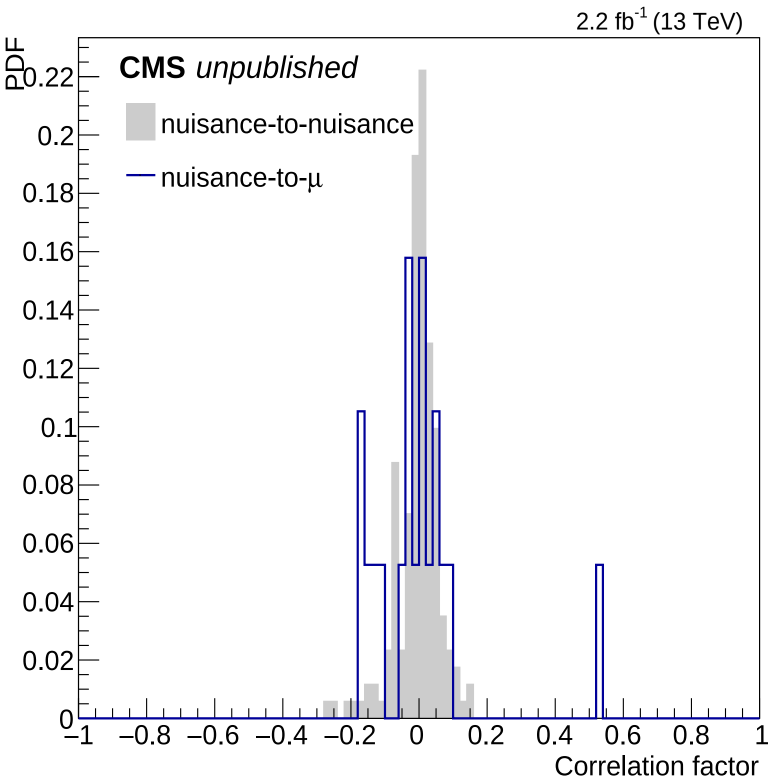


Figure 8.14: Distribution of the inter-nuisances correlation and correlation of the signal strength and nuisances using shape analysis. Only nuisances with an impact > 0.005 are shown.

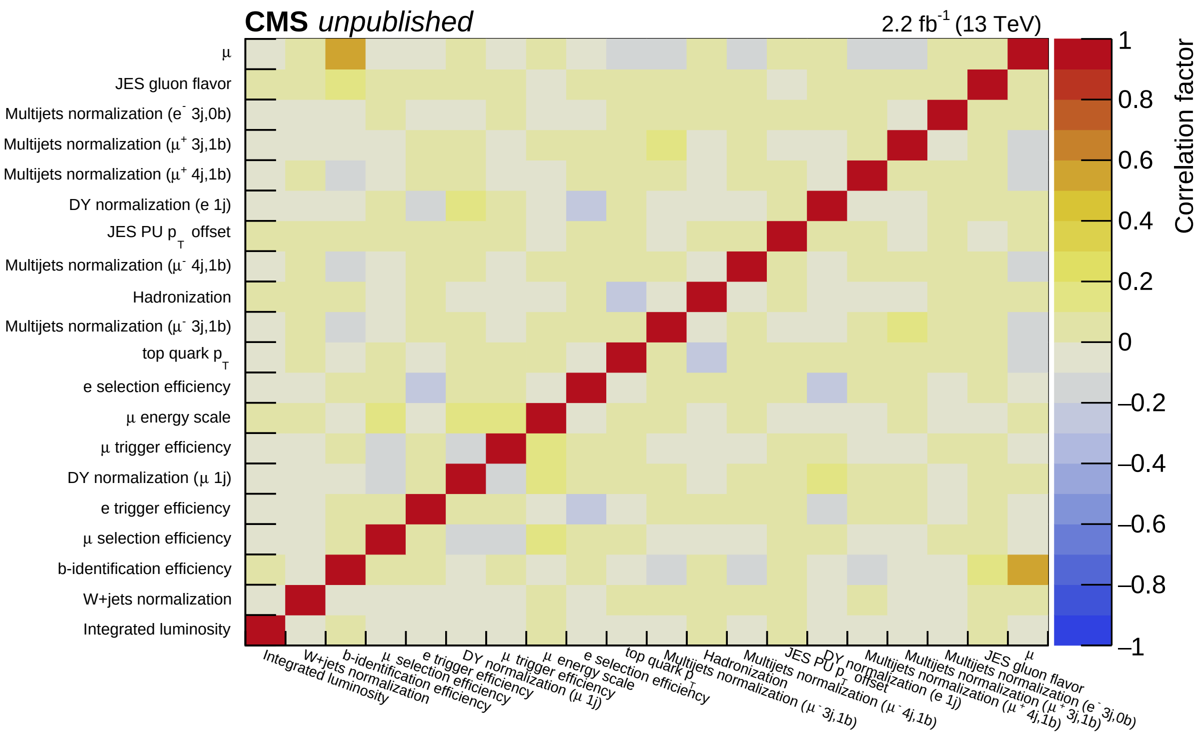


Figure 8.15: Postfit correlation matrix. The top row shows the correlation of the nuisances with the signal strength μ using shape analysis. A diamond signals rate-only uncertainties. Only nuisances with an impact > 0.005 are shown.

Table 8.6: List of nuisances with larger impact on the fit in the shape analysis. The middle column shows the estimated impact, while the right-most column shows the correlation with the signal strength in the fit.

Source	$\Delta\mu$	$\rho(\mu)$
Integrated luminosity	0.024	-0.03
Trigger/selection efficiency		
e trigger	0.010	-0.03
μ trigger	0.010	-0.04
μ eff.	0.012	-0.04
e eff.	0.007	-0.02
Background modelling		
W+jets (norm.)	0.017	0.09
DY$\rightarrow$ $\mu\mu$+1j (norm.)	0.01	0.03
DY$\rightarrow$ ee+1j (norm.)	0.005	0.01
Multijets μ^{-}+3j,1t	0.005	-0.17
Multijets μ^{-}+4j,1t	0.005	-0.18
Multijets μ^{+}+4j,1t	0.004	-0.14
Multijets μ^{-}+4j,1t	0.004	-0.14
Multijets μ^{+}+3j,1t	0.004	-0.13
Multijets e^{-}+3j,0t	0.004	0.01
$t\bar{t}$ modeling		
top p_T	0.006	-0.11
Hadronizer (shape)	0.005	0.08
Experimental		
μ scale	0.009	0.01
b identification probability (rate)	0.014	0.53
JES Pileup PtRef (rate)	0.005	0.06
Gluon response (rate)	0.004	0.04

8.4 Extrapolation to full phase space

The PLR fit is used to extract the number of $t\bar{t}$ signal events in the fiducial phase space region. However, the measurement of inclusive cross section requires an extrapolation, from the fiducial to the full phase space regions. Defining acceptance as:

$$A_{fid} = \frac{N_{fid}}{N_{gen}} \tag{8.6}$$

where $N_{fid} = \mathscr{L} \cdot \varepsilon \cdot \sigma_{fid}$ is the number of events in the fiducial region and $N_{gen} = \mathscr{L} \cdot \sigma_{tot}$ the number of generated events in the full fase space. The relation between the fiducial and full phase space cross sections is written as:

$$\sigma(t\bar{t}) = \frac{\sigma(t\bar{t})_{fid}}{A_{fid}} \ , \tag{8.7}$$

and the total cross section written as:

$$\sigma(t\bar{t}) = \frac{N_{fid}}{A_{fid} \cdot \varepsilon \cdot \mathscr{L}} \ , \tag{8.8}$$

In the Equation 8.8 N_{fid}, the numbar of events in fiducial region and selection efficiency ε, are determined experimentally. The ε comprises trigger, lepton, jet, b-jet selection efficiencies and is determined from control regions (tag-and-probe, QCD multijet samples) and further constrained in-situ in the fit. The $\mathscr{L}$ is determined independently [187] and incorporated into the fit as an uncertainty, which is also taken into account for the MC-based predictions. Also its effect needs to be externalized while determining the total signal cross section. Finally the acceptance is taken from MC and therefore studied in simulation.

In order to check the variation of acceptance against different choices of QCD scales and PDFs, different MC generator samples are studied. In the study, the event-based generator level weights of different MC samples are used as an alternative weight in the $t\bar{t}$ simulations. The number of events that pass a pre-selection criteria are counted. Given offline, one isolated lepton and one jet is required, the same requirements is replicated at generator level by requiring one p_T >25 GeV final state lepton and at least one generator level jet (anti k_T, R=0.4) with p_T >25 GeV. Both objects are required to have $|\eta| < 2.5$. Table 8.7 summarizes the uncertainties which are considered in the extrapolation to the full phase space.

Table 8.7: Summary of the uncertainties considered in the extrapolation to the full phase space.

Uncertainty	Procedure
PDF+α_S	RMS of the 100 NNPDF3.0 variations added in quadrature with PDF set with $\Delta\alpha_S = \pm 0.001$
QCD scale	Envelope of the variations $(\mu_R, \mu_F) = [(1/2, 1), (2, 1/2), (1, 1/2), (1, 2), (1/2, 1/2), (2, 2)]$
PS scale	Difference between nominal PS scale POWHEG +PYTHIA 8 and 2,1/2 variations
Hadronizer	Difference between PYTHIA8 and HERWIG ++
MC generator	Difference between POWHEG and MG5_AMC@NLO both interfaced to PYTHIA 8
Integrated luminosity	2.7%

In this study, the branching ratios in the MG5_AMC@NLO and MadGraph samples are also corrected. As W decays to leptons with branching ratio of 0.1086 [202], this is used to re-scale the geometric branching ratio assumed in both MG5_AMC@NLO and MadGraph of 1/9. Similar procedure applies to W hadronic decays which are expected to occur with a branching ratio of 0.6741 [202] while MG5_AMC@NLO and MadGraph assume 2/3. The correction applied is therefore $[0.1086/(1/9)]\cdot[0.6741/(2/3)] = 0.9883$ for the semi-leptonic channel. After applying this correction both POWHEG, MG5_AMC@NLO and MadGraph are found to be in agreement with respect to the acceptance prediction for this analysis.

Table 8.8 shows the evolution of the acceptance for different (inclusive) jet multiplicity requirements. The mass dependence of the acceptance is 0.00023/GeV (0.00065/GeV) after the requirement of at least 1 (4) jets. The variation of the QCD scale at parton shower level has no effect except for events with at least 4 jets. The fluctation of +0.003/-0.006 is expected from the up/down variation of the scale. The difference of hadronization between PYTHIA 8 and HERWIG ++ seems to be stable for MG5_aMC@NLO samples at the level of 1%.

Table 8.9 and Figure 8.16 summarize the evolution of the acceptance as function of the inclusive jet multiplicity.

8.4.1 Major uncertainties

The sources of uncertainties include PDF, α_s and QCD scales. For the PDF, the root mean square (RMS) of the 100 replicas of NNPDF30_nlo_as_0118 while for α_s, the envelope of the difference with respect to NNPDF30_nlo_as_0117 and NNPDF30_nlo_as_0119 is assigned as an uncertainty. For the QCD scales, the envelope of the individual variations of μ_R/μ_F by factors of 0.5 and 2 and the combined variation of both scales by a factor of 0.5 and 2.

The dominant source of uncertainty is expected to be due to PDFs. The total estimated uncertainty (adding linearly the sources) is expected to be $\mathcal{O}(0.37\%)$ ($\mathcal{O}(0.78\%)$) for events with at least 1 (4) jets.

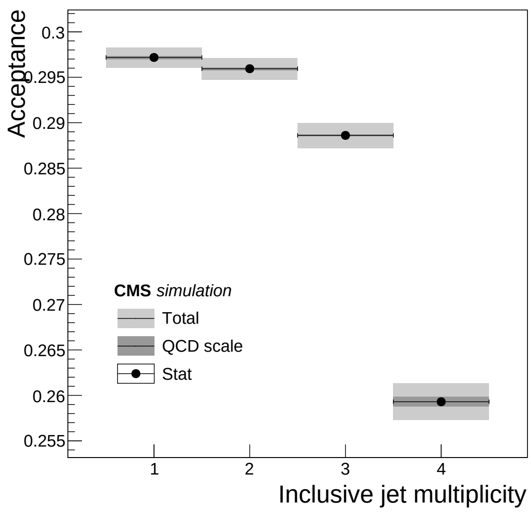


Figure 8.16: Acceptance as function of the inclusive jet multiplicity, as estimated with the $t\bar{t}$ POWHEG +PYTHIA 8 (reference simulation).

In this study, the acceptance is estimated to be 0.2972±0.0001 (stat)±0.0043 (syst), being the systematic uncertainty estimated by varying the QCD scale at matrix element (±0.0017), by taking the dispersion of the NNPDF3.0 PDF replicas

(± 0.0006), and by varying the parton shower used to interface the matrix element generator, i.e. PYTHIA 8 vs HERWIG++, (± 0.0039). The total uncertainty associated to the extrapolation is therefore estimated to be 1.5%.

Table 8.8: Estimated acceptance along with statistical uncertainty using different $t\bar{t}$ generators and hadronizers. Uncertainties are due to the limited statistics in the simulation.

Generator+Hadronizer	1 jet	2 jets	3 jets	4 jets
POWHEG +PYTHIA 8 m=169.5GeV	0.2964±0.0002	0.2950±0.0002	0.2874±0.0002	0.2572±0.0002
POWHEG +PYTHIA 8 m=172.5GeV	0.2972±0.0001	0.2959±0.0001	0.2886±0.0001	0.2593±0.0001
POWHEG +PYTHIA 8 m=175.5GeV	0.2978±0.0002	0.2966±0.0002	0.2897±0.0002	0.2611±0.0002
POWHEG +PYTHIA 8 scale up	0.2972±0.0002	0.2960±0.0002	0.2886±0.0002	0.2602±0.0002
POWHEG +PYTHIA 8 scale down	0.2971±0.0002	0.2958±0.0002	0.2880±0.0002	0.2578±0.0002
MG5_aMC@NLO+PYTHIA 8	0.2999±0.0003	0.2987±0.0002	0.2921±0.0003	0.2654±0.0002
MG5_aMC@NLO+HERWIG ++	0.3038±0.0003	0.3027±0.0003	0.2957±0.0002	0.2680±0.0002

Table 8.9: The acceptance and uncertainties on acceptance, QCD scales, PDF and α_s for POWHEG +PYTHIA 8 (reference simulation).

	1 jet	2 jets	3 jets	4 jets
Acceptance	0.2972±0.0001	0.2959±0.0001	0.2886±0.0001	0.2593±0.0001
QCD scales	0.0017	0.0016	0.0011	0.0005
PDF	0.0005	0.0006	0.0008	0.001
α_S	0.0003	0.0003	0.0003	0.0004

8.5 Cross section extraction

Figure 8.17 summarizes the results on signal strength as discussed in the previous sections, and Figure 8.18 summarizes the expected and observed impacts of the nuisances employed in the fit in the visible phase space region.

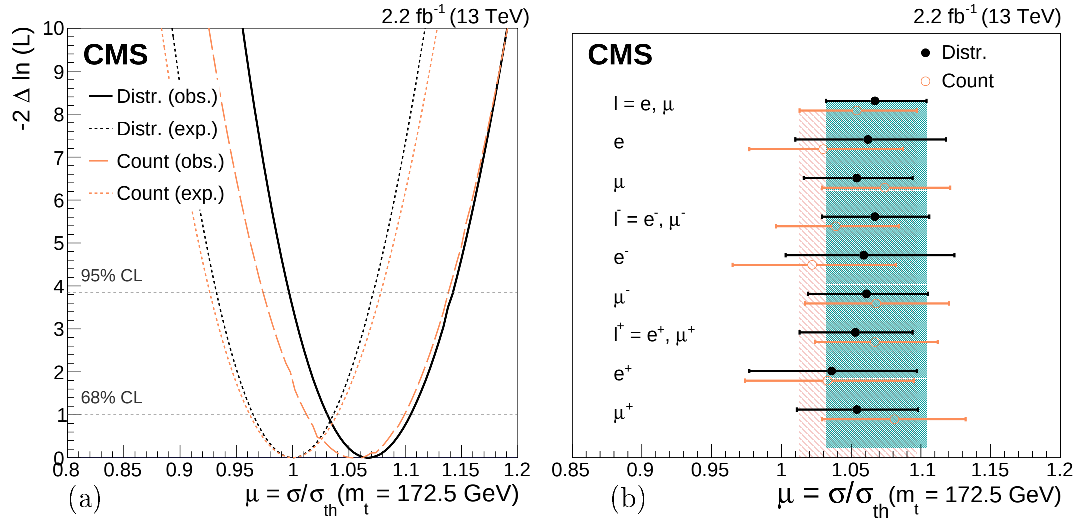


Figure 8.17: (a) The observed (solid curve) and expected (dashed curve) variation of the likelihood as a function of the signal strength μ for the distribution-based analysis. The expected curve is obtained by performing the fit using simulated events with $m_t = 172.5\ GeV$. For comparison, the corresponding curves for the counting cross-check analysis are also shown. The two horizontal lines represent the values in the PLR that are used to determine the 68% and 95% confidence level (CL) intervals for the signal strength. (b) Comparison of the values of the signal strength extracted for different combinations of events for the distribution-based default analysis (solid circles) and the cross-check counting analysis (open circles). The horizontal bars represent the total uncertainties, except the beam energy uncertainty. The shaded bands represent the uncertainty in the final combined signal strength obtained from the distribution-based and cross-check analyses.

The total (fiducial) cross section is measured by multiplying the signal strength and the theory cross section (and the estimated acceptance). The uncertainties in the acceptance are added in quadrature to the total cross section, which is extrapolation to the full phase space.

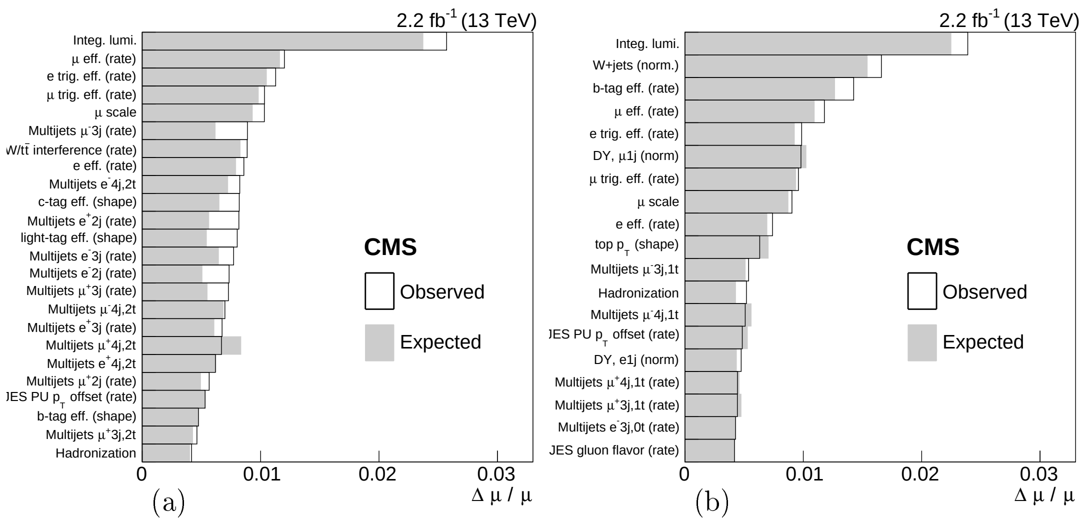


Figure 8.18: Summary of the impact of the most relevant nuisance parameters in the count (a) and analysis of distributions (b). The observed (hollow bars) impacts are compared to the expected (gray bars).

Chapter 9

Top quark mass measurement

As the $t\bar{t}$ cross section also depends on the top quark pole mass [203], therefore, it can be re-interpreted to measure this pole mass. The TOP++ (v2.0) program [168] and the CT14 NNLO PDF [204] set are used to parameterize the dependance of the cross section on the mass. The parametrization used is:

$$\sigma(m_t) = \sigma(m_{\mathrm{ref}}) \times \left(\frac{m_{\mathrm{ref}}}{m_t}\right)^4 \times \left[1 + a_1\left(\frac{m_t}{m_{\mathrm{ref}}} - 1\right) + a_2\left(\frac{m_t}{m_{\mathrm{ref}}} - 1\right)^2\right], \qquad (9.1)$$

where $m_{ref} = 172.5$ GeV is the mass used as reference, m_t is the pole mass to determine, a_1 and a_2 are coefficients determined after performing the calculations with various m_t hypotheses. The effects induced by the choice of factorization and renormalization scales $\mu_{\mathrm{R}}/\mu_{\mathrm{F}}$, the uncertainty in the PDF+α_{S}, and uncertainties in the beam energy are evaluated by recomputing the cross section after fluctuating these parameters within their uncertainties. The resulting typical uncertainties in $\sigma(m_t)$ amounts to $^{+2.5\%}_{-3.7\%}$, $^{+4.3\%}_{-4.4\%}$, and $\pm 0.23\%$, respectively. The latter reflects a 0.1% uncertainty in the beam energy at which the data used in this analysis has been collected [205]. Figure 9.1 shows the result of the parameterization of Equation 9.1 compared to the theory predictions obtained with TOP++.

