# Constraining the Top-Higgs Yukawa Coupling with $t\bar{t}$ Production at $\sqrt{s} = 13$ TeV with CMS Experiment

by

Yi-Ting Duh

Submitted in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Supervised by Professor Regina Demina

Department of Physics and Astronomy

Arts, Sciences and Engineering

School of Arts and Sciences

University of Rochester

Rochester, New York

2018

# Table of Contents

# Biographical Sketch

The author was born and raised in Taiwan. She received her Bachelor of Science degree in Physics from National Taiwan Normal University and Master of Arts degree in Physics from National Taiwan University. She started her PhD program at the University of Rochester and later joined the group of Professor Regina Demina to pursue research in CMS experiment under her direction. Beginning in 2014, she was stationed at the European Center for Nuclear Research (CENR) in Geneva, Switzerland and conducted her research as a member of the CMS collaboration.

The following publications were a result of work conducted during doctoral study:

- Measurement of differential cross sections for top quark pair production and associated jets using the lepton+jets final state in proton-proton collisions at 13 TeV, TOP-17-002, PRD 97 (2018) 112003
- Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV, TOP-16-008, PRD 95 (2017) 092001

# Acknowledgements

I would like to acknowledge the many kind individuals who have invested significant professional and emotional resources to help me complete this work. I convey my deepest gratitude to my advisor Professor Regina Demina, for training me in research methods and ensuring I had every opportunity to succeed. My thanks are also extended to Professor Aran Garcia-Bellido for his engagement of this analysis. Many thanks to all the people I have worked with over the course of the last six years. In particular, thanks to Otto Hindrichs for his constant help and support, and to my colleagues Kin Ho Lucien Lo, Mauro Verzetti, and Mario Galanti for many fruitful discussions. Finally, thanks to my parents for all their patience and support.

# Abstract

A measurement of the top quark Yukawa coupling from the $\mathrm{t\bar{t}}$ differential production cross sections in the semi-leptonic channel is presented. The $\mathrm{t\bar{t}}$ production cross section is sensitive to the square of the Yukawa coupling via weak corrections that enter when Higgs bosons are exchanged between the final state top and anti-top quarks. These weak corrections have a negligible effect in the inclusive $\mathrm{t\bar{t}}$ production cross section, but they lead to strong distortions of differential distributions near the threshold energy for pair-production (for low relative velocities of the top quarks). The differential distributions are therefore sensitive to anomalous values of the top Yukawa coupling and allow us to set an upper limit based on precision measurements.

This analysis is based on data collected by the CMS experiment in the LHC at 13 TeV corresponding to an integrated luminosity of 35.8 $\mathrm{fb}^{-1}$. In addition to a top reconstruction with at least four jets in the final state, a novel technique to reconstruct the $\mathrm{t\bar{t}}$ system with one missing jet is developed to enhance the experimental sensitivity in the low invariant mass region. We compare the data yields in the invariant mass of the top-antitop system, $M_{\mathrm{t\bar{t}}}$, the rapidity difference between top and antitop, $|y_{\mathrm{t}} - y_{\mathrm{\bar{t}}}|$, and the number of reconstructed jets with the

templates simulated for different values of the top Yukawa coupling and extract an upper limit on the top quark Yukawa coupling of 1.62 (Expected) at 95% CL.

# Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professors Regina Demina (advisor), Lynne Orr, and Steven Manly of the Department of Physics and Astronomy, and Professor David Mccamant of the Department of Chemistry.

The Compact Muon Solenoid (CMS) experiment at Large Hadron Collider (LHC) was designed and constructed before 2008. The author's contributions to the CMS experiment of the commissioning and operation were centered around the Silicon Strip Tracker subsystem. She helped to maintained the Tracker DAQ system and attended data-taking in the early stage of the LHC 'Run 2' starting on 2015. Analysis infrastructure where carried out by members of the CMS Collaboration, including: luminosity measurement; trigger efficiency measurement; lepton identification and isolation efficiency measurements; jet energy scale measurement; jet resolution measurement. Event selection criteria are based on recommendations of the CMS Collaboration. A variety of problems related to $t\bar{t}$ event reconstruction and systematic uncertainties were solved in collaboration with Otto Hindrichs, who developed the original reconstruction framework used in this analysis. All other work for the dissertation was completed by the author independently.