To measure the pole mass, the likelihood function is re-parametrized, μ is transformed into a functional form depending on the top quark pole mass, as

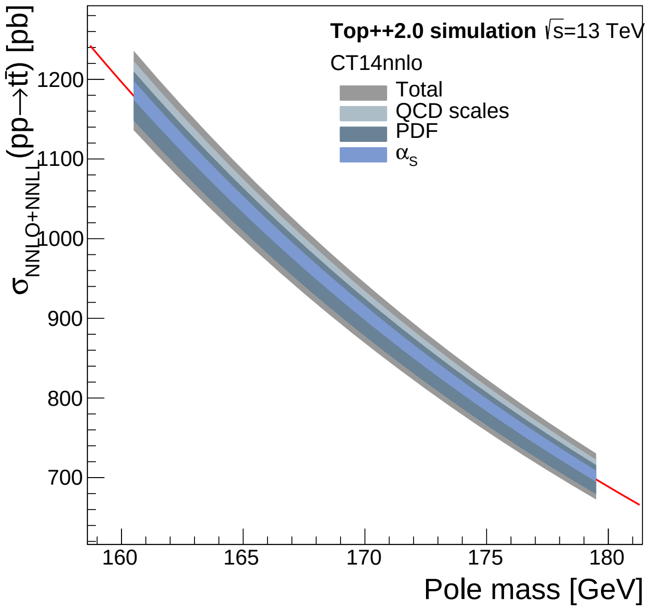


Figure 9.1: Evolution of the $t\bar{t}$ cross section at NNLO+NNLL with the top quark mass, using the CT14 NNLO PDF.

$$\mu(m_t) = \frac{\sigma(m_t)}{\sigma_{th}} \cdot \frac{\delta A}{A(m_t)} \tag{9.2}$$

where the last factor ($\frac{\delta A}{A(m_t)}$), is a mass-dependent correction to the acceptance A. Using simulated $t\bar{t}$ samples with different top quark masses, the acceptance is estimated to vary by 0.08% per $\Delta m_t = 1$ GeV. Figure 9.2 compares the parameterisation of Equation 9.2 and the acceptance correction. The uncertainties in the extrapolation, as well as the theoretical uncertainties that affect the cross section as a function of m_t due to the choices of μ_R/μ_F, PDF, α_S, and beam energy, are added as extra nuisance parameters in the fit for the pole mass. With the exception of μ_R/μ_F, which is defined through a log-uniform probability distribution, the remaining uncertainties are assigned a log-normal function [206]. After repeating the maximum-likelihood fit, the pole mass extracted is:

$$m_t = 170.6\ ^{+2.7}_{-2.7}\ GeV\ ,$$

where the quoted uncertainty contains both the statistical and systematic contributions. The result agrees with that obtained using the NNPDF3.0 NNLO

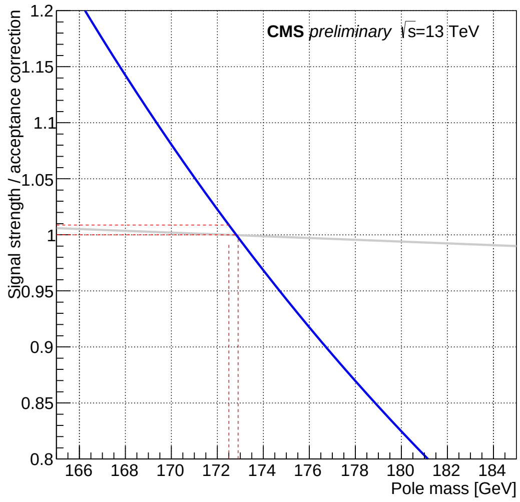


Figure 9.2: Model used for the extraction of m_t from the cross section measurement.

PDF [207]: $m_t = 170.3\ ^{+2.6}_{-2.7}$ GeV. The latter is only used as a cross-check, as the NNPDF3.0 PDF includes top-quark-related data in the determination of the proton PDFs. In both cases, the best-fit value is chosen by fixing the nuisance parameter associated with the choice of the μ_{R} and μ_{F} ratio to its post-fit value, and repeating the scan of the likelihood. This procedure is adopted to resolve the almost degenerate behavior of the likelihood induced through the use of a log-uniform parton distribution function (pdf) assigned to the choice of the μ_{R} and μ_{F} ratio.

Figure 9.3 shows the variation of the likelihood as a function of the top quark pole mass. For comparison, the expected likelihood from the Asimov set of nuisance parameters at $m_t = 172.5\ GeV$ is shown. Figure 9.4 compares different PDF sets, and the cut-in-categories with the shape analysis.

The shape analysis innovates by constraining in-situ the MC top mass and treating it as a nuisance, as proposed in [1]. Figure 9.5 depicts the improved sensitivity of the shape analysis to the MC mass with respect to a simple counting experiment. In both cases, the post-fit is observed to be in perfect agreement with the pre-fit. The (-1,1)-values of the nuisance correspond to a ± 3 GeV variation in the pole mass. The interpolation of the shape is based on a vertical morphing algorithm and the function used is non-linear, in particular for values in the [-1,1] range.

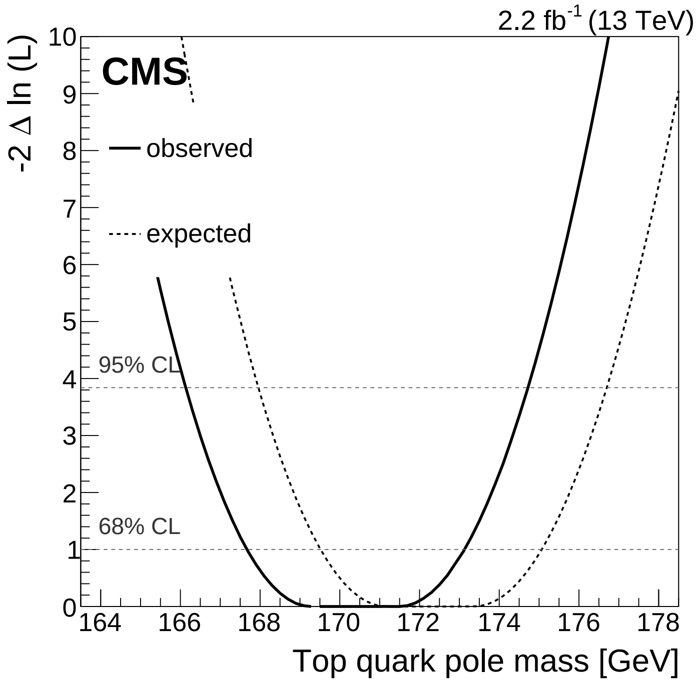
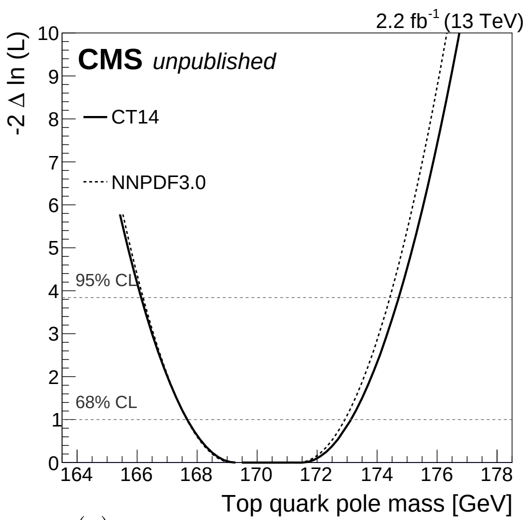
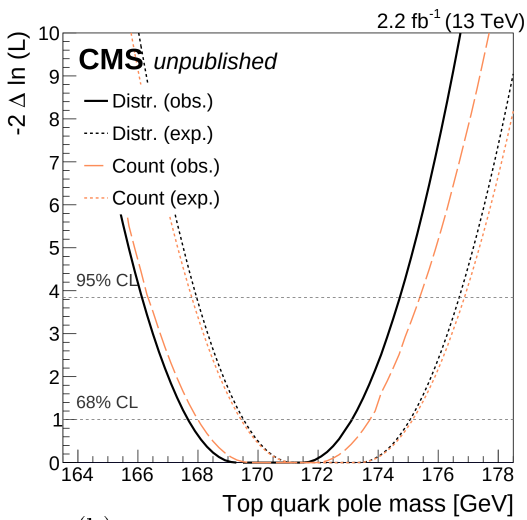


Figure 9.3: Dependence of the likelihood on the top quark pole mass (solid curve). The expected dependence from the simulation, using the a priori set of nuisance parameters with their expected values at $m_t = 172.5$ GeV, is shown for comparison as the dotted curve. The changes in the likelihood corresponding to the 68% and 95% confidence levels (CL) are shown by the dashed lines.



(a)



(b)

Figure 9.4: Dependence of the likelihood on the top quark pole mass (solid curve). The expected dependence from the simulation, using the a priori set of nuisance parameters with their expected values at $m_t = 172.5$ GeV, is shown for comparison as the dotted curve. The changes in the likelihood corresponding to the 68% and 95% confidence levels (CL) are shown by the dashed lines. Different PDF sets are compared (a) with the cut-in-categories and shape analysis (b).

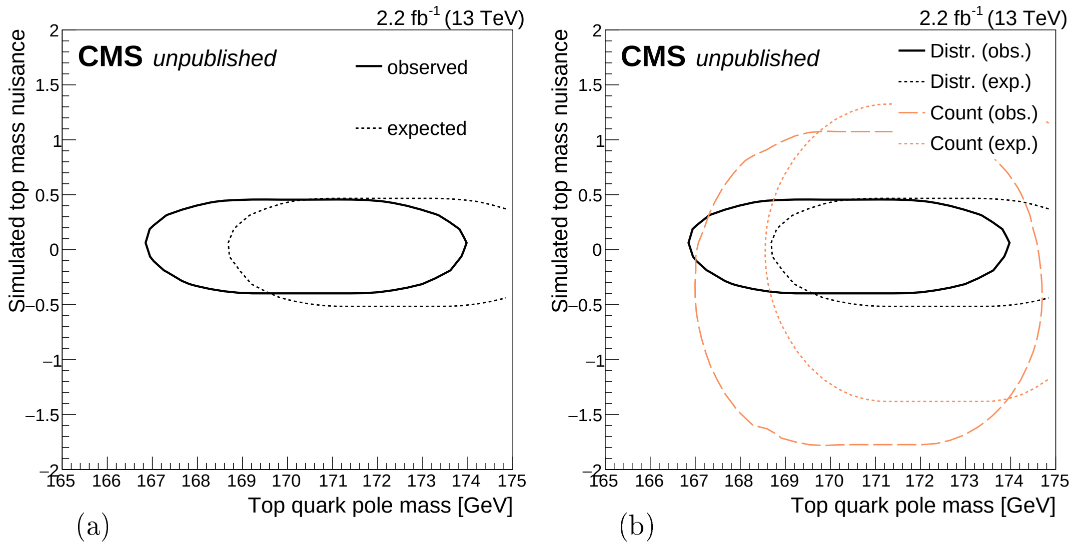


Figure 9.5: Two dimensional contours of the PLR corresponding to $2\Delta \ln L = 2.30$ (68% CL for a 2D fit) in the top quark pole mass versus MC mass nuisance plane. The expected (for $m_t = 172.5\ GeV$) and observed, using the CT14 PDF set, are compared for the analysis of the distributions (a) and the cut-in-categories and analysis (b).

The impact of each source of systematic uncertainty in the values corresponding to the fit is estimated using a similar procedure to the one described above for the cross section measurement. Table 9.1 summarizes the estimated uncertainties in the determination of m_t from the measured cross section.

Table 9.1: The source and value of the systematic uncertainties in the measurement of the pole mass of top quark (m_t).

Source	Δm_t [GeV]
Uncertainties from the fit in the fiducial region	-2.21 /$+2.45$
Extrapolation to the full phase space	-0.69 /$+1.09$
Beam energy	-0.08 /$+0.12$
$\mu_{\mathrm{R}}/\mu_{\mathrm{F}}$ and PDF$+\alpha_{\mathrm{S}}$	-0.89 /$+1.09$
Total	$+2.7$ /-2.7

Conclusion

This chapter is dedicated to the results for $t\bar{t}$ production corss section, measured at $\sqrt{s} = 13$ TeV in the semi-leptonic final state containing one isolated lepton (electron or muon) and at least one jet. The total integrated luminosity used for this study is 2.2 fb^{-1} for both the electron and muon channels. In addition, the pole mass is extracted at NNLO using the CT14 PDF set [90].

The alternative possibilities to measure the $t\bar{t}$ cross section have also been studied, including a complete assessment of the main systematic uncertainties. By performing a simultaneous fit to event distributions in 44 independent categories (section 8.3.2), the strength of the $t\bar{t}$ signal relative to the NNLO+NNLL [168] computation is measured. The signal strength is measured in a region of phase-space where the lepton has $p_T > 30$ GeV and $|\eta| < 2.1$, and at least one jet has $p_T > 30$ GeV and $|\eta| < 2.5$. The resulting visible $t\bar{t}$ cross section in this phase-space region is determined with an overall 3.6% relative uncertainty, a summary of results is shown in Table 9.2.

Table 9.2: Summary of the cross sections measured in this analysis.

Phase space	Analysis	Cross section [pb]
Fiducial	Count	$205.6\ \pm 0.4\ (stat.)\ ^{+6.9}_{-6.4}\ (syst.)\ \pm 4.7\ (lumi.)$
	Distr.	$208.2\ \pm 0.4\ (stat.)\ ^{+5.5}_{-4.9}\ (syst.)\ \pm 4.8\ (lumi.)$

The extrapolation to the full phase space is performed by using the acceptance estimated from the $t\bar{t}$ simulation. Using POWHEG [79], the acceptance is measured to be $0.2345 \pm 0.0001\,(\text{stat})^{+0.0044}_{-0.0043}\,(\text{syst})$ GeV, where the systematic uncertainty is due to change in $\mu_\text{R}/\mu_\text{F}$ (± 0.0017), with the CT14 PDF and α_S uncertainties ($^{+0.0009}_{-0.0007}$) GeV [204] and also changing the parton shower algorithm results into

($\pm$0.0039)% uncertainty. The total uncertainty associated with the extrapolation is estimated to be 1.6%. This uncertainty is added in quadrature to the systematic uncertainty obtained in the fitted fiducial region where the measurement is extrapolated to the full phase-space.

Summing the statistical (0.3%), systematic (2.7%), and integrated luminosity (2.3%) uncertainties in quadrature, a total relative uncertainty in the $t\bar{t}$ cross section of 3.9% is obtained. The final results are summarized in Table 9.3, the results measured for the muon and electron channels from the 13 TeV data are in agreement with the predictions of NNLO+NNLL [168, 208].

Table 9.3: Summary of the cross sections measured in this analysis [209].

Phase space	Analysis	Cross section [pb]
Full	Count	$876.9 \pm 1.7\ (stat.)\ ^{+29.5}_{-27.5}\ (syst.) \pm 16.1\ (extr.) \pm 20.2\ (lumi.)$
	Distr.	$887.7 \pm 1.7\ (stat.)\ ^{+23.0}_{-20.7}\ (syst.) \pm 16.3\ (extr.) \pm 20.4\ (lumi.)$

Figure 9.6 shows the comparison of this study with the other CMS measurements in dilepton channel and all jets channel.

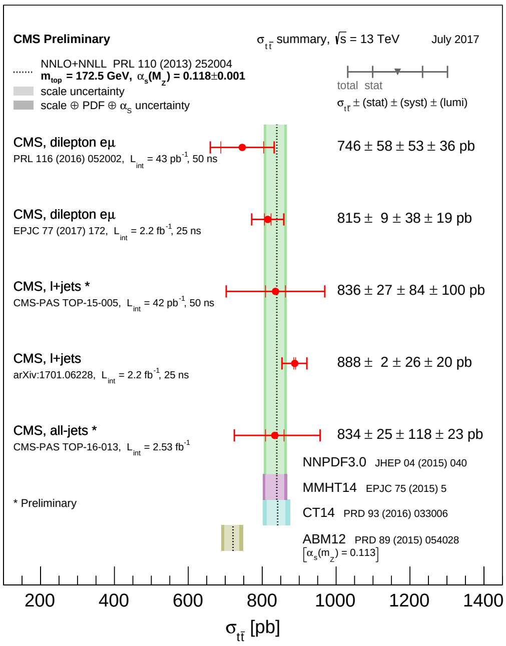


Figure 9.6: Summary of top quark pair production cross section [210].

The $t\bar{t}$ cross section $\sigma_{t\bar{t}}$ can be reinterpreted to extract the pole mass m_t of the top quark by using the dependence of the $\sigma_{t\bar{t}}$ on this parameter. The TOP++ program [168] and the CT14 NNLO PDF [204] is used to parametrize this dependence. After repeating the maximum-likelihood fit, the pole mass extracted is as [209],

$$m_t = 170.6 \, ^{+2.5}_{-2.9} \; GeV \; ,$$

where the quoted uncertainty contains both the statistical and systematic contributions. The result is in agreement with that obtained using the NNPDF3.0 NNLO PDF [163], i.e. $m_t = 172.3 \, ^{+2.4}_{-2.7}$ GeV. The latter is only used as a cross-check, as the NNPDF3.0 PDF includes top-quark-related data in the determination of the proton PDFs. In both cases, the best-fit value is determined by fixing the nuisance parameter associated with the choice of the μ_{R} and μ_{F} ratio to its post-fit value, and repeating the scan of the likelihood. This procedure is adopted to resolve the almost degenerate behavior of the likelihood, induced through the use of a log-uniform pdf assigned to the choice of the μ_{R} and μ_{F} ratio.

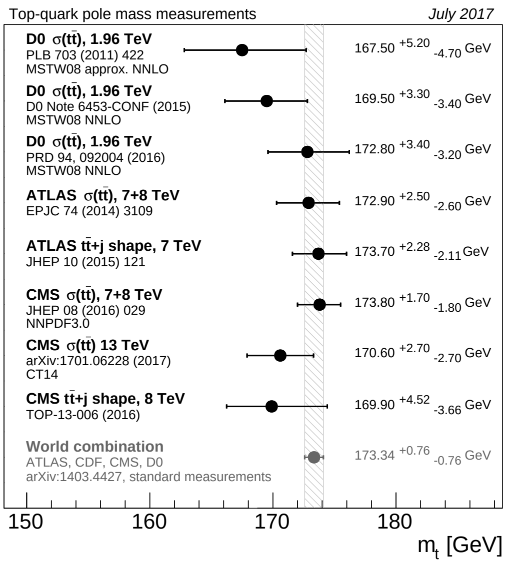


Figure 9.7: Summary of top quark pole mass [210].

The results obtained from this analysis are amongst the most precise measurements of the top quark cross section and can be used as a base line for many precise measurements i.e. extraction of the strong coupling constant α_s, ratio of the cross section at different center-of-mass energies, precise measurement of PDFs etc. The results can also be used in supersymmetry (SUSY) to study the supersymmetric top quark partner (stop). The precision of the results presented here, can further be improved for new LHC data by improving the techniques to constrain the lepton trigger, identification and isolation uncertainties.

Bibliography

[1] J. Kieseler, K. Lipka, and M. Sven-Olaf., Calibration of the top-quark Monte-Carlo mass. (2015), doi:10.1103/PhysRevLett.116.162001, arXiv:1511.00841.

[2] M. L. Mangano and J. Rojo., Cross Section Ratios between different CM energies at the LHC: opportunities for precision measurements and BSM sensitivity. *JHEP* 08 (2012) 010, doi:10.1007/JHEP08(2012) 010, arXiv:1206.3557.

[3] G. Ciezarek, M. F. Sevilla, B. Hamilton, R. Kowalewski, T. Kuhr, V. Luth and Y. Sato, A challange to lepton universality in B-meson decays. (2017), doi:10.1038/nature22346, arXiv:1703.01766.

[4] M. Czakon, A. Mitov, M. Papucci, J. T. Ruderman, and A. Weiler., Closing the stop gap. *Phys. Rev. Lett.*, 113(2014), no.20,201803, doi:10.1103/PhysRevLett.113.201803, arXiv:1407.1043.

[5] K. Beernaert, Measurement of spin correlations in $t\bar{t}$ systems in the muon+jets channel using a matrix element method with the CMS detector at the LHC. PhD thesis, Ghent University (2015).

[6] S. F. Novaes, Standard Model: An Introduction. (2000), arXiv:hep-ph/0001283.

[7] Y. Omar, Indistinguishable particles in quantum mechanics: An Introduction. (2005), arXiv:quant-ph/0511002.

[8] O. W. Greenberg., Spin and unitary spin independence in a paraquark model of baryons and mesons. Phys. Rev. Lett., doi:10.1103/PhysRevLett.13.598, 13, 598-602, (1964).

[9] M. Y. Han and Y. Nambu., Three-triplet model with double SU(3) symmetry. Phys. Rev. Lett.139, B1006-B1010, (1965).

[10] M. Gell-Mann, H. Fritzsch and H. Leutwyler., Advantages of the colour octet gluon picture. *Phys. Rev. Lett. B47*, 365–368, doi:10.1016/0370-2693(73)90625-4, (1973).

[11] C. Amsler et al., Review of particle physics. *Phys. Lett.*, B667 (2008) 1-1340 doi:10.1016/j.physletb.2008.07.018.

[12] T. Aoyama, M. Hayakawa, T. Kinoshita et al., Quantum electrodynamics calculation of lepton anomalous magnetic moments: Numerical approach to the perturbation theory of QED. PTEP 2012 (2012) 01A107. doi:10.1093/ptep/pts030.

[13] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716 (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[14] CDF Collaboration, Observation of top quark production in $p\bar{p}$ collisions. *Phys. Rev. Lett.*, 74 (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[15] D0 Collaboration, Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.8 TeV. *Phys. Rev. Lett.*, 74 (1995) 2422–2426, doi:10.1103/PhysRevLett.74.2422, arXiv:hep-ex/9411001.

[16] R. Feynman, F. Morinigo, and W. Wagner et al., Feynman lectures on gravitation (Frontiers in physics S). ISBN-13: 978-0813340388.

[17] A. Diaferio, S. Schindler, and K. Dolag., Clusters of galaxies: Setting the stage. *Space Science Reviews* 134 (2008) 7–24, doi:10.1007/s11214-008-9324-5, arXiv:0801.0968.

[18] Planck Collaboration, Planck 2013 results. XVI. Cosmological parameters. doi:10.1051/0004-6361/201321591, arXiv:1303.5076.

[19] A. Blondel, E. Graverini, N. Serra, M. Shaposhnikov., Search for Heavy Right Handed Neutrinos at the FCC-ee. arXiv:1411.5230.

[20] S. Herb, D. Hom, L. Lederman, J. Sens, and H. Snyder et al., Observation of a dimuon resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions, *Phys. Rev. Lett.*, 39 (1977) 252–255, doi:10.1103/PhysRevLett.39.252.

[21] CDF Collaboration, Observation of top quark production in $p\bar{p}$ collisions. *Phys. Rev. Lett.*, 74 (1995) 2626-2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[22] D0 Collaboration, Observation of the top quark. *Phys. Rev. Lett.*, 74 (1995) 2632-2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.

[23] D0 Collaboration, A direct measurement of the total decay width of the top quark. doi:10.1103/PhysRevLett.111.202001, arXiv:1308.4050.

[24] A. Salam, J. Goldstone and S. Weinberg., Broken symmetries. *Prog. Rev.*, 127, 965–970 (1962).

[25] S. Peris, R. D. Peccei and X. Zhang., Non Standard couplings of the top-quark and precision measurements of the electroweak theory. *Nucl. Phys.*, B349, 305–322, (1991), doi:10.1016/0550-3213(91)90327-T.

[26] E. Laenen, J. Smith, and W. L. Van Neerven., Top quark production cross section. *Phys. Lett.*, B321 :254–258, (1994). doi:10.1016/0370-2693(94)90473-1, arXiv:hep-ph/9310233.

[27] S. Moch and P. Uwer., Theoretical status and prospects for top-quark pair production at hadron colliders. *Phys. Rev. D*, 78:034003, Aug. (2008), doi:10.1103/PhysRevD.78.034003, arXiv:0804.1476.

[28] E. Perez and E. Rizvi, The quark and gluon structure of the proton. (2012), arXiv:1208.1117v2.

[29] M. Czakon, P. Fiedler and A. Mitov., Total top-quark pair-production cross section at hadron colliders through $O(\alpha_S^4)$. *Phys. Rev. Lett.*, 110:252004, (2013), doi:10.1103/PhysRevLett.110.252004.

[30] TOP LHC working group. twiki page.

[31] F. Cascioli, S. Kallweit, P. Maierhofer and S. Pozzorini, A unified NLO description of top-pair and associated Wt production. *Eur. Phys. J. C*, vol. 74, no. 3, (2013), doi:10.1140/epjc/s10052-014-2783-9, arXiv:1312.0546.

[32] S. Frixione, E. Laenen, P. Motylinski, B. Webber and C. D. White., Single top hadroproduction in association with a W boson. *JHEP* 07 (2008) 029, doi:10.1088/1126–6708/2008/07/029, arXiv:0805.3067.

[33] T. M. P. Tait., The tW^- mode of single top production. *Phys. Rev. D* 61 (1999) 034001, doi:10.1103/PhysRevD.61.034001.

[34] W. Bernreuther., Top quark physics at the LHC. *J. Phys. G*, G35:083001, (2008), doi:10.1088/0954-3899/35/8/083001, arXiv:0805.1333.

[35] CMS Collaboration, Measurement of the ratio B(t→Wb)/B(t→Wq). CMS Physics Analysis Summary CMS-PAS-TOP-12-035, (2012).

[36] V. Khachatryan et al., Measurement of the t-channel single top quark production cross section and of the $|V_{tb}|$ CKM matrix element in pp Collisions at $\sqrt{s}$ = 8 TeV. *JHEP*, (2014) doi:10.1007/JHEP06(2014)090 arXiv:1403.7366.

[37] K. A. Olive et al., Review of particle physics. Chin.Phys. vol. C38 p. 090001, (2014).

[38] F. Jegerlehner, M. Yu. Kalmykov, and B. A. Kniehl., Selfconsistence of the Standard Model via the renormalization group analysis. *J. Phys. Conf. Ser.*, 608(1:012074, (2015), arXiv:1412.4215.

[39] CMS Collaboration, Measurement of b-tagging efficiency using $t\bar{t}$ events. CMS Physics Analysis Summary CMS-PAS-BTV-11-003, (2011).

[40] CMS Collaboration, Identification of b-quark jets with the CMS experiment. *JINST* 8 (2013) 04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[41] L. Evans and P. Bryant., LHC Machine. *JINST* 3 (2008) doi:10.1088/1748-0221/3/08/S08001.

[42] CMS Collaboration, The CMS experiment at the CERN LHC. *JINST* 3 (2008) doi:10.1088/1748-0221/3/08/S08004.

[43] Fermilab, Design report Tevatron 1 project. *FERMILAB DESIGN* 01 (1983).

[44] C. Bulletin, The particle suppliers. Les fournisseurs de particules. https://cds.cern.ch/record/1255151.

[45] In Pictures, The Higgs boson phenomenon. http://www.pcworld.idg.com.au/slideshow/429544/pictures-higgs-boson-phenomenon/?image=13.

[46] C. Lefevre, LHC: the guide (English version). Guide du LHC (version anglaise). (February 2009).

[47] ATLAS Collaboration, ATLAS: Detector and physics performance technical design report. Volume 1, CERN-LHCC-99-14, *ATLAS-TDR-14*, (1999).

[48] ATLAS Collaboration, ATLAS: Detector and physics performance technical design report. Volume 2, CERN-LHCC-99-15, *ATLAS-TDR-15*, (1999).

[49] CMS Collaboration, CMS Physics:Technical Design Report Volume 1: Detector Performance and Software, Technical Design Report CERN-LHCC-2006-001, *CMS-TDR-008-1*, Geneva, (2006).

[50] CMS Collaboration, CMS technical design report, volume II: Physics performance, *J. Phys*. G34 (2007) 995-1579. doi:10.1088/0954-3899/34/6/S01.

[51] ALICE Collaboration, Heavy ion collisions at the LHC: the Alice experiment, arXiv:hep-ph/9612221.

[52] ALICE Collaboration, ALICE technical design report: Detector for high momentum PID, CERN–LHCC–98–19, (1998).

[53] LHCb Collaboration, LHCb technical design report: Reoptimized detector design and performance, CERN–LHCC–2003–030, (2003).

[54] LHCf Collaboration, O. Adriani et. al., The LHCf detector at the CERN Large Hadron Collider, *JINST 3 (2008) S08006*.

[55] TOTEM Collaboration, G. Anelli et. al., The TOTEM experiment at the CERN Large Hadron Collider, *JINST 3 (2008) S08007*.

[56] M. D. Negra et. al., CMS physics technical design report. CERN/LHCC 2006–001, *CMS TDR 8.1, (2006)*.

[57] V Karimaki., The CMS tracker system project technical design report. CERN-LHCC–98–006. *CMS–TDR–5, (1997)*.

[58] CMS Collaboration, S. Chatrchyan et. al., Commissioning and performance of the CMS pixel tracker with cosmic ray muons. *JINST 5 (2010) T03007*, doi:10.1088/1748-0221/5/03/T03007, arXiv:0911.5434.

[59] D. Sprenger, M. Weber, R. Adolphi, R. Brauer, L. Feld, et al., Validation of Kalman Filter alignment algorithm with cosmic-ray data using a CMS silicon strip tracker endcap. *JINST* 5 (2010) P06007, doi:10.1088/1748-0221/5/06/P06007, arXiv:1003.5460.

[60] CMS Collaboration, CMS Physics : Technical design report Volume 1: Detector Performance and Software. Technical Design Report CMS. CERN, Geneva, 2006. http://cdsweb.cern.ch/record/922757.

[61] CMS ECAL Collaboration, W. Lustermann., The electromagnetic calorimeter of CMS, summary and status, *J. Phys. Conf. Ser.*, 160 (2009) 012044 hrefhttp://iopscience.iop.org/article/10.1088/1742-6596/160/1/012044/metadoi:10.1088/1742-6596/160/1/012044.

[62] CMS Collaboration, The CMS electromagnetic calorimeter project: Technical Design Report. CERN, Geneva, (1997). http://cdsweb.cern.ch/record/349375.

[63] CMS Collaboration, Search for heavy stable charged particles with 100 inverse picobarns and 1 inverse femtobarn in the CMS experiment. CMS Physics Analysis Summary CMS-PAS-EXO-08-003, (2009).

[64] CMS Collaboration, Missing transverse energy performance of the CMS Detector. *JINST*, (2011) doi:10.1088/1748-0221/6/09/P09001, arXiv:1106.5048.

[65] G. L. Bayatian et al., CMS physics: Technical design report. CERN–LHCC–2006–001, CMS–TDR–008–1, (2006).

[66] CMS Collaboration, The TriDAS Project, Technical Design Report, Volume 1: The Trigger Systems. CERN/LHCC 2000–038, *CMS-TDR-6.1*, (2000).

[67] CMS Collaboration, The TriDAS Project, Technical Design Report, Volume 2: Data Acquisition and High–Level Trigger. CERN/LHCC 02–26, CMS-TDR-6.2, (2002).

[68] T. Lenzi, G. De Lentdecker., Development and study of different muon track reconstruction algorithms for the level-1 trigger for CMS muon upgrade with GEM detectors https://inspirehep.net/record/1236817/plots, arXiv:1306.0858.

[69] WLCG Collaboration, LHC computing grid: Technical Design Report. CERN-LHCC-2005-024, LCG-TDR-001, (2005).

[70] CMS Luminosity-Public Results. LumiPublicResults.

[71] T. Sjőstrand., Monte Carlo Tools. arXiv:0911.5286, (2009).

[72] M. H. Seymour and M. Marx., Monte Carlo event generators. arXiv:1304.6677, (2013).

[73] S. Hőche., Introduction to parton-shower event generators. https://inspirehep.net/record/1328513/plots arXiv:1411.4085, (2015).

[74] T. Stelzer and W. F. Long., Automatic generation of tree level helicity amplitudes. *Nucl. Phys. Proc. Suppl.*, 37B:158–162, (1994) doi:10.1016/0010-4655(94)90084-1, arXiv:hep-ph/9401258.

[75] T. Sjostrand, S. Mrenna, and P. Z. Skands., A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, (2008) doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[76] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hőche et al., General purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, (2011) doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[77] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, vol. 1407, p. 079, (2014), doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[78] R. Frederix and S. Frixione., Merging meets matching in MC@NLO. *JHEP*, vol. 1212, p. 061, (2012), doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[79] S. Frixione, P. Nason, and G. Ridolfi., A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, vol. 0709, p. 126, (2007), doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[80] S. Frixione and B. R. Webber., Matching NLO QCD computations and parton shower simulations. *JHEP*, vol. 0206, p. 029, (2002), doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[81] S. Frixione, P. Nason and B. R. Webber., Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, vol. 0308, p. 007, (2003), doi:10.1088/1126-6708/2003/08/007, arXiv:hep-ph/0305252.

[82] P. Nason., A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, vol. 0411, p. 040, (2004), doi:10.1088/1126-6708/2004/11/040, hep-ph/0409146.

[83] S. Frixione, P. Nason and C. Oleari., Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, vol. 0711, p. 070, (2007), doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[84] S. Alioli, P. Nason, C. Oleari, and M. Re., A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, vol. 1006, p. 043, (2010), doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[85] E. Boos et al., COMPHEP 4.4: Automatic computations from lagrangians to events. *Nucl. Instrum. Meth.*, A534:250–259, (2004), doi:10.1016/j.nima.2004.07.096, arXiv:hep-ph/0403113.

[86] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjőstrand., Parton fragmentation and string dynamics. *Phys. Rept.*, 97 (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[87] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri et al., HERWIG 6.5 release note. (2002), arXiv:hep-ph/0210213.

[88] T. Gleisberg, S. Hőche, F. Krauss, M. Schőnherr, S. Schumann, et al., Event generation with SHERPA 1.1. *JHEP*, vol. 0902, p. 007, (2009), doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

[89] GEANT4 Collaboration, GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506, (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[90] S. Dulat, T. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump and C. P. Yuan., New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev.*, D 93, 033006, (2016), doi:10.1103/PhysRevD.93.033006, arXiv:1506.07443.

[91] G. Altarelli and G. Parisi., Asymptotic freedom in parton language. *Nucl. Phys.*, B126:298, (1977), doi:10.1016/0550-3213(77)90384-4.

[92] V. N. Gribov and L. N. Lipatov., Deep inelastic e-p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, (1972).

[93] Yu. L. Dokshitser., Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics. *Sov. Phys. JETP*, 46:641–653, (1977).

[94] T. Kinoshita., Mass singularities of feynman amplitudes. *J. Math. Phys.*, 3:650–677, (1962), doi:10.1063/1.1724268.

[95] T. D. Lee and M. Nauenberg., Degenerate systems and mass singularities. *Phys. Rev.*, 133:B1549–B1562, (1964).

[96] S. Catani and M. H. Seymour., A General algorithm for calculating jet cross-sections in NLO QCD. *Nucl. Phys.*, B485:291–419, (1997), doi:10.1016/S0550-3213(96)00589-5, arXiv:hep-ph/9605323.

[97] David A. Kosower., Antenna factorization of gauge theory amplitudes. *Phys. Rev.*, D57:5410–5416, (1998), doi:10.1103/PhysRevD.57.5410, arXiv:hep-ph/9710213.

[98] F. Cascioli, P. Maierhofer and S. Pozzorini., Scattering amplitudes with open loops. *Phys. Rev. Lett.*, 108:111601, (2012), doi:10.1103/PhysRevLett.108.111601, arXiv:1111.5206.

[99] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde et al., An automated implementation of on-shell methods for one-loop amplitudes. *Phys. Rev.*, D78:036003, (2008), doi:10.1103/PhysRevD.78.036003, arXiv:0803.4180.

[100] V. Hirschi, R. Frederix, S. Frixione, M. V. Garzelli, F. Maltoni et al., Automation of one-loop QCD corrections. *JHEP*, 1105:044, (2011), doi:10.1007/JHEP05(2011)044, arXiv:1103.0621.

[101] V. V. Sudakov., Vertex parts at very high-energies in quantum electrodynamics. *Sov. Phys. JETP*, 3 (1956) 65–71.

[102] J. Alwall et al., A standard format for Les Houches Event Files. *Comput. Phys. Commun* 176 (2007) 300–304, doi:10.1016/j.cpc.2006.11.010, arXiv:hep-ph/0609017.

[103] S. Frixione, F. Stoeckli, P. Torrielli, B. R. Webber and C. D. White., The MCatNLO 4.0 Event Generator. (2010). arXiv:1010.0819.

[104] S. Hőeche et al., Matching parton showers and matrix elements. (2006), arXiv:hep-ph/0602031.

[105] L. Mangano, M. Mauro, F. Piccinini, M. Treccani., Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP* 07 (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[106] B. Andersson, S. Mohanty and F. Soderberg., The lund fragmentation process for a multi-gluon string according to the area law. *Eur. Phys. J*, 21 (2001)

631–647, doi:10.1007/s100520100757, arXiv:0106185.

[107] M. G. Bowler., e^+e^- production of heavy quarks in the string model. *Z. Phys.*, C11:169, (1981).

[108] CMS Collaboration, Underlying event measurements with leading particles and jets in pp collisions at $\sqrt{s}$ = 13 TeV. CMS Physics Analysis Summary CMS-PAS-FSQ-15-007, (2015).

[109] M. Sandhoff and P. Z. Skands., Colour annealing - A toy model of colour reconnections. (2005). *FERMILAB-CONF-05-518-T*.

[110] S. Argyropoulos and T. Sjőstrand., Effects of color reconnection on $t\bar{t}$ final states at the LHC. *JHEP*, 11:043, (2014), doi:10.1007/JHEP11(2014)043, arXiv:1407.6653.

[111] CMS Collaboration, Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s}$ = 2.76 TeV. (2015), doi:10.1140/epjc/s10052-016-4083-z, arXiv:1512.06212.

[112] CMS Collaboration, Particle-Flow event reconstruction in CMS and performance for jets, taus, and MET. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[113] CMS Collaboration, Commissioning of the Particle-Flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-002 (2010).

[114] S. Chatrchyan et al., Description and performance of track and primary vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, (2014), doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[115] E. Chabanat and N. Estre., Deterministic annealing for vertex finding at CMS,. C04-09-27, p.287-290, (2005).

[116] R. Fruhwirth, W. Waltenberger, and P. Vanlaer., Adaptive vertex fitting. *J. Phys.*, G34:N343, (2007).

[117] F. Beaudette., Performance of the Particle Flow Algorithm in CMS. Presentation at ICHEP–35^{th} International Conference on High Energry Physics, Paris, July (2010)..

[118] A. Wolfgang, B. Mangano, T. Speer and T. Todorov., Track reconstruction in the CMS tracker. CMS-NOTE-2006-041, (2006).

[119] R. E. Kalman., A new approach to linear filtering and prediction problems. *ASME–Journal of Basic Engineering 82(1)*, 35–45 (1960).

[120] S. Chatrchyan et al., The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, (2013), doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.

[121] S. Chatrchyan et al., Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, (2012), doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[122] M. Maes., Measurement of the top quark pair production cross section at the LHC with the CMS experiment. PhD thesis, PhD thesis, (2013).

[123] CMS Collaboration, Particle-flow commissioning with muons and electrons from J/Ψ and W events at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-003, (2010).

[124] S. Baffioni et al., Electron reconstruction in CMS. *Eur. Phys. J.* C49 (2007),1099–1116, doi:10.1140/epjc/s10052-006-0175-5.

[125] CMS Collaboration, Electron reconstruction and identification at $\sqrt{s} = 7$ TeV. CMS Physics Analysis Summary CMS-PAS-EGM-10-004, (2010).

[126] CMS Collaboration, Electron Reconstruction within the Particle Flow Algorithm. CMS Analysis Note AN-10-034, (2010).

[127] W. Adam, R. Frűhwirth, A. Strandlie, and T. Todorov., Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics* 31, N9 (2005).

[128] A. Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss et al., TMVA–Toolkit for Multivariate Data Analysis. PoS, ACAT:040, (2007), arXiv:physics/0703039.

[129] M. Cacciari and G. P. Salam., Dispelling the N^3 myth for the k_T jet-finder. *Phys. Lett. B*, 641:57, (2006), doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.

[130] M. Cacciari, G. P. Salam, and G. Soyez., The anti-k_T jet clustering algorithm. *JHEP*, 04:063, (2008), doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[131] S. D. Ellis and D. E. Soper., Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, (1993), doi:10.1103/PhysRevD.48.3160, arXXiv:hep-ph/9305266.

[132] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber., Better jet clustering algorithms. *JHEP*, 9708:001, (1997), doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[133] M. Wobisch and T. Wengler., Hadronization corrections to jet cross-sections in deep inelastic scattering. (1999). arXiv:hep-ph/9907280.