This work has been enabled by a wide variety of independent software projects, including: simulation packages, the analysis libraries, Root, RooFit, and RootStat; free and open source software like the GNU/Linux system and the Python programming language. Computing facilities utilized for this analysis are located at Fermi National Accelerator Laboratory (Fermilab), and the European Center for Nuclear Research (CERN).

Graduate study was supported by the U.S. Department of Energy, Office of Science, under Award Number DE-SC0008475 as well as by teaching assistantships in the University of Rochester Department of Physics and Astronomy.

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Introduction

The Standard Model (SM) representing our understanding of the universe at the smallest length scale. It describes three out of four fundamental forces in nature (electromagnetic, weak and strong interaction), via which all known elementary particles interact. SM provided accurate predictions for a wide variety of experimental results and predicted the existence of fundamental particles, such as the top quark observed in 1995 [1,2] and Higgs boson discovered in 2012 [3,4]. Although the SM has been very successful, it fails to explain several experimental anomalies. For example, it does not provide a viable candidate for dark matter [5–7], and the non-zero neutrino mass. More importantly, it does not include gravity.

To investigate physics at high energy scales, accelerators collide particles head on at high energies. The Large Hadron Collider (LHC) at CERN was built to collide protons at the highest ever achieved energy. Around the accelerator ring, two detectors, the Compact Muon Solenoid (CMS) and the A Toroidal LHC Apparatus (ATLAS), were built to detect resulting particles from collisions. The LHC and CMS detector are described in Chapter 3. Event reconstruction is introduced in Chapter 4.

| Fermion type | Generation 1st | Generation 2nd | Generation 3rd | $T$ | $T_3$ | $Y$ | $Q$ | Color charge |
|---|---|---|---|---|---|---|---|---|
| Leptons | $\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$ | $\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$ | $\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$ | $\frac{1}{2}$ | $\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ | $-1$ | $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ | 0 |
| | $e_R$ | $\mu_R$ | $\tau_R$ | 0 | 0 | $-2$ | $-1$ | 0 |
| Quarks | $\begin{pmatrix} u \\ d \end{pmatrix}_L$ | $\begin{pmatrix} c \\ s \end{pmatrix}_L$ | $\begin{pmatrix} t \\ b \end{pmatrix}_L$ | $\frac{1}{2}$ | $\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ | $\frac{1}{3}$ | $\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$ | $r, g, b$ |
| | $u_R$ | $c_R$ | $t_R$ | 0 | 0 | $\frac{4}{3}$ | $\frac{2}{3}$ | $r, g, b$ |
| | $d_R$ | $s_R$ | $b_R$ | 0 | 0 | $-\frac{2}{3}$ | $-\frac{1}{3}$ | $r, g, b$ |

**Table 1.1:** Summary of the three generations of fermions and their properties. $T$, $T_3$ and $Y$ represent total weak isospin, the third component of weak isospin and hypercharge respectively. $Q = T_3 + \frac{1}{2}Y$ represents the electric charge [8].

## 1.2 The Standard Model

In SM, all matter particles are described as spin-$\frac{1}{2}$ fermionic fields, whereas force carriers are represented by spin-1 bosonic fields. In addition, a spin-0 field, called the Higgs field, is included in the SM to give masses to all matter and force fields.