[134] CMS Collaboration, Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST* 6, (2011) P11002 doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[135] A. Perloff (on behalf of the CMS Collaboration), Pileup measurement and mitigation techniques in CMS. *J. Phys.: Conf. Ser.*, 404 (2012) 012045, doi:10.1088/1742-6596/404/1/012045.

[136] M. Cacciari and G. P. Salam., Pileup subtraction using jet areas. *Phys. Lett. B*, 659 (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[137] M. Cacciari, G. P. Salam and G. Soyez., FastJet User Manual. *Eur. Phys. J.*, C72:1896, (2012), doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[138] D. Rathjens., Jet energy calibration and a search for supersymmetry with vector boson fusion channel like sign di-T_h final states. PhD thesis (2015).

[139] H. Kirschenmann., Jet energy scale corrections and their impact on measurements of the top-quark mass at CMS. PhD thesis, U. Hamburg, Dept. Phys., (2014). DESY-THESIS-2014-031.

[140] Jet Energy Corrections and Resolution performances plots for Run2–June 2016 (LHCP), twiki page 2016.

[141] N. Saoulidou and M. Zielinski., Jet energy resolution measurement. twiki page, 2011.

[142] K. Goebel., Probing supersymmetry based on precise jet measurements at the CMS experiment. PhD thesis, U. Hamburg, Dept. Phys., (2015). DESY-THESIS-2015-003.

[143] CMS Collaboration, Identification of b-quark jets with the CMS experiment. *JINST* 8 (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[144] The D0 Collaboration, Observation of single top quark production. March (2009), arXiv:0903.0850.

[145] CMS Collaboration, Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, (2013).

[146] M. Seidel., Precise measurement of the top-quark mass at the CMS experiment using the ideogram method. PhD thesis, U. Hamburg, Dept. Phys., (2015).

[147] CMS Collaboration, Missing transverse energy performance of the CMS detector. *JINST* 6, (2011) 09001, doi:10.1088/1748-0221/6/09/P09001, arXiv:1106.5048.

[148] Missing Transverse Energy Analysis. twiki (2015).

[149] CMS Collaboration, MET performance in 8 TeV data. CMS Physics Analysis Summary CMS-PAS-JME-12-002 (2013).

[150] V. Khachatryan et al., Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *JINST*, 10(02):P02006, (2015), doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511v2.

[151] BRIL Collaboration, BRIL Work Suite, (2015).

[152] LHCtopWG Collaboration, NLO single-top channel cross sections : ATLAS-CMS recommended predictions for single-top cross sections using the Hathor v2.1 program. twiki, (2015).

[153] GEN POG Collaboration, Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing. twiki, (2015).

[154] GEN POG Collaboration, McM Monte-Carlo Request Management. site, (2015).

[155] LHCtopWG Collaboration, NNLO+NNLL top-quark-pair cross sections : ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program (M. Czakon, A. Mitov, 2013). twiki, (2015).

[156] CMS Collaboration, Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 116 (2016). doi:10.1103/PhysRevLett.116.052002, arXiv:1510.05302.

[157] N. Kidonakis., Top quark production, in proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013), p. 139. (2014). doi:10.3204/DESY-PROC-2013-03/Kidonakis, arXiv:1311.0283.

[158] K. Melnikov and F. Petriello., Electroweak gauge boson production at hadron colliders through $O(\alpha_s^2)$, *Phys. Rev. D*, 74 (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.

[159] J. M. Campbell and R. K. Ellis., MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205 (2010) 10, doi:10.1016/j.nuclphysbps.2010.08.011, arXiv:1007.3492.

[160] J. M. Campbell, R. K. Ellis, and C. Williams., Vector boson pair production at the LHC. *JHEP* 07 (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[161] T. Sjőstrand, S. Mrenna, and P. Skands., PYTHIA 6.4 physics and manual. *JHEP* 05 (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[162] T. Sjőstrand et al., An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191 (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[163] F. Demartin et al., The impact of PDF and alphas uncertainties on Higgs Production in gluon fusion at hadron colliders. *Phys. Rev. D* 82 (2010) doi:10.1103/PhysRevD.82.014002, arXiv:1004.0962.

[164] CMS Collaboration, Underlying event tunes and double parton scattering. CMS Physics Analysis Summary CMS–PAS–GEN–14–001, (2016).

[165] P. Skands, S. Carrazza, and J. Rojo., Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C 74 (2014), no. 8, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[166] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk., Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP* 03 (2013) 015, doi:10.1007/JHEP03(2013)015, arXiv:1212.3460.

[167] M. Bahr et al., HERWIG++ physics and manual. *Eur. Phys. J.*, C 58 (2008) 639, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[168] M. Czakon and A. Mitov., TOP++: a program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185 (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[169] J. Allison et al., Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53 (2006) 270, doi:10.1109/TNS.2006.869826.

[170] PVT Collaboration, Utilities for accessing pileup information for data. twiki, (2015).

[171] C. Battilana et al., Reference muon id, isolation and trigger efficiencies for Run-II. twiki, (2015).

[172] I. Kravchenko et al., Electron scale factors for Run 2. twiki, (2015).

[173] D. Trocino et al., Baseline muon selections for Run-II MC (CMSSW_7_2_X and above)–Preliminary. twiki, (2015).

[174] I. Kravchenko et al., Cut Based Electron ID for Run 2. twiki, (2015).

[175] I. Iashvili et al., Recommended jet energy corrections and uncertainties for data and MC. twiki page, (2015).

[176] N. Saoulidou and M. Zielinski., Jet energy resolution measurement. twiki page, (2011).

[177] BTV Collaboration, BTag Calibration Documentation. twiki, (2015).

[178] D. Dobur et al., Measurement of the top pair-production in association with a Z boson in pp collisions at 13 TeV. CMS Analysis Note CMS AN-2016/031, (January, 2016).

[179] M. Ivova., Measurement of ttZ production cross section in the 4 lepton channel at $\sqrt{s} = 13$ TeV.

[180] BTV Collaboration, Usage of b-tag objects for 13 TeV data with 50ns bunch spacing and 74X PromptReco. twiki, (2015).

[181] J. Ocariz., Probability and statistics for particle physicists. (2014), arXiv:1405.3402.

[182] E. Basso., The Drell-Yan process as a testing ground for parton distributions up to LHC. (2015), arXiv:1509.07988.

[183] CMS Collaboration, Determination of the top quark mass from the m(lb) distribution in dileptonic $t\bar{t}$ events at $\sqrt{s}$=8 TeV, CMS Physics Analysis Summary CMS-PAS-TOP-14-014, (2014).

[184] CMS Collaboration, Measurement of the pp → ZZ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at $\sqrt{s}$=8 TeV. (2014), doi:10.1016/j.physletb.2014.11.059, arXiv:1406.0113.

[185] M. Voutilainen et al., Jet energy scale uncertainty sources. twiki (2012).

[186] , CMS Collaboration, Identification of c-quark jets at the CMS experiment. (2016), CMS Physics Analysis Summary CMS PAS BTV-16-001.

[187] CMS Collaboration, CMS Luminosity measurement for the 2015 data taking period. (2017), CMS Physics Analysis Summary CMS-PAS-LUM-15-001.

[188] M. Gorner et al., p_T (top-quark) based reweighting of ttbar MC. twiki (2013).

[189] A. S. Belyaev, E. E. Boos, and L. V. Dudko., Single top quark at future hadron colliders:Complete signal and background study. *Phys. Rev. D* 59 (1999) (075001), doi:10.1103/PhysRevD.59.075001.

[190] C. D. White, S. Frixione, E. Laenen, and F. Maltoni., Isolating Wt production at the LHC. *JHEP* 11 (2009) 074, doi:10.1088/1126–6708/2009/11/074, arXiv:0908.0631.

[191] S. Catani, G. Ferrera and M. Grazzini., W boson production at hadron colliders:the lepton charge asymmetry in NNLO QCD. (2010), arXiv:1002.3115.

[192] S. Westhoff., Review of top charge asymmetries for LHC run II. (2017), arXiv:1703.01983.

[193] C. Kom, and W. J. Stirling., Charge asymmetry in W+jets production at the LHC. *Eur. Phys. J.*, C69, (2010) 10.1140/epjc/s10052-010-1353-z, arXiv:1004.3404.

[194] G. Cowan, K.Cranmer, E.Gross, and O.Vitells., Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71 (2011) 1554, doi:10.1140/epjc/s10052–011–1554–0, arXiv:1007.1727.

[195] L. Moneta, K. Belasco, K. Cranmer, G. Schott, W. Verkerke, M. Wolf, D. Piparo, A. Lazzaro, S. Kreiss., The RooStats project. arXiv:1009.1003.

[196] G. Petrucciani et al., Documentation of the RooStats based statistics tools for Higgs PAG. twiki, (2012).

[197] CMS Collaboration, Reference muon id, isolation and trigger efficiencies for Run-II. twiki page, 2016.

[198] J. Salfeld-Nebgen, New Luminosity Calibrations for 2015 Data.

[199] J. Wenninger and E. Todesco, Large Hadron Colloder momentum calibration and accuracy. Technical Report CERN-ACC-2017-007, CERN, Geneva, (Feb.2017).

[200] Minuit Collaboration, MINUIT. web-page, (2006).

[201] P. Silva and Q. Hassan, Datacards for TOP-16-006. gitlab repository, (2017).

[202] Particle Data Group, J. Beringer et al., Review of Particle Physics., *Phys. Rev. D* 86 (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[203] D0 Collaboration, Measurement of the inclusive $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV and determination of the top quark pole mass. (2016), doi:10.1103/PhysRevD.94.092004, arXiv:1605.06168.

[204] S. Dulat et al., New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D* 93 (2016) 033006, doi:10.1103/PhysRevD.93.033006, arXiv:1506.07443.

[205] J. Wenninger, Energy Calibration of the LHC Beams at 4 TeV. Technical Report CERN–ATS–2013–040, CERN, (2013).

[206] CMS Collaboration, Determination of the top-quark pole mass and strong coupling constant from the $t\bar{t}$ production cross section in the pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B* 728 (2014) 496, doi:10.1016/j.physletb.2013.12.009, arXiv:1307.1907.

[207] NNPDF Collaboration, Parton distributions for the LHC Run II. *JHEP* 04 (2015) 040, doi:10.1007/JHEP04 (2015)040, arXiv:1410.8849.

[208] CMS Collaboration, Measurement of the $t\bar{t}$ production cross section using events in the $e\mu$ final state in the pp collisions at $\sqrt{s}$ =13 TeV. (2016). arXiv:1611.04040. Submitted to *Eur. Phys. J. C.*

[209] CMS Collaboration, Measurement of the $t\bar{t}$ production cross section using events with one lepton and at least one jet in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP* 09 2017 051, doi:10.1007/JHEP09(2017)051, arXiv:1701.06228.

[210] Top Quark Physics Summary Figures. twiki page

Appendix A

Fit results for trigger efficiency

Appendix-A shows the fit results for trigger efficiency for different jet multiplicies in data and simulation correspondingly. The fits in the inclusive and pass regions describe fairly well the data, however in the fail region the radiative tail in the $Z \to ee$ peak is not being fully captured. This reflects in the results being obtained as a cross-check of the efficiencies which are shown 6.2.2.

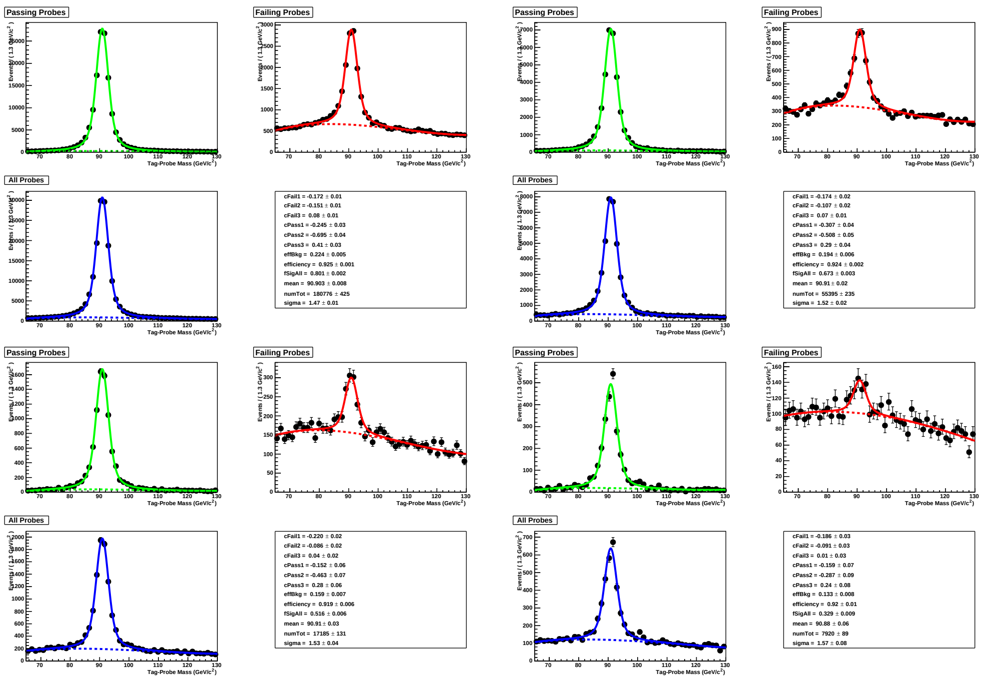


Figure A.1: Tag and probe fits in data $Z \to \mu\mu$ plus (a) =1 jet, (b) =2 jets, (c) =3 jets and (d) $\geq$ 4 jets events.

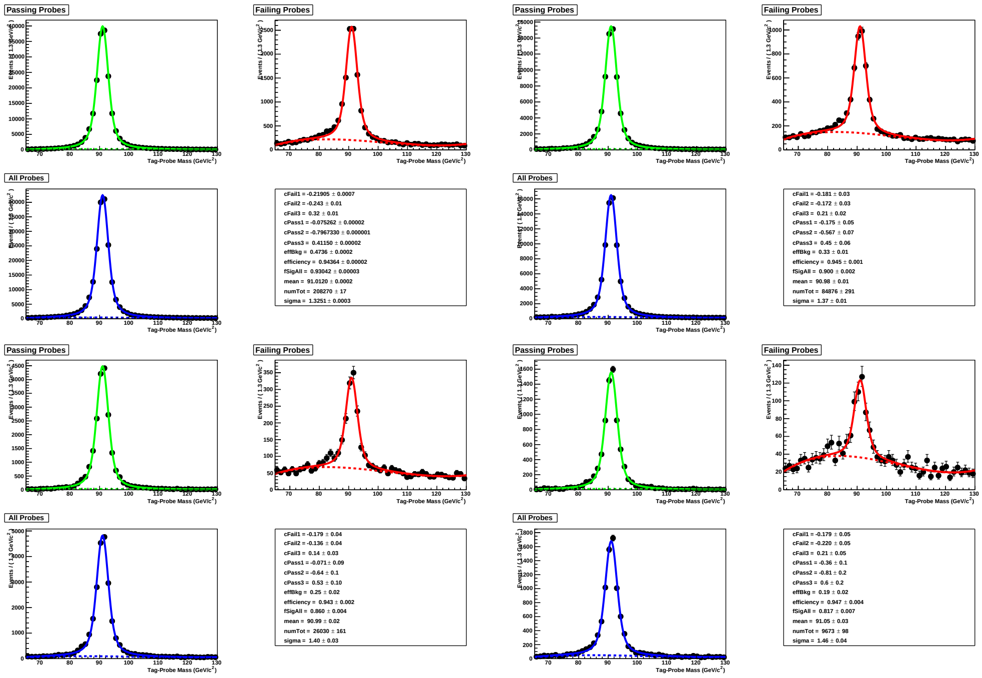


Figure A.2: Tag and probe fits in simulated $Z \to \mu\mu$ plus (a) =1 jet, (b) =2 jets, (c) =3 jets and (d) $\geq$ 4 jets events.

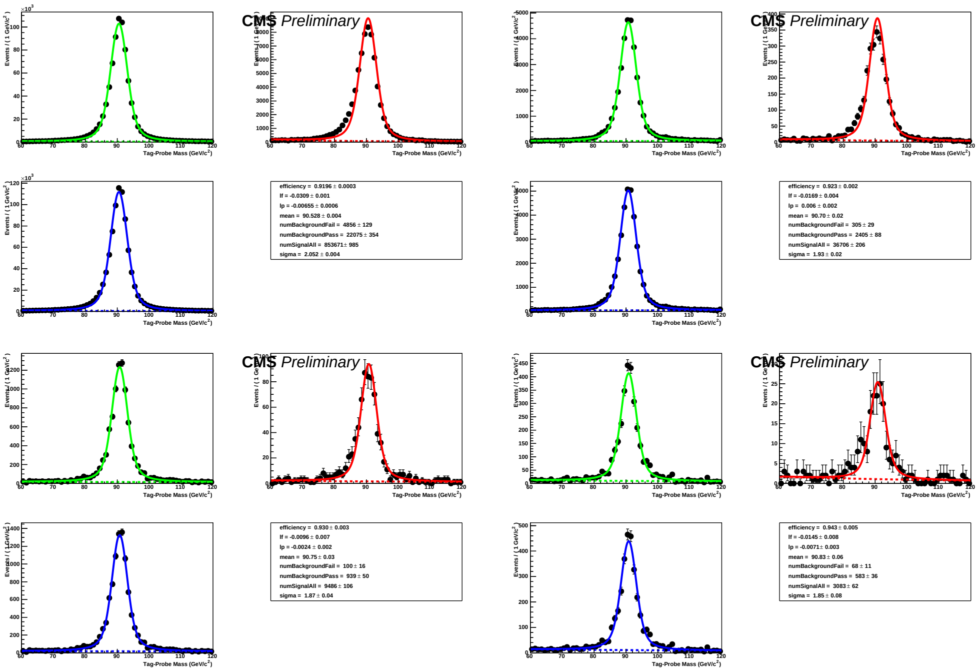


Figure A.3: Tag and probe fits in data $Z \to ee$ plus (a) =1 jet, (b) =2 jets, (c) =3 jets and (d) $\geq$ 4 jets events.

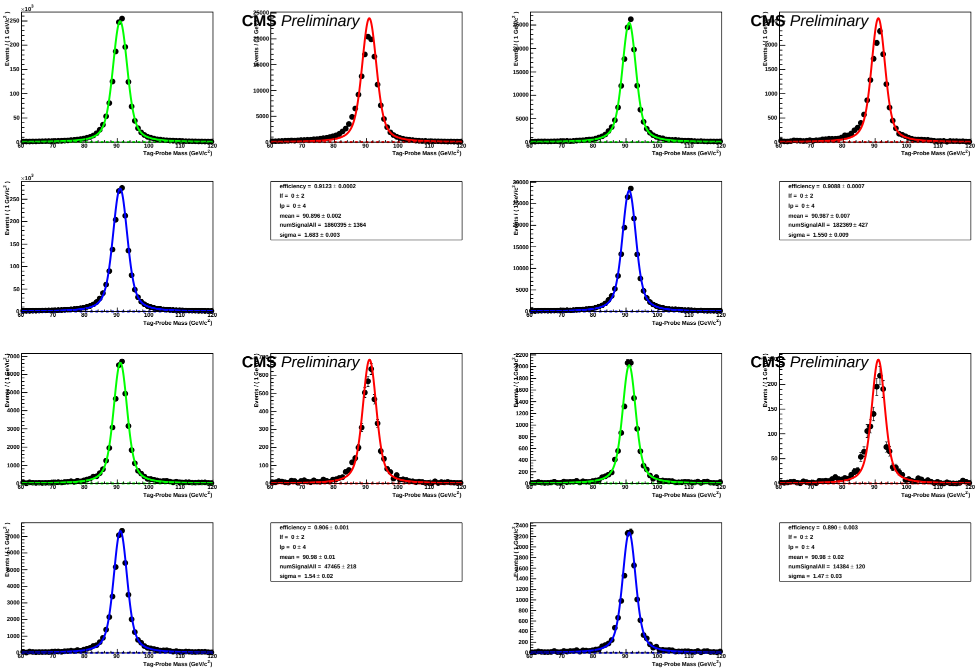


Figure A.4: Tag and probe fits in simulated $Z \to ee$ plus (a) =1 jet, (b) =2 jets, (c) =3 jets and (d) $\geq$ 4 jets events.

Appendix B

H_T distributions in the W+jets simulation

Figure B.1 shows the effect of the main theory uncertainties on the W+jets simulation with aMC@NLO. Different QCD scale choices are expected to induce both shape and rate variations on the H_T distribution which can be at the level of 10%. At very high H_T, and low jet multiplicity, the statistical uncertainty of the sample can be $> 30\%$.

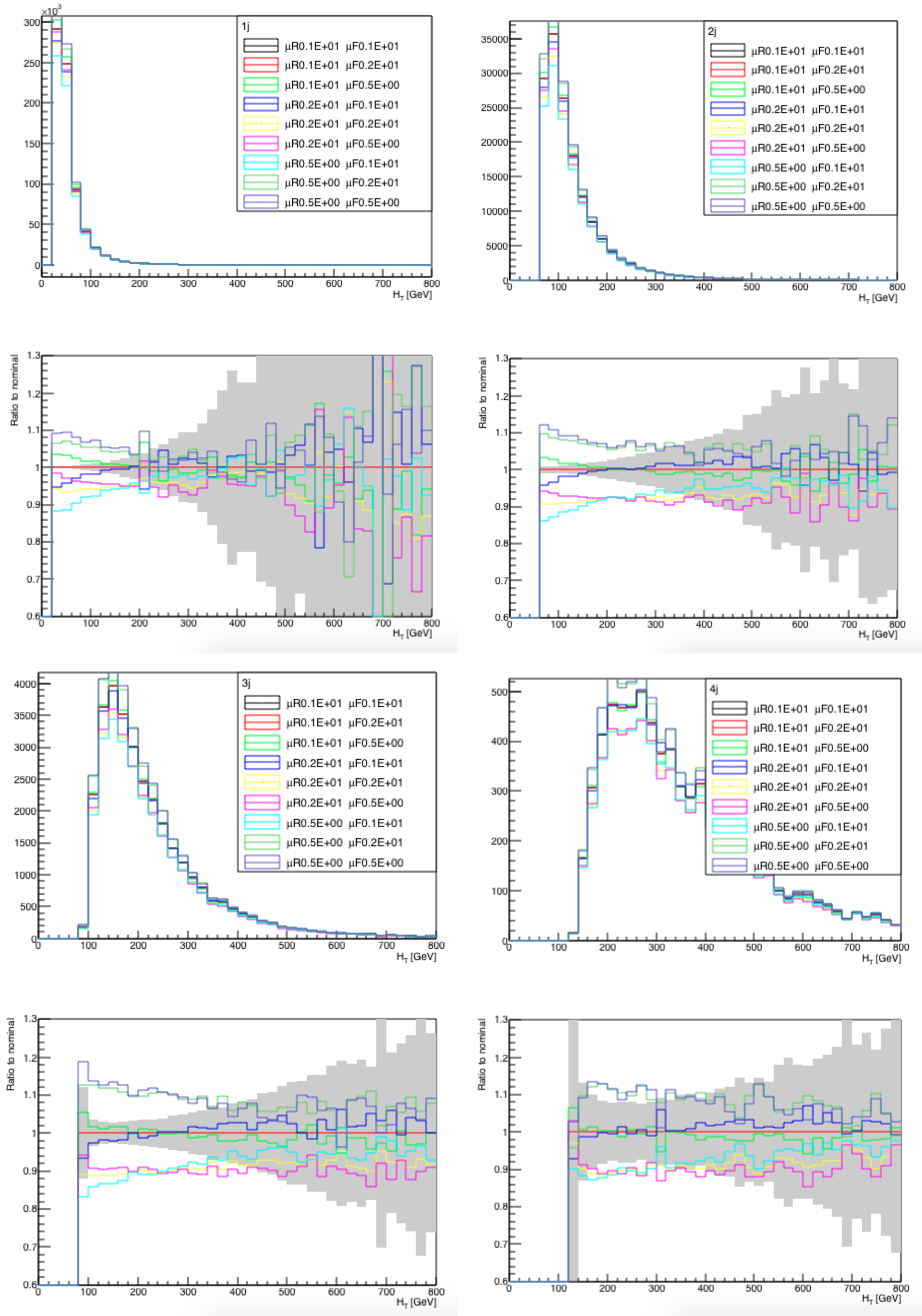


Figure B.1: Distributions for the H_T variable for e^+ events in the W+jets aMC@NLO simulation. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown. The bottom panels show the ratio of the distributions with varied QCD scale at matrix-element level, with respect to the nominal scale choice. The shaded band represents the intrinsic relative statistical uncertainty of the simulated prediction in each bin.

Appendix C

Control distributions

This section describes the pre-fit control distributions for different variables in events with 1, 2, 3 or at least 4 jets. The lepton p_T and η are displayed in Figures C.1, C.2, C.3 and C.4. The shapes of the distributions are overall fairly reproduced as well as the rates prior to any fit. The QCD background model is able to capture the electron fakes generated mostly in the barrel to endcap transition of the ECAL, as well as in the forward ECAL region. The largest disagreement is found in the e+1j region due to electrons in the endcap.

The sum p_T of the selected jets (H_T) and the invariant mass of the $t\bar{t}$ system (M($t\bar{t}$)) are shown in Figures C.5, C.6, C.7 and C.8. A slight trend can be observed towards softer H_T in data, with respect to predictions. The trend is expected, however to be covered by QCD scale choices in the W+jets simulation and at very high H_T by the statistical uncertainty of the MC itself. Figure B.1 in Section B show the breakdown of each one of these effects. Fair agreement is observed for M($t\bar{t}$), in particular in the signal region.

The b-tagging discriminator (CSV) is shown in Figures C.9 and C.10. Fair agreement is observed between data and simulation in all categories.

The $\not\!\!E_T$ distributions are shown in Figures C.11 and C.12. Fair agreement is observed between data and predictions.

The b-jet multiplicity are shown in Figure C.13. A good agreement is found between data and expectations. A slight excess is common in both channels for

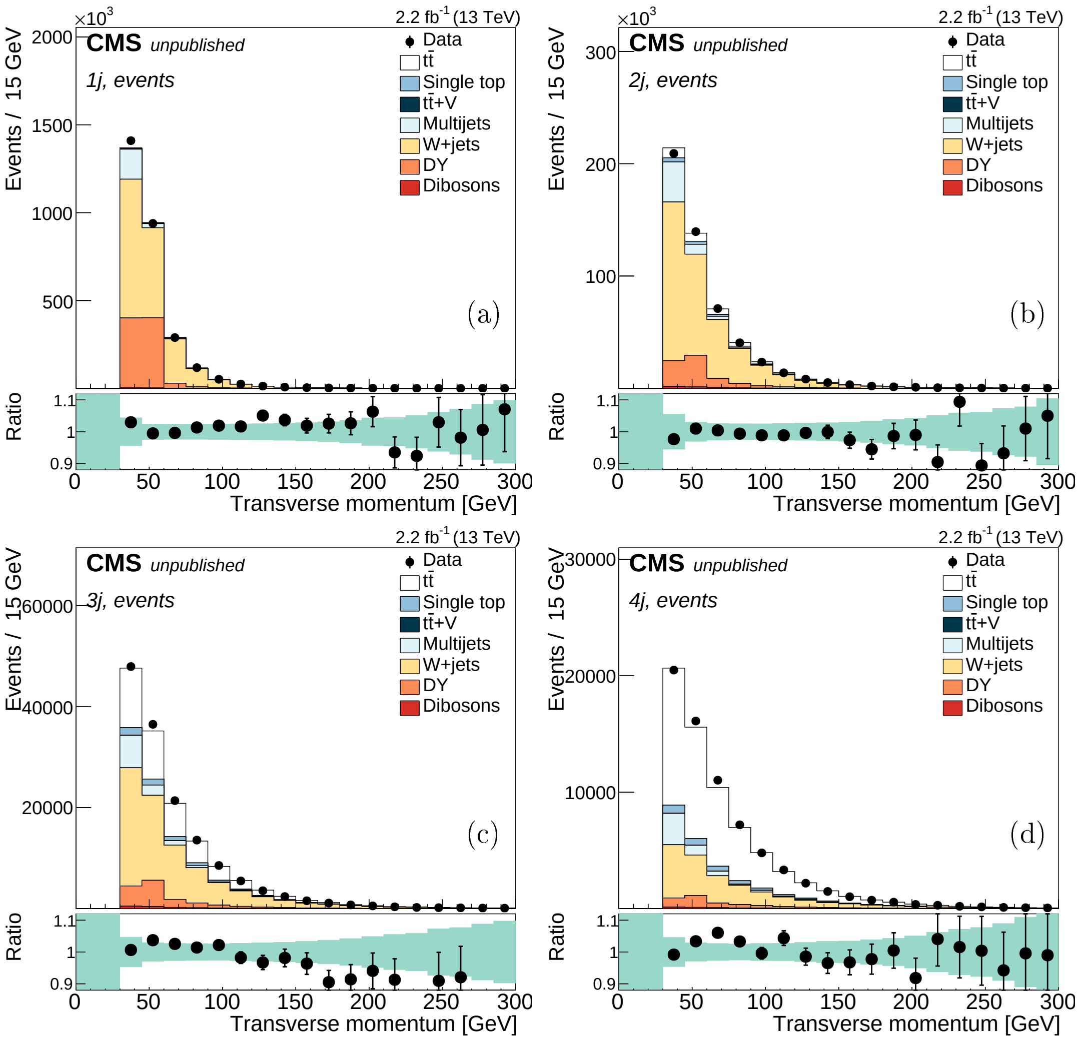


Figure C.1: Distributions for the lepton pt variable for μ events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

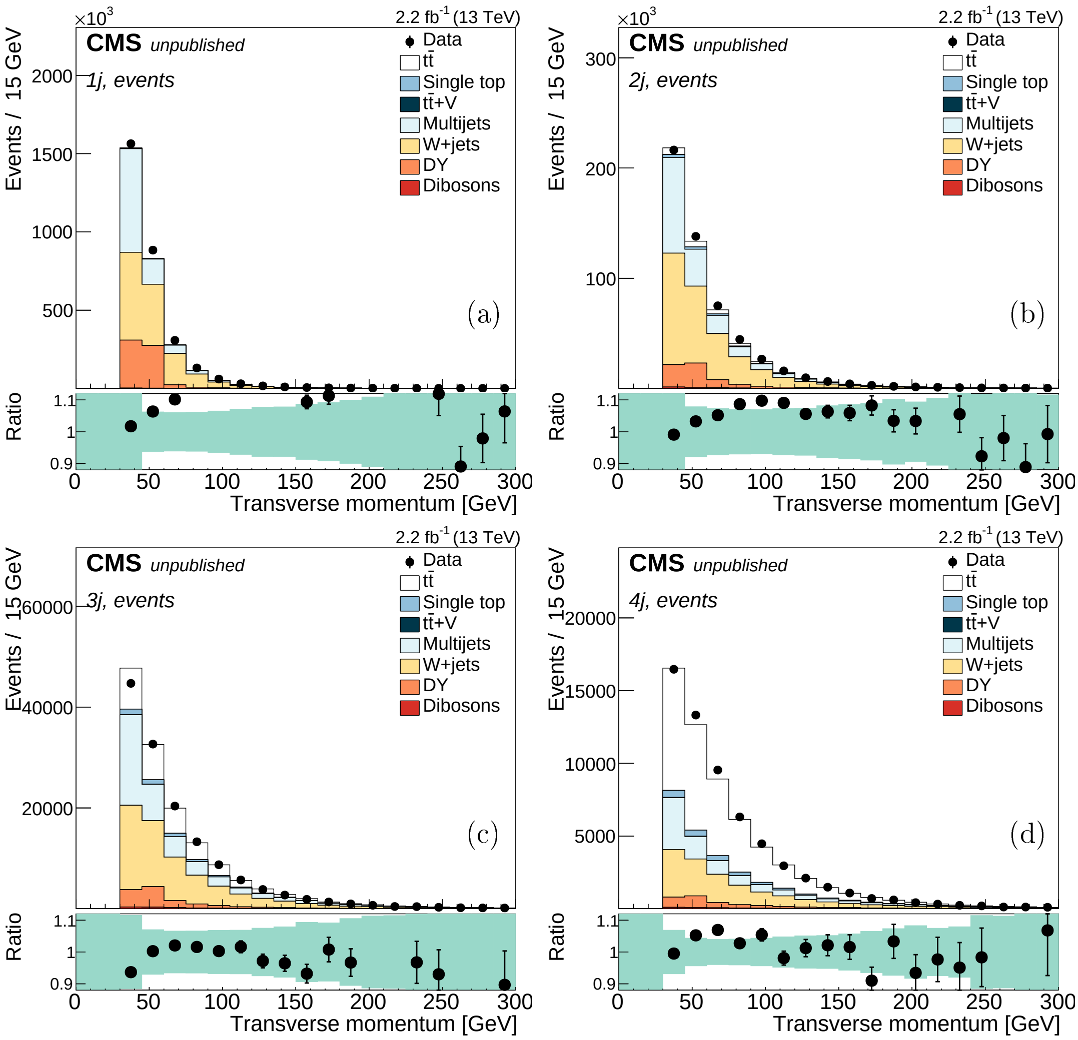


Figure C.2: Distributions for the lepton pt variable for e events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

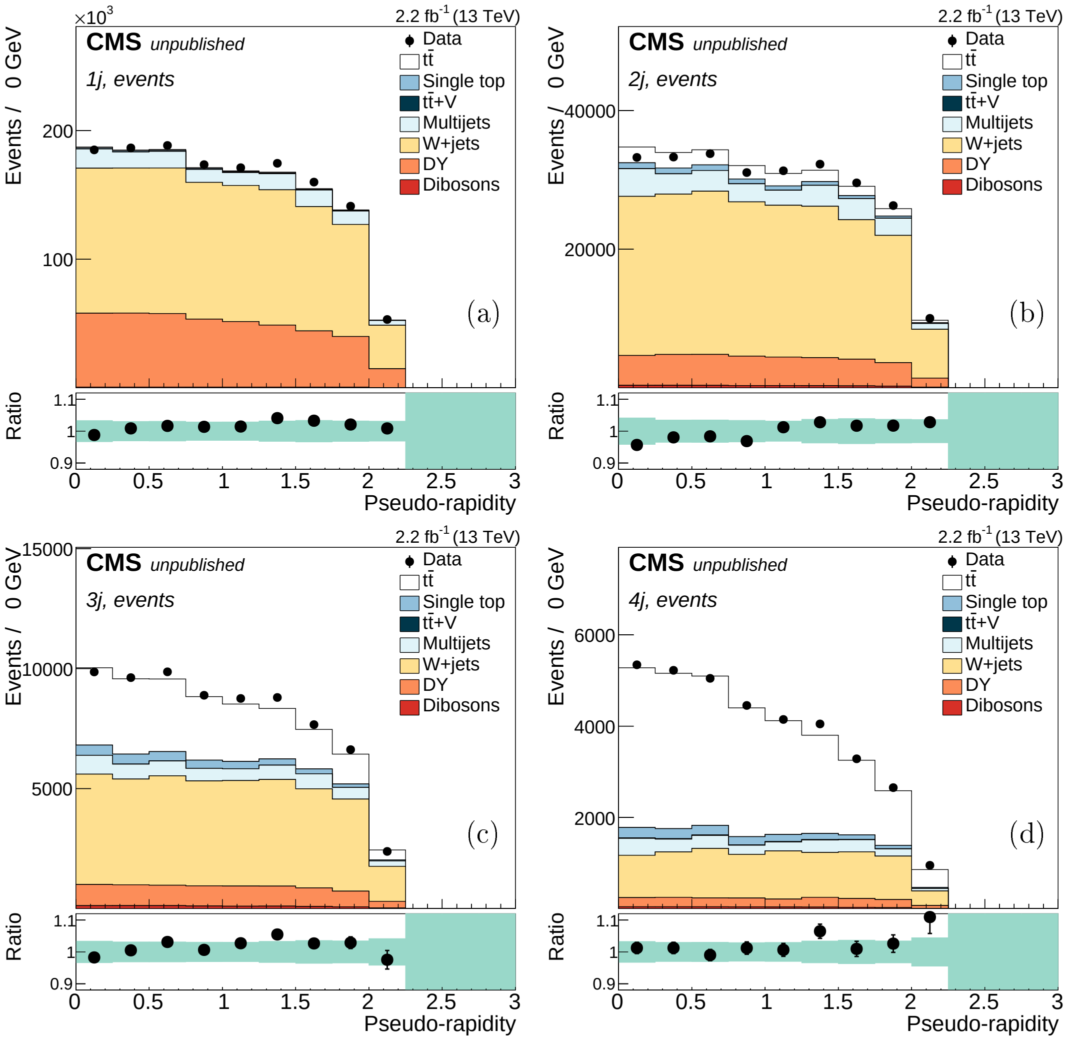


Figure C.3: Distributions for the lepton eta variable for μ events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

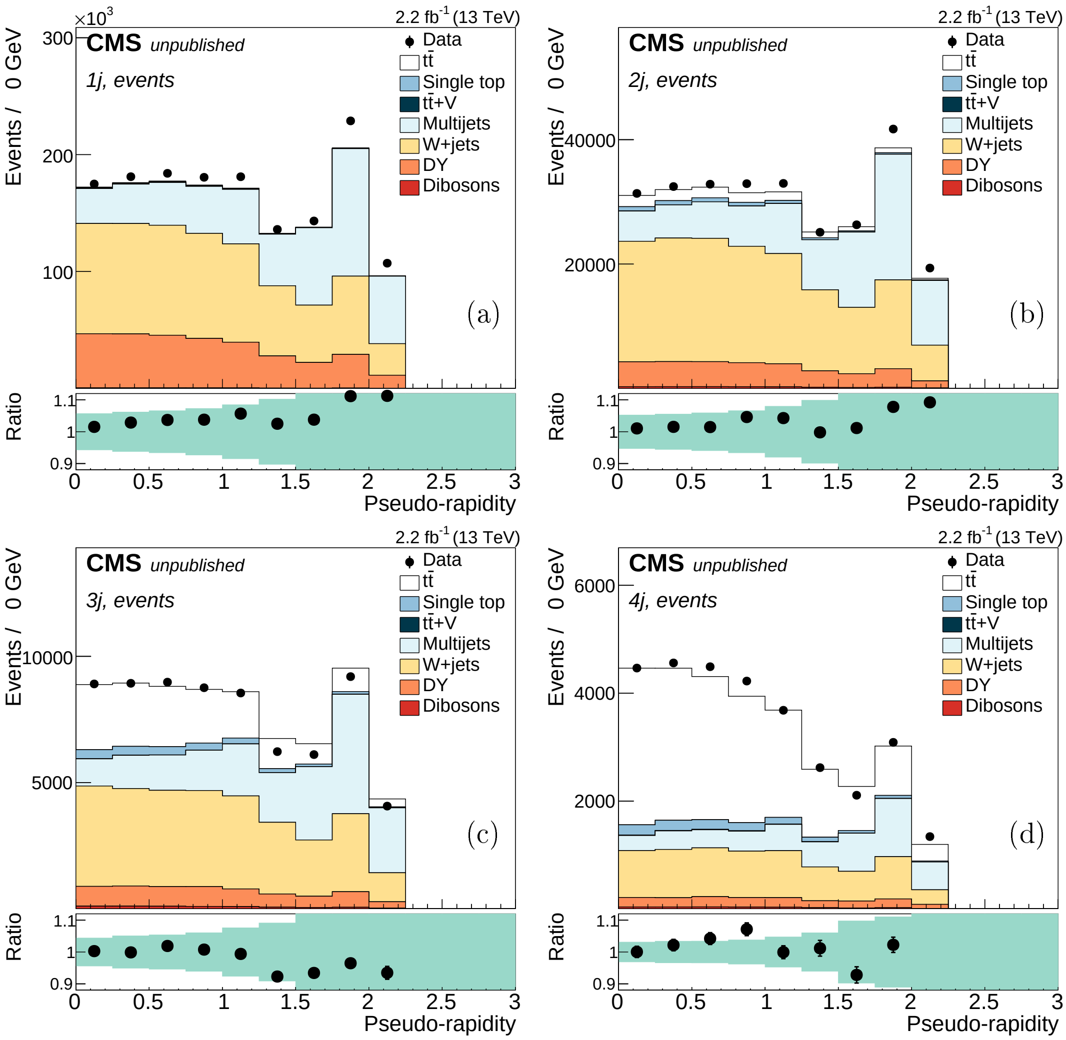


Figure C.4: Distributions for the lepton eta variable for e events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