Spin-$\frac{1}{2}$ fermions are classified as quarks or leptons. One major difference between leptons and quarks is that only quarks carry color charges, and thus can interact via strong interaction. Both leptons and quarks carry electromagnetic, $Q$ and weak charges, or weak isospin, $T$ that behaves like normal spin, its third projection being $T_3$ and hypercharge, $Y$. Table 1.1 summarizes various quantum numbers of each fermion in the SM. Four spin-1 gauge bosons act as force carriers for the electromagnetic, weak and strong interaction. Furthermore, Table 1.2 shows properties of four gauge bosons and the Higgs boson.

## 1.3 Strong interaction

The strong interaction can be mathematically formulated as a $SU(3)$ gauge theory, called quantum chromodynamics (QCD). This theory gives rise to eight massless

| Boson type | $T$ | $T_3$ | $Y$ | $Q$ | Color charge | Force |
|---|---|---|---|---|---|---|
| photons | 0 | 0 | 0 | 0 | 0 | Electromagnetic |
| $Z$ | 0 | 0 | 0 | 0 | 0 | Weak |
| $W$ | 1 | $\pm1$ | 0 | $\pm1$ | 0 | Weak |
| gluons | 0 | 0 | 0 | 0 | $r\bar{g}, r\bar{b}, g\bar{r}, g\bar{b}, b\bar{r}, b\bar{g}, \frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g}), \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$ | Strong |
| $H$ | $\frac{1}{2}$ | $-\frac{1}{2}$ | $+1$ | 0 | 0 | - |

**Table 1.2:** Summary of gauge boson and their properties [8].

gauge fields called gluons. Gauge invariance dictates gluons to be massless but does not impose any constraints on quark fields. Each quark field can carry a color charge, red($r$), green ($g$) or blue($b$). The force carriers, gluons, also have color charge. As a result QCD has an interesting property, called asymptotic freedom [9], arising from color charge screening in vacuum. This means that the strong coupling strength decreases for higher energy scales, in contrast to the electromagnetic force. This important property ensures the validity of perturbative QCD at high energy scales [10].

## 1.4 Electroweak interaction

The electromagnetic and weak interaction are collectively described by a single $SU(2) \times U(1)$ gauge theory in the SM. The $SU(2)$ gauge group gives rise to three gauge fields $W^i_\mu$ ($i = 1, 2, 3$ and $\mu$ is the Lorentz index), which only couple to the left handed chiral component of the lepton field. In addition, $U(1)$ gauge group results in the hypercharge gauge field $B_\mu$.

The four electroweak gauge field mix to give observable gauge boson fields, W, Z and photon field [11],

$$W^\pm_\mu = \frac{1}{2}\left(W^1_\mu \mp iW^2_\mu\right) \tag{1.1}$$

$$Z_\mu = \cos\theta_W W^3_\mu - \sin\theta_W B_\mu \tag{1.2}$$

$$A_\mu = \sin\theta_W W^3_\mu + \cos\theta_W B_\mu \tag{1.3}$$

where $A_\mu$, $Z_\mu$ and $W_\mu^\pm$ are the photon, Z boson and W boson fields respectively. The Weinberg angle $\theta_W$ is expressed in terms of the $SU(2)$ and $U(1)$ gauge coupling.

Experimentally observed $W$ and $Z$ bosons [12–15] are massive. Thus, the theory requires an extra mechanism to generate masses for gauge bosons. This is discussed in the next section.

## 1.5 The Higgs Mechanism

The formulation of strong and electroweak interactions as gauge theories requires the gauge bosons ($W$ boson, $Z$ boson, photon and gluon) to be massless. It is not compatible with the experimental observation of massive $W$ and $Z$ bosons [12–15]. To provide masses for the electroweak gauge bosons, the Higgs mechanism [16–18] was proposed as a special case of spontaneous symmetry breaking. Spontaneous symmetry breaking is a phenomenon in which a physical system respects a certain symmetry, but the ground state (or multiple degenerate ground states) of this system does not.