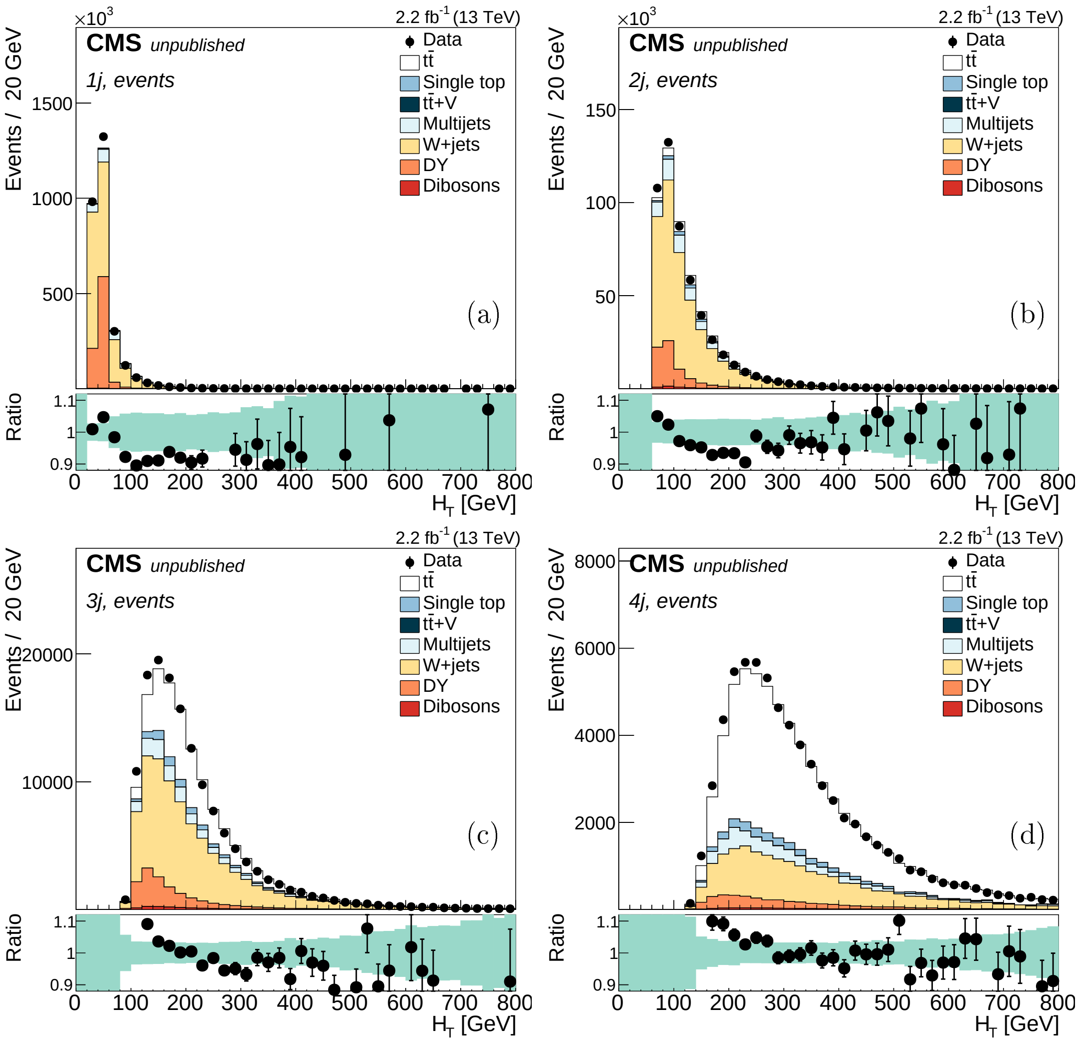


Figure C.5: Distributions for the H_T variable for μ events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

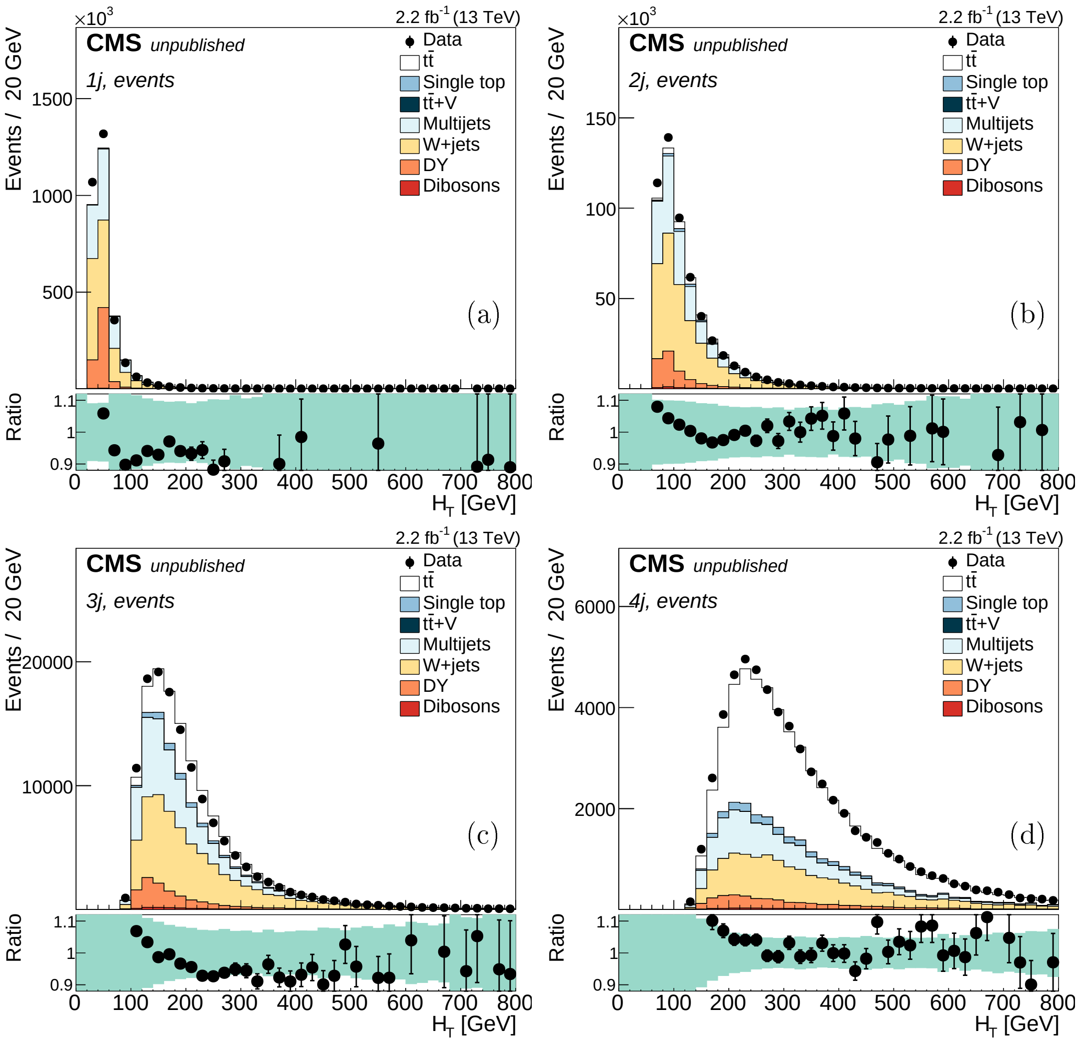


Figure C.6: Distributions for the H_T variable for e events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

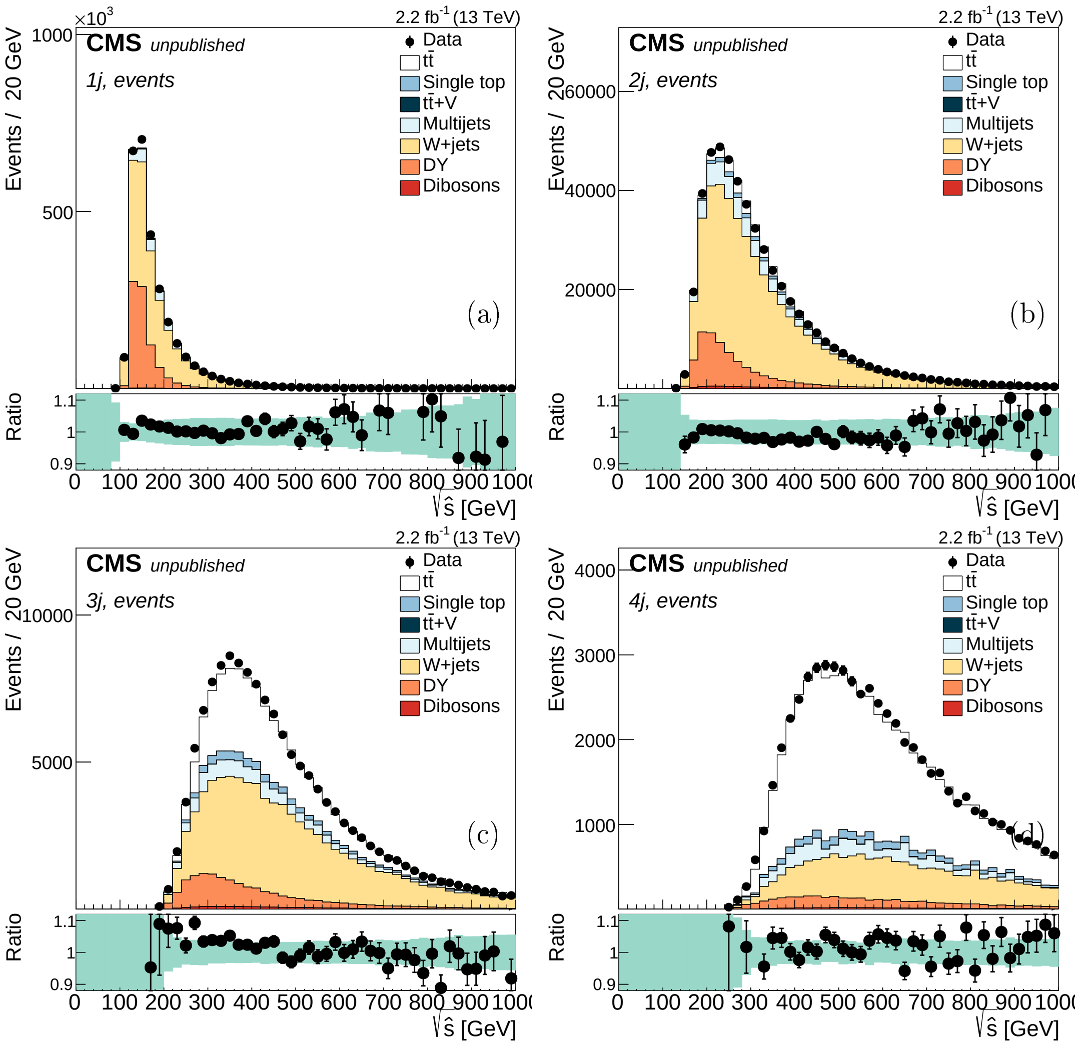


Figure C.7: Distributions for the M($t\bar{t}$) variable for μ events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

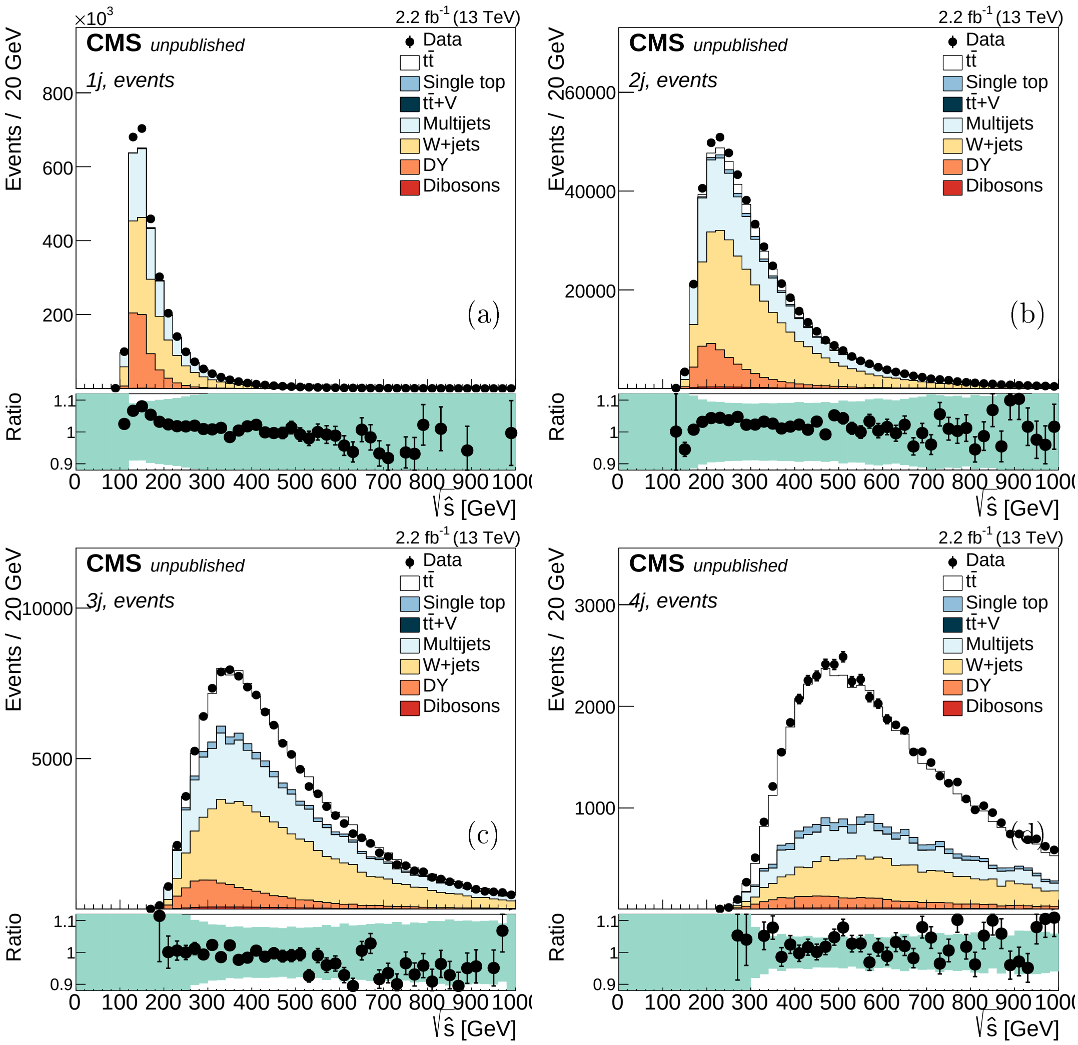


Figure C.8: Distributions for the M($t\bar{t}$) variable for e events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

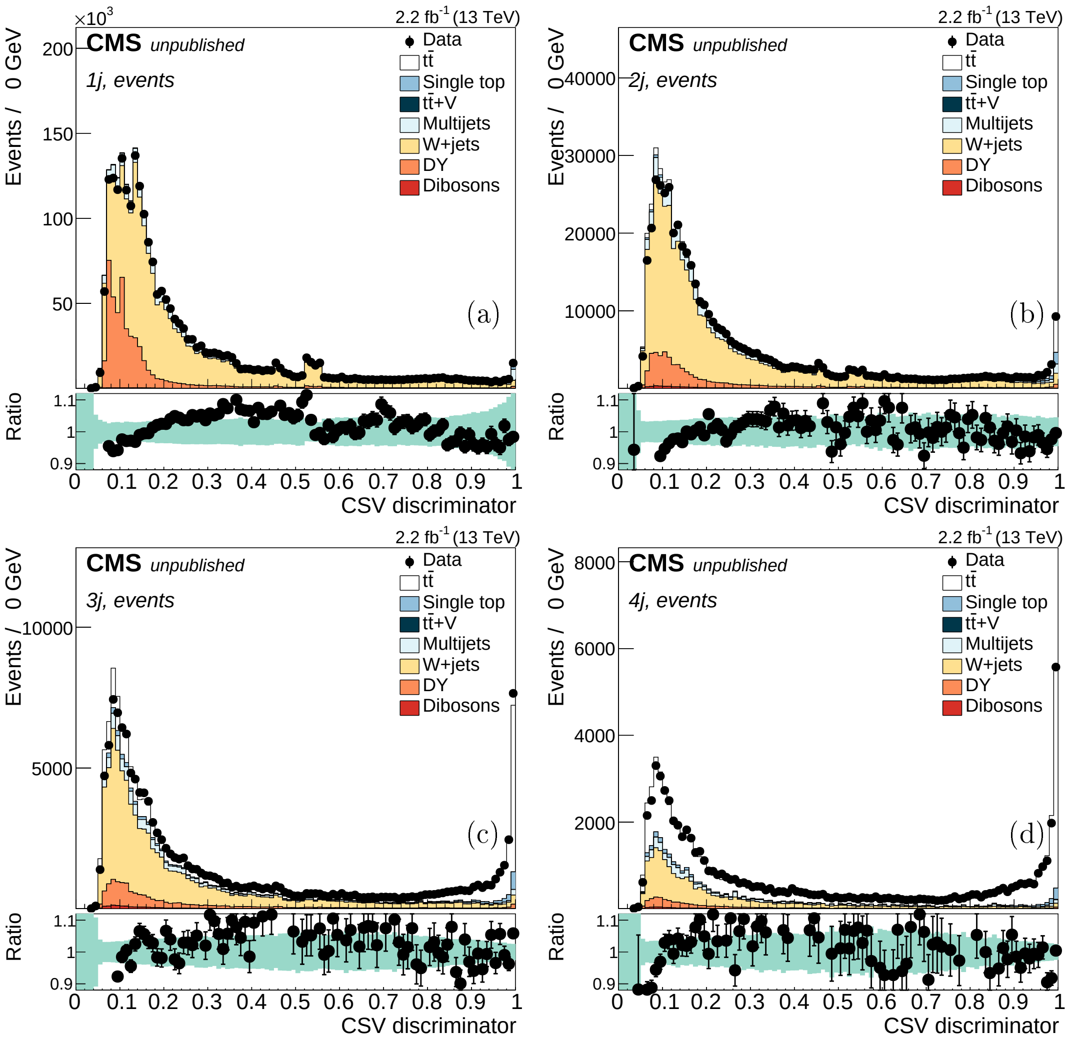


Figure C.9: Distributions for the CSV discriminator for μ events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

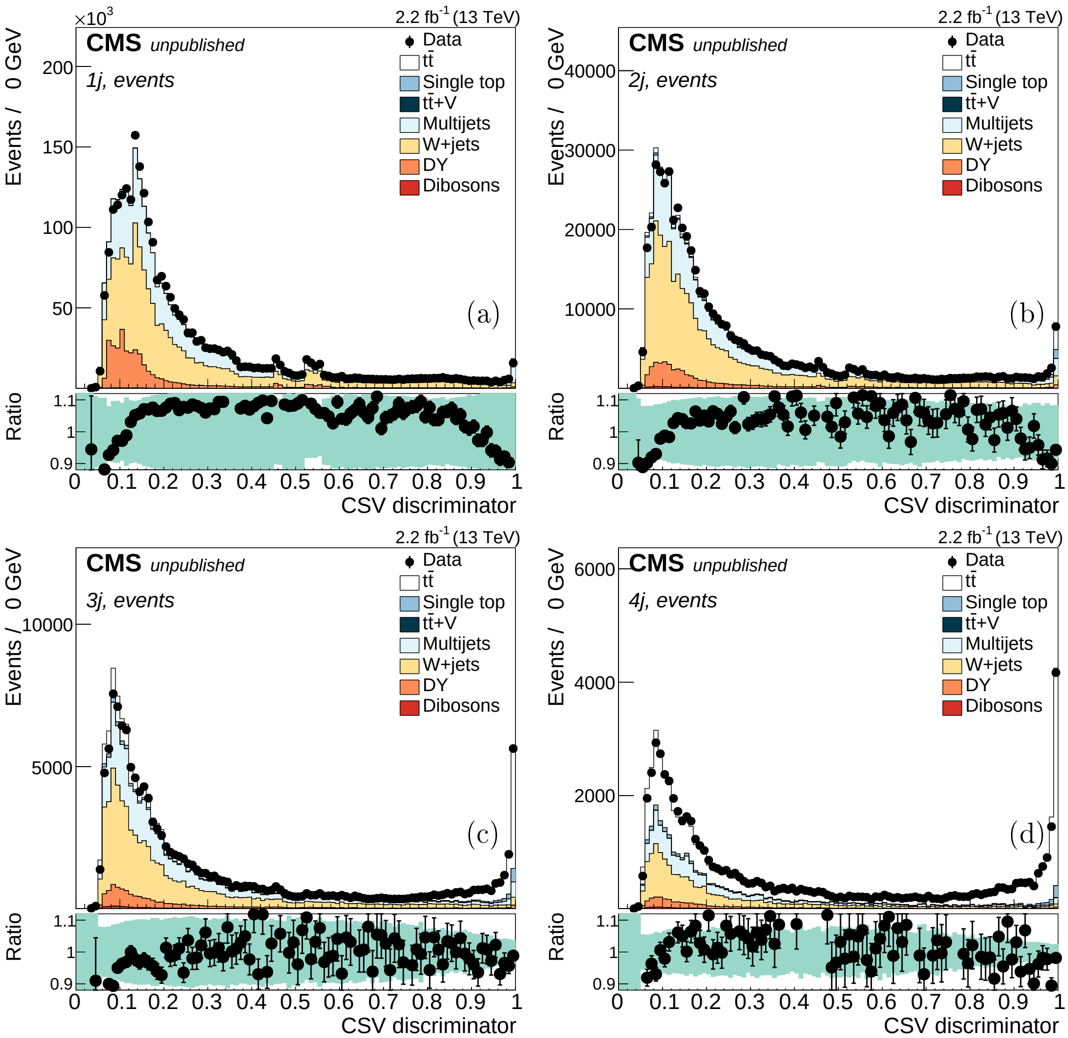


Figure C.10: Distributions for the CSV discriminator for e events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

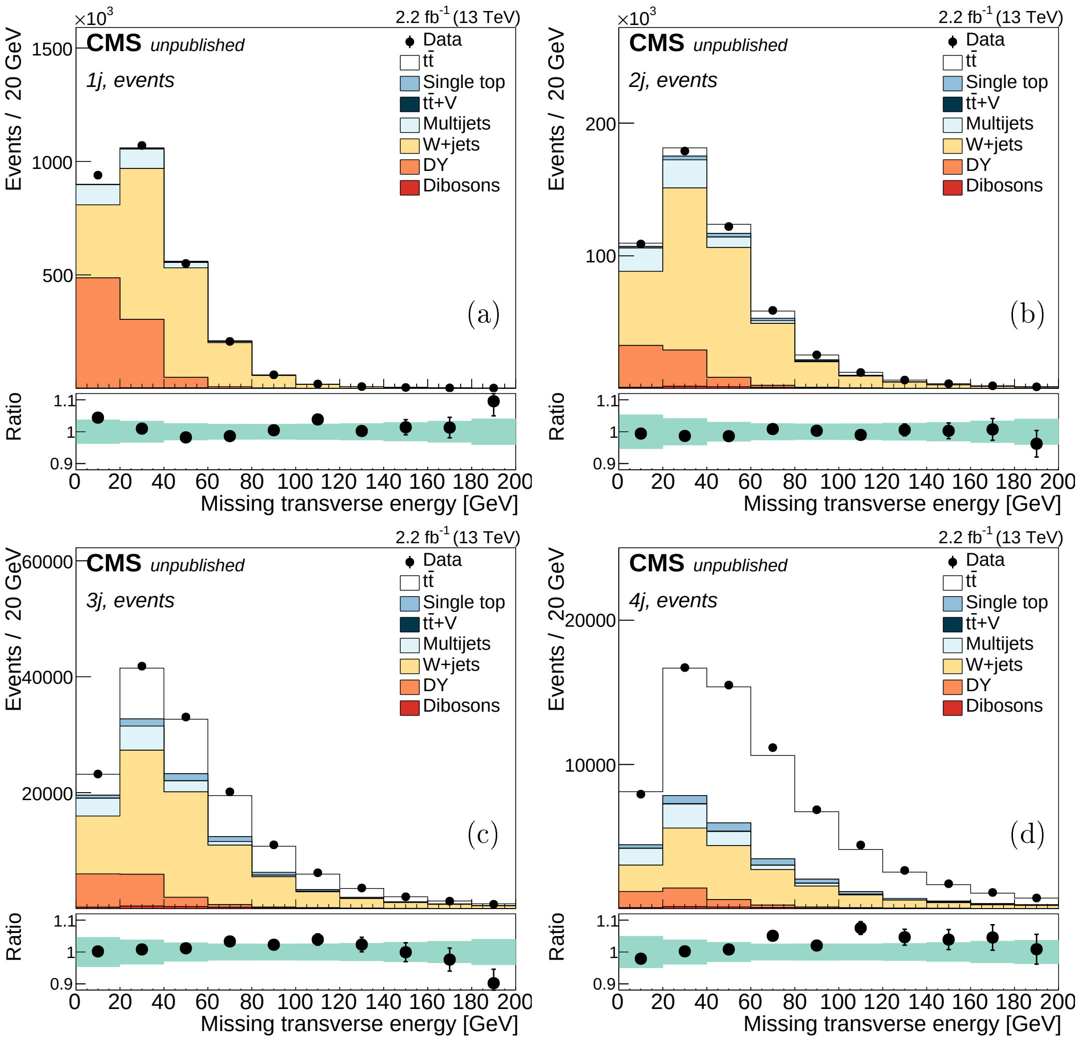


Figure C.11: Distributions for the $\not\!\!E_T$ in μ events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

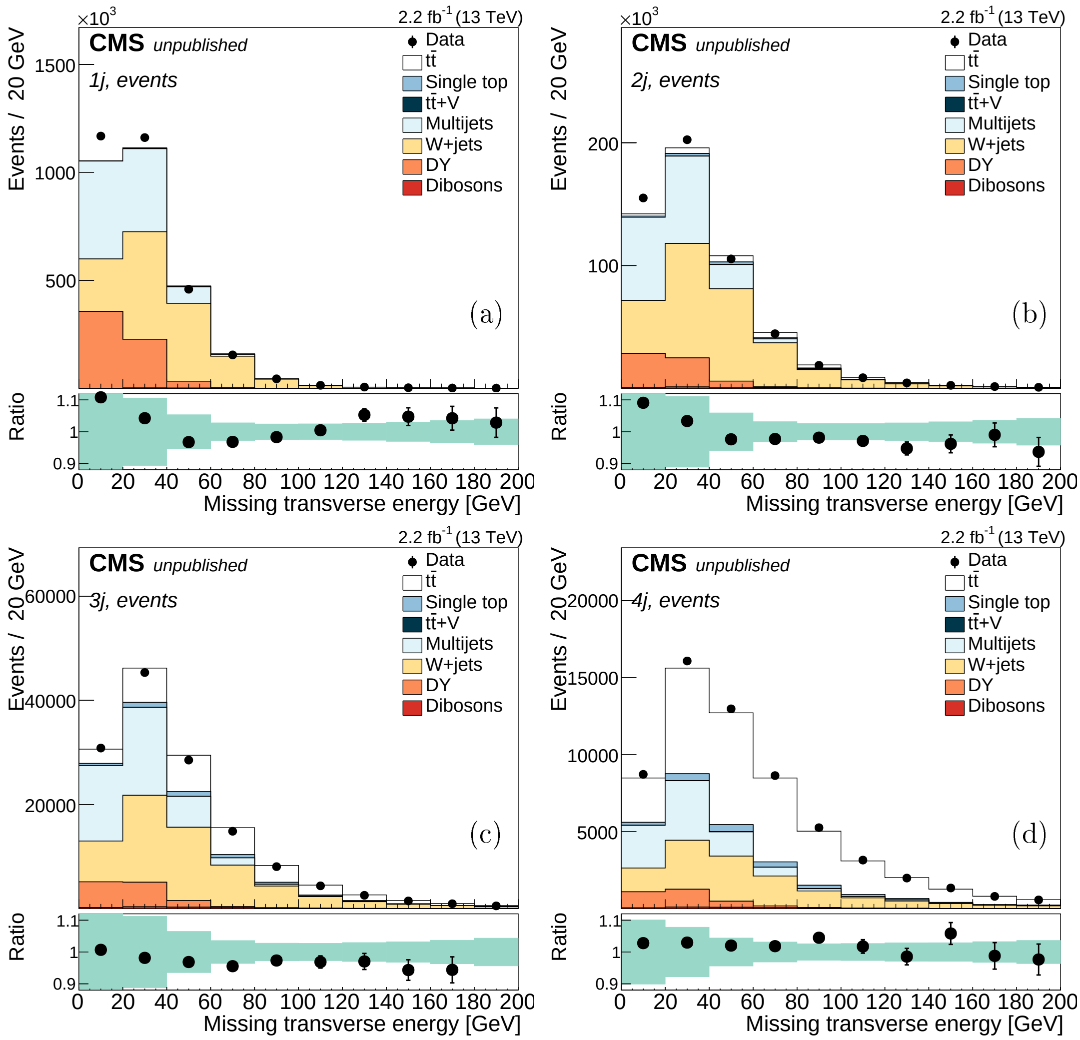


Figure C.12: Distributions for the $\not\!\!E_T$ in e events. Events with 1 (a), 2 (b), 3 (c) or at least 4 (d) jets are shown.

events with 3 or 4 jets and 2 b-tags.

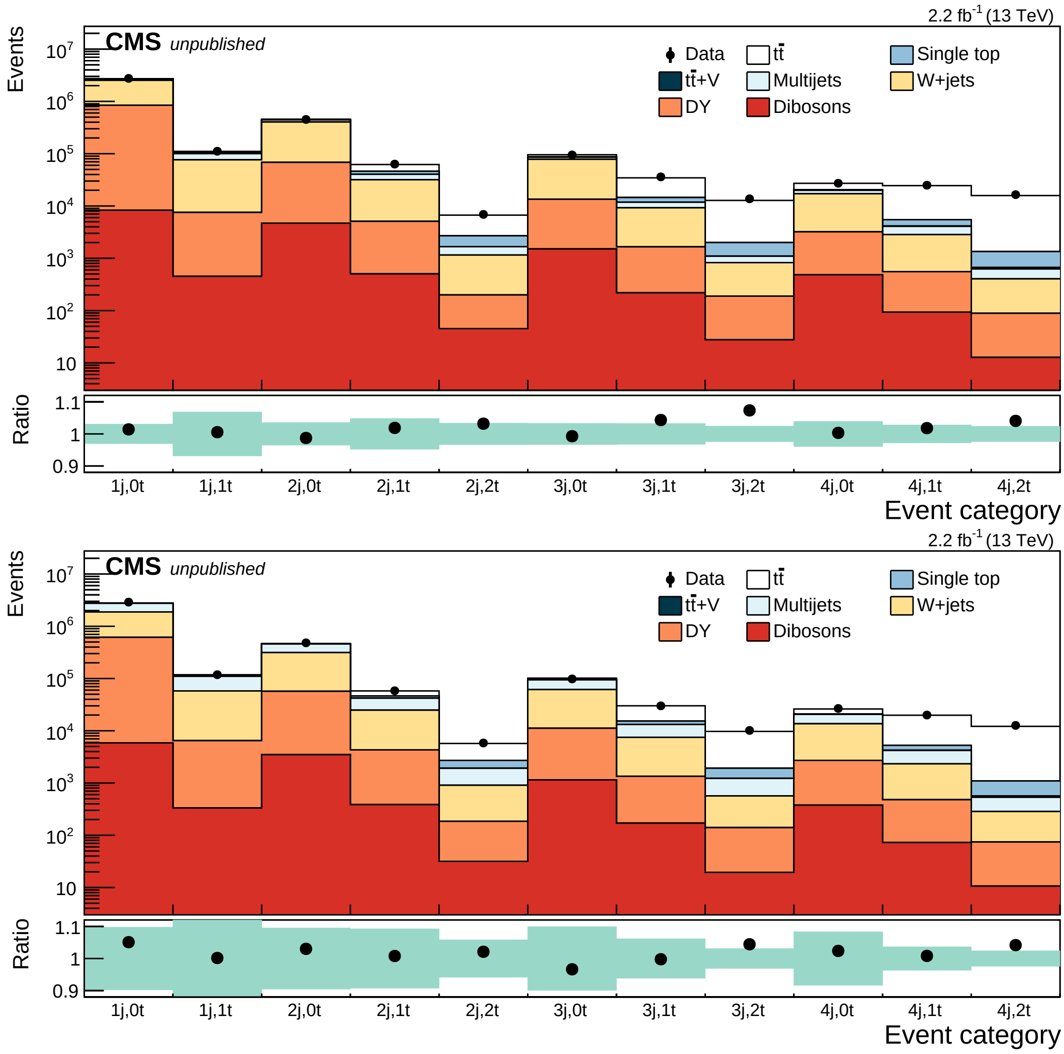


Figure C.13: Distributions for the number of b tags in μ (top) and e (bottom) events.

The prefit distributions for the $M(\ell b)$ and min $M(\ell, b)$ are shown in Figures C.14, C.15, C.16 and C.17 correspondingly. Fair agreement is observed for all categories. In some case the QCD estimate tends to over predict pairs at very low mass, this is however covered by the normalization of the QCD multijets, and it affects only a handful of events.

Figures C.18, C.19, C.20 and C.21 show extra control distributions for the ΔR variable computed between the lepton and the b-jet candidate chosen to compute the $M(\ell, b)$ variables shown before. One can observe that in muon+$1b$-tag events the QCD multijet prediction obtained from the control region tends to have slightly

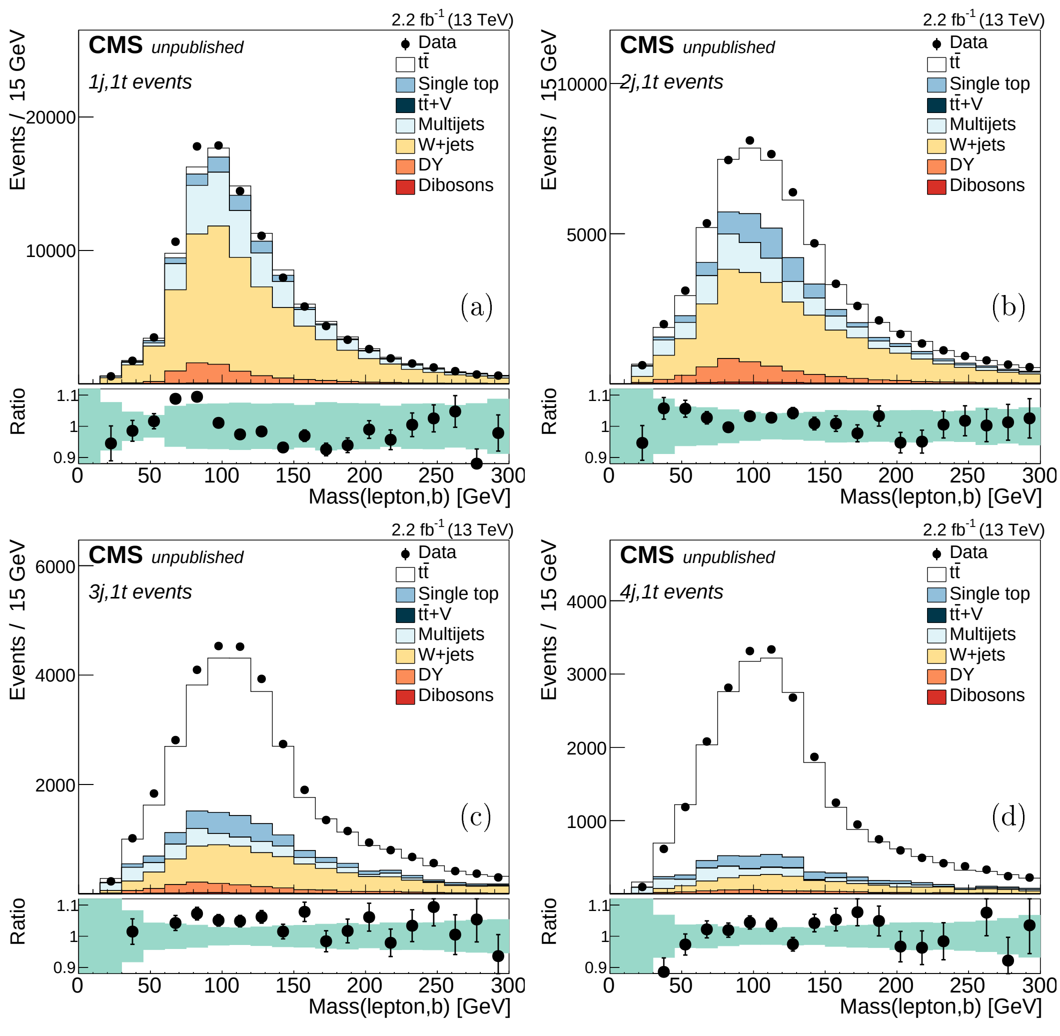


Figure C.14: Distributions for the $M(\ell b)$ variable in μ events. From left to right events with 1,2,3 or at least 4 jets and one b-tag.

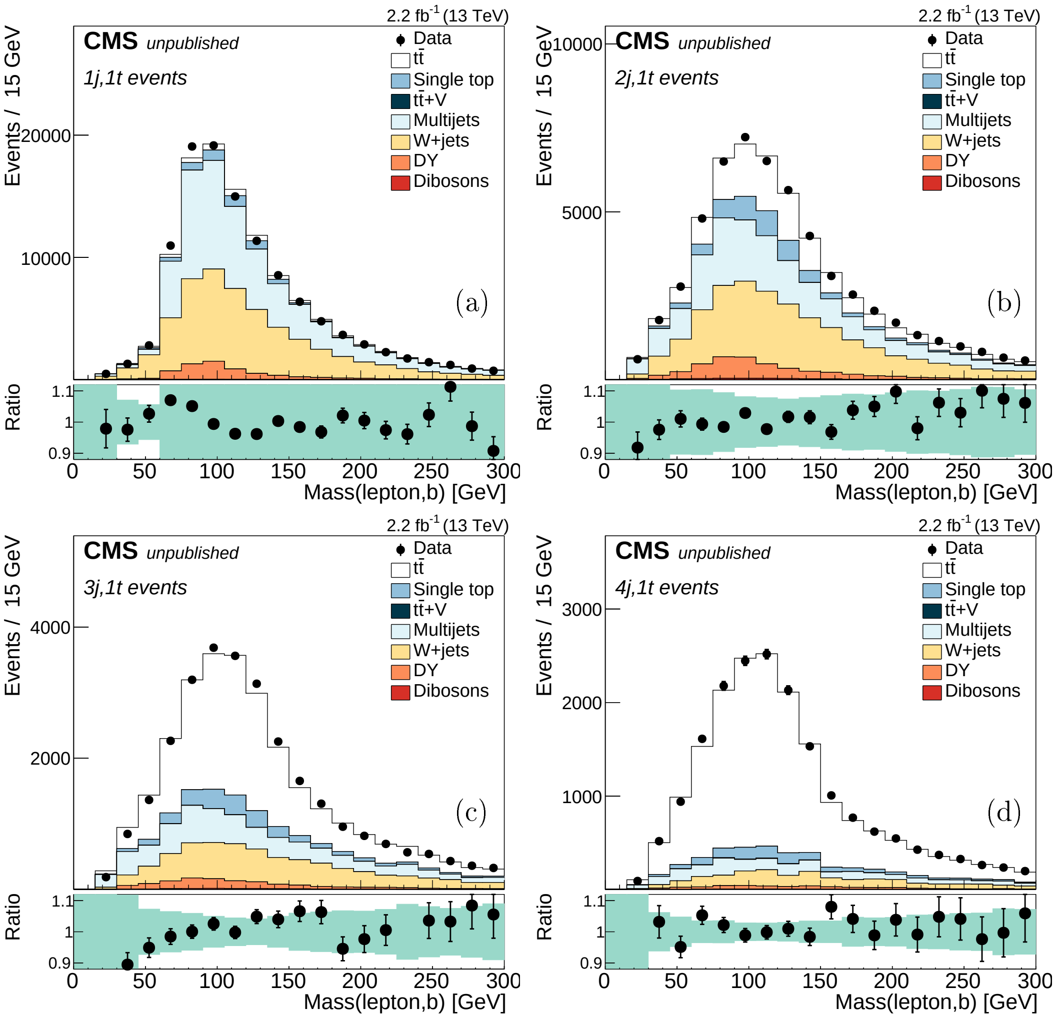


Figure C.15: Distributions for the $M(\ell b)$ variable in e events. From left to right events with 1,2,3 or at least 4 jets and one b-tag.