In addition to the gauge and matter fields in SM, a complex scalar field that is an $SU(2)$ doublet is introduced

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.4}$$

where the quantum numbers of $\phi^+$ and $\phi^0$ are summarized in Table 1.3. The field is described by the following Lagrangian:

$$L_{\text{Higgs}} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \tag{1.5}$$

with $D_\mu = \partial_\mu + igT^aW_\mu^a + ig^{'}B_\mu$. The potential $V(\phi)$ is proposed to be

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.6}$$

with $\mu^2 < 0$ and $\lambda > 0$. This potential leads to infinite classical ground state solutions with $|\phi|^2 = -\frac{\mu^2}{2\lambda}$. Quantum-mechanically, the ground state becomes a

| Field | $T$ | $T_3$ | $Y$ | $Q$ |
|---|---|---|---|---|
| $\phi^+$ | $\frac{1}{2}$ | $+\frac{1}{2}$ | 1 | 1 |
| $\phi^0$ | $\frac{1}{2}$ | $-\frac{1}{2}$ | 1 | 0 |

**Table 1.3:** Summary of quantum numbers of the Higgs field [8].

degenerate set of vacuum states with a non-zero expectation value $\langle 0| \, |\phi|^2 \, |0\rangle = -\frac{\mu^2}{2\lambda}$. The vacuum state can be conveniently chosen in the unitary gauge such that the scalar field satisfies

$$\langle 0| \, \phi \, |0\rangle = \begin{pmatrix} 0 \\ \sqrt{\frac{\mu^2}{2\lambda}} \end{pmatrix} \tag{1.7}$$

As a result, Higgs field can be expanded around its minimum

$$\phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix} \tag{1.8}$$

with $v$ as the vacuum expectation value. Since the scalar field is an $SU(2)$ doublet and has a hypercharge, it interacts with the $W$ and $Z$ bosons, as described by the following Lagrangian:

$$L_{\mathrm{EW}} = \frac{1}{2}(\partial^\mu H)(\partial_\mu H) - \frac{1}{4}g^2 W^+ W^- (v+H)^2 + \frac{1}{8}g_Z^2 ZZ(v+H)^2 - V\left[\frac{1}{2}(v+H)^2\right] \tag{1.9}$$

where the $v^2$ terms provide the mass terms for $W$ and $Z$ boson in a gauge-invariant way. The gauge fields for the $W$ and $Z$ boson are defined in Equation 1.3.

The Higgs mechanism allows the introduction of massive electroweak gauge bosons in SM without violating gauge invariance. Almost fifty years after the original papers on the Higgs mechanism, the Higgs boson was discovered by the ATLAS and CMS collaboration with a mass of 125 GeV [3, 4].

## 1.6 Top-Higgs Yukawa coupling

The Higgs boson is also postulated to generate mass terms for leptons and quarks. After spontaneous symmetry breaking as described in Section 1.5, each fermion acquires a mass through a Yukawa-type coupling with the Higgs boson, with the coupling strength being proportional to the fermion mass. The top quark, as the heaviest known elementary particle discovered so far, interacts with the Higgs bosons with the largest Yukawa coupling. This interaction can be probed by producing the Higgs in the fusion of a top quark-antiquark pair or through Higgs radiation from a top quark, as known as the ttH production. Figure 1.1 and 1.2 shows the corresponding diagrams for the ttH production. Observation of the ttH production has been reported in 2018 [19].