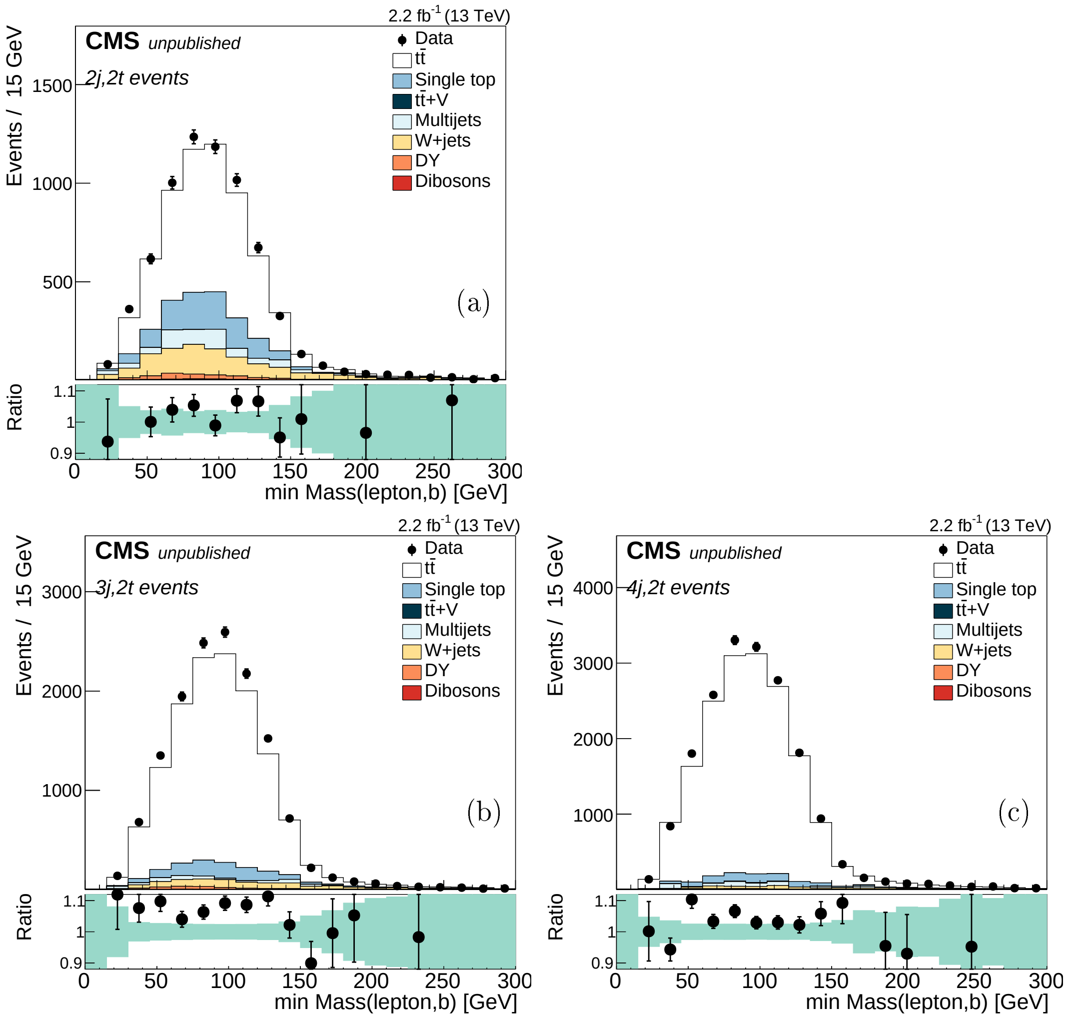


Figure C.16: Distributions for the min $M(\ell b)$ variable in μ events. From left to right events with 2, 3 or at least 4 jets and at least two b-tags.

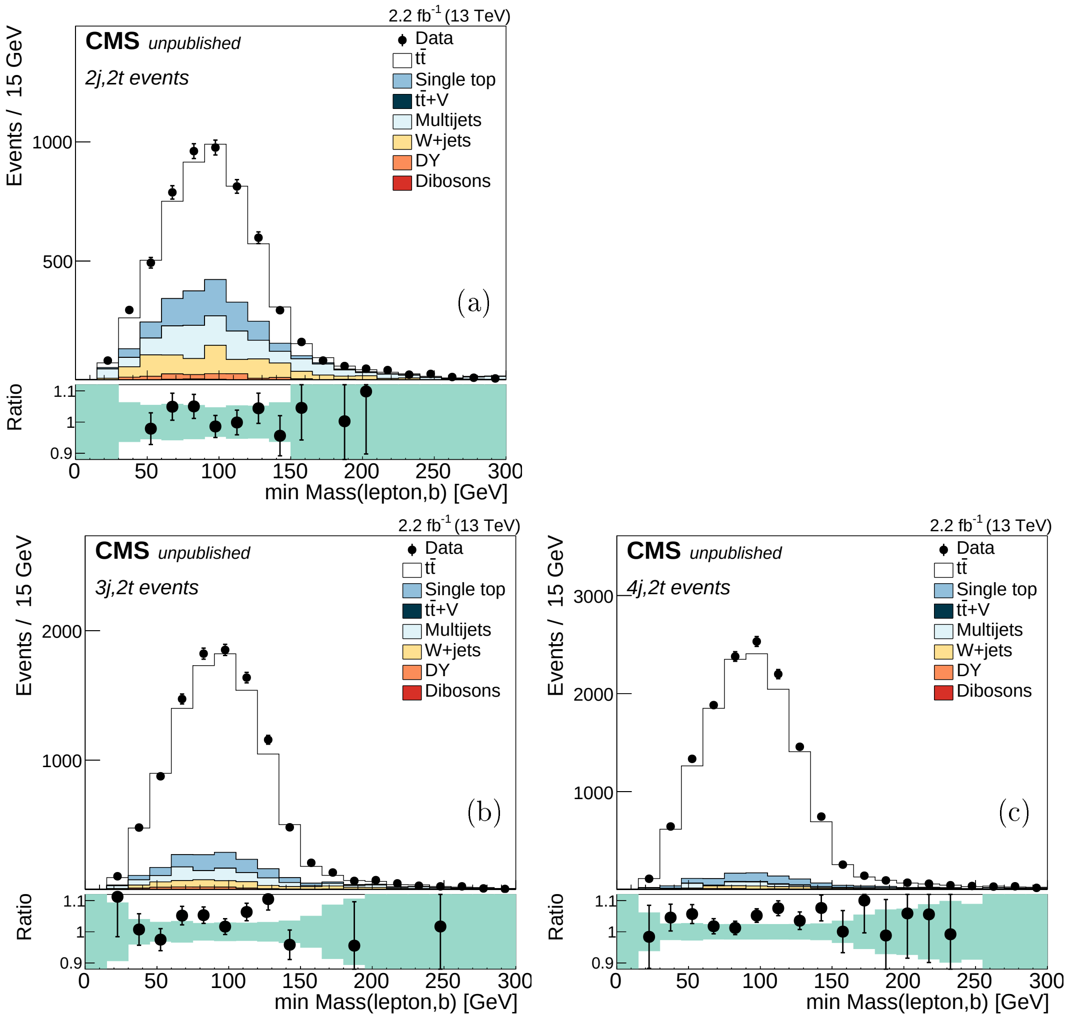


Figure C.17: Distributions for the $\min M(\ell b)$ variable in e events. From left to right events with 2, 3 or at least 4 jets and at least two b-tags.

higher contamination from close by lepton-b-tagged jet pairs. Nevertheless, as this is expected to affect only a small portion of the events, and it is expected to be covered by the normalization and shape uncertainties for the QCD multijets background, we don't apply an extra cut or re-weighting of the QCD template.

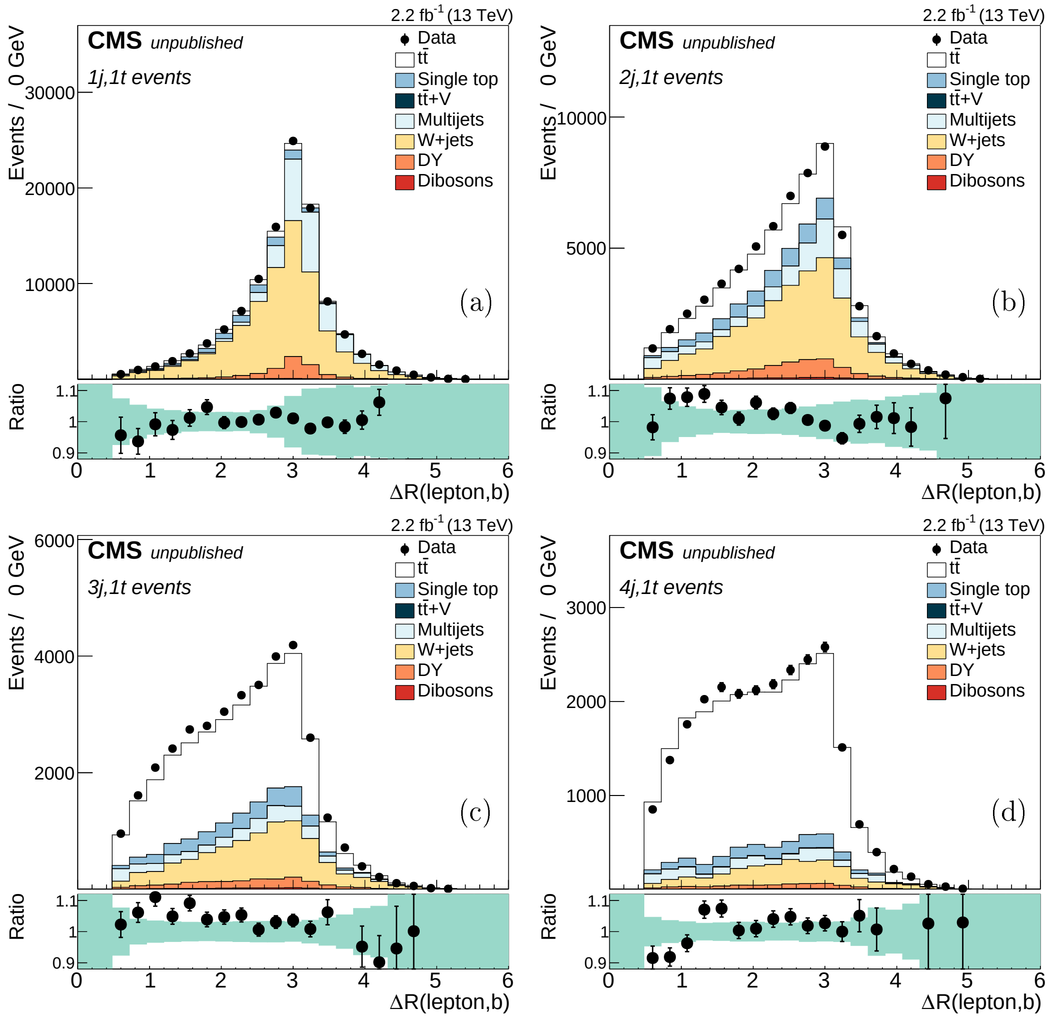


Figure C.18: Distributions for the $\Delta R(\ell b)$ variable in μ events. From left to right events with 1, 2, 3 or at least 4 jets and one b-tag.

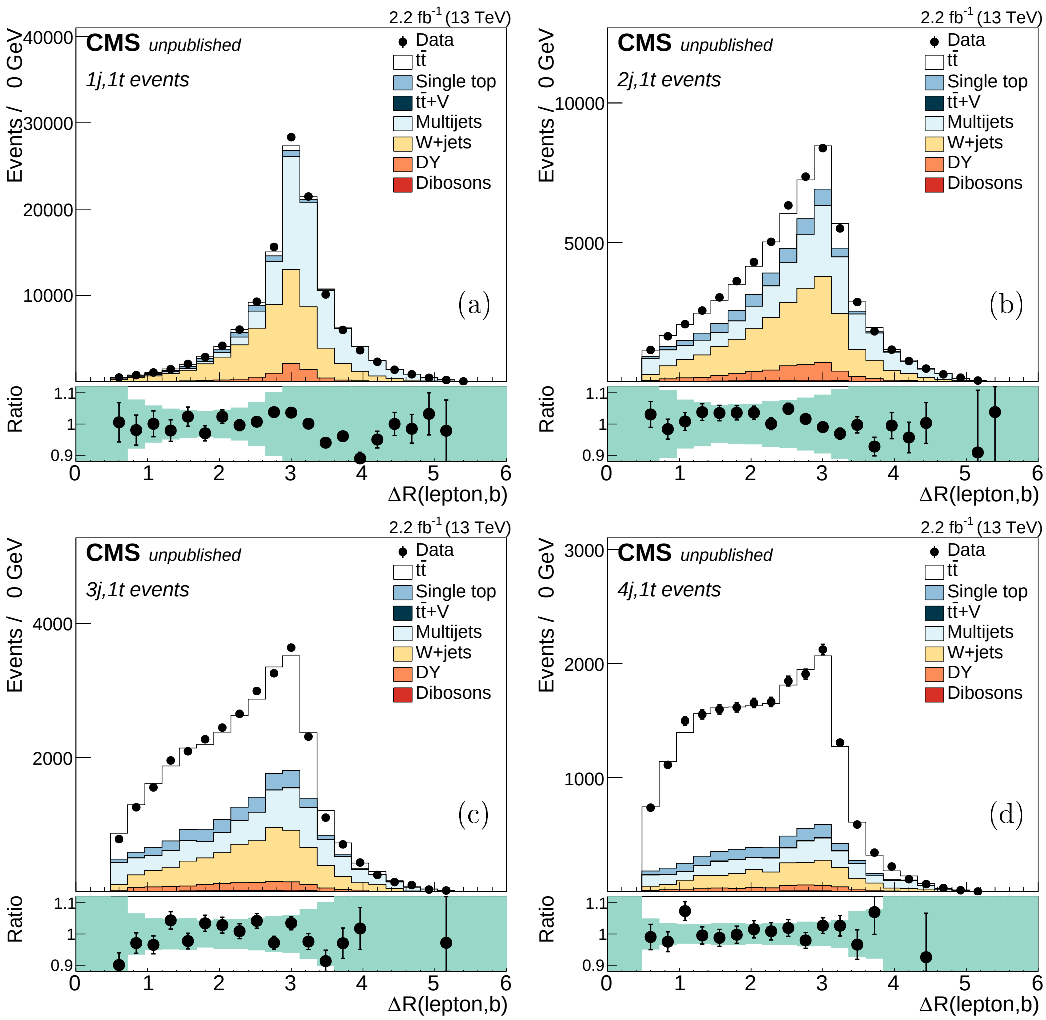


Figure C.19: Distributions for the $\Delta R(\ell b)$ variable in e events. From left to right events with 1,2,3 or at least 4 jets and one b-tag.

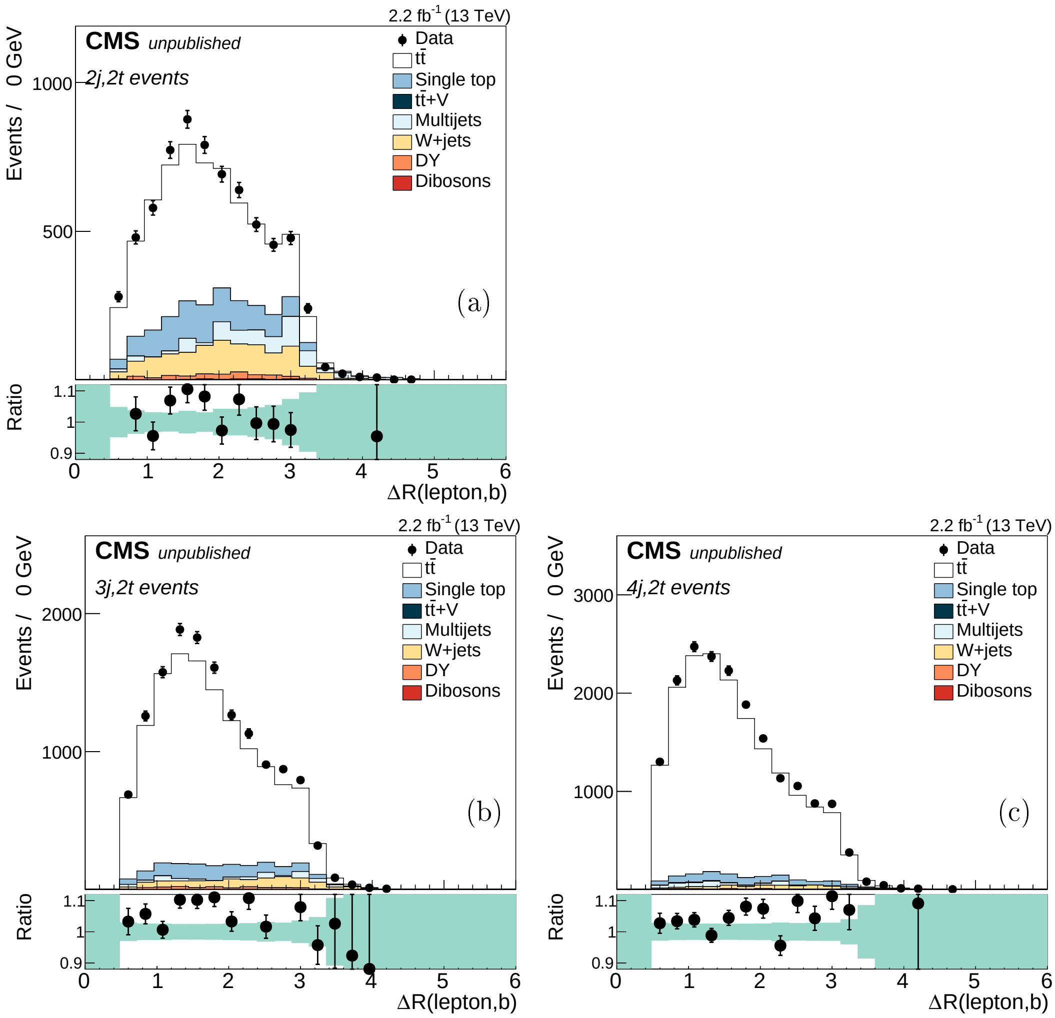


Figure C.20: Distributions for the $\Delta R(\ell b)$ variable in μ events. From left to right events with 2, 3 or at least 4 jets and at least two b-tags.

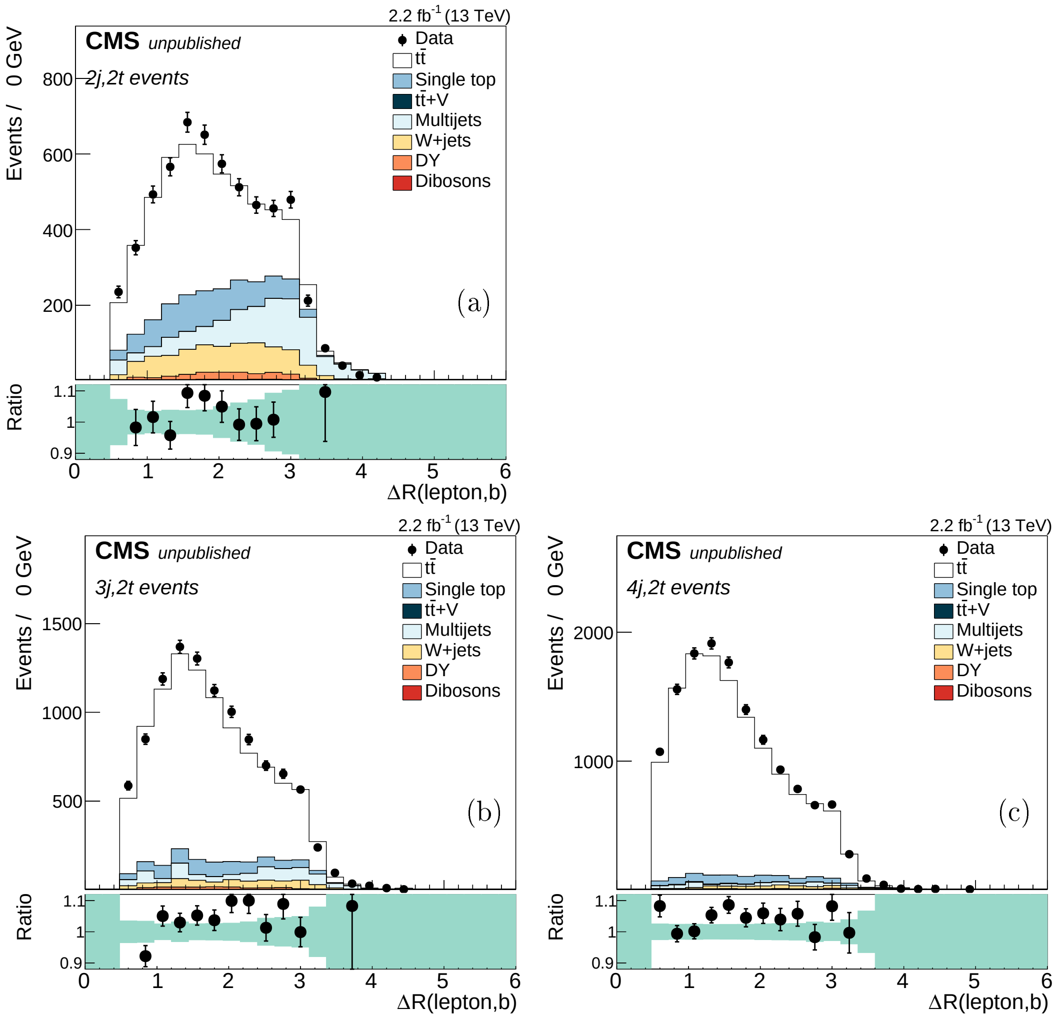


Figure C.21: Distributions for the $\Delta R(\ell b)$ variable in e events. From left to right events with 2, 3 or at least 4 jets and at least two b-tags.