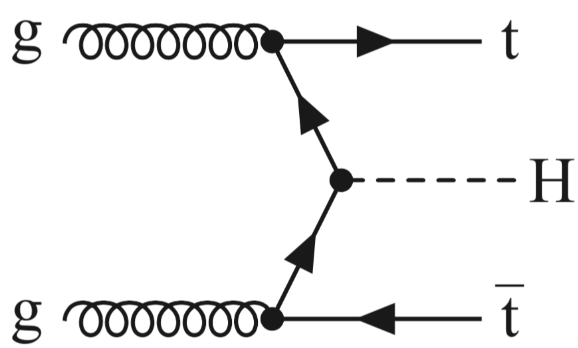


**Figure 1.1:** Production of Higgs boson via the fusion of a top quark-antiquark pair.

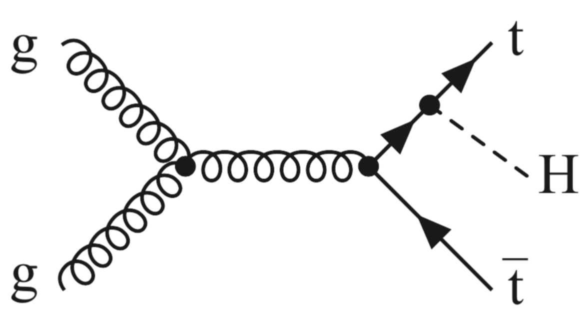


**Figure 1.2:** Production of Higgs boson via radiation from a top quark.

In this thesis, we are interested in measuring the top-Higgs Yukawa coupling using virtual Higgs exchange in top pair productions. A deviations of the Yukawa coupling from the SM value can induce strong distortions in top kinematic distributions.

## 1.7 Top quark physics in LHC

The top quark is the heaviest known elementary particle discovered so far. Since its discovery in 1995 by the CDF and D0 experiments at the Tevatron, substantial efforts have been spent on studying its properties. The LHC provides a unique opportunity to precisely measure the top quark production mechanisms and properties, as with the unprecedented center-of-mass energy of $\sqrt{s} = 13$ TeV and large instantaneous luminosity of the accelerator, the LHC can produce top quarks abundantly.

### 1.7.1 Top quark pair production

In LHC, top quark pairs are predominantly produced via QCD processes. The main production mechanisms are quark-antiquark annihilation and gluon fusion. Figure 1.3 shows the leading order Feynman diagrams contributing to top quark pair production in proton-proton collisions. Quantum amplitudes and hence partonic cross sections represented by these diagrams are convoluted by parton distribution functions (PDF) to compute the cross sections of top pair production. Due to the large strong coupling constant ($\approx 0.1$), higher order corrections can give sizable contributions to the calculations. Due to asymptotic freedom, at higher energy scales, the calculations are more reliable as the strong coupling constant becomes smaller.

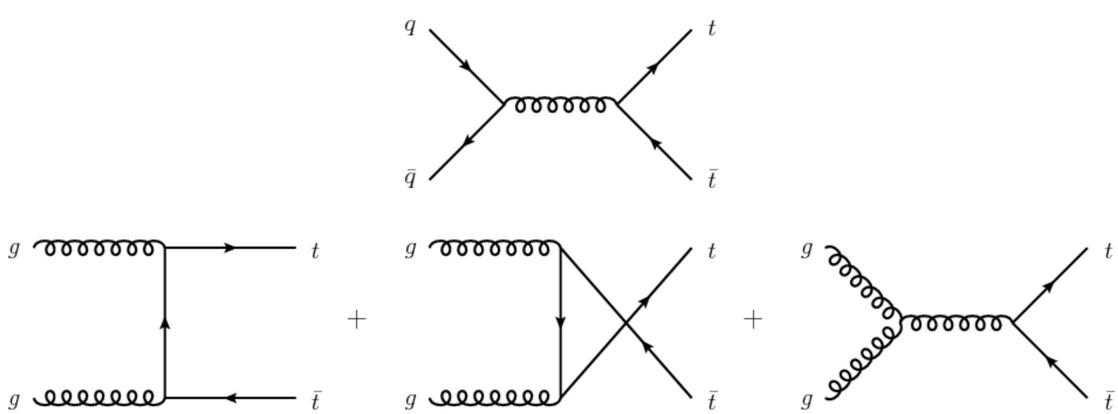


**Figure 1.3:** Lowest order diagrams contributing to top quark pair production, in quark-antiquark annihilation (top) or gluon-gluon fusion (bottom).

### 1.7.2 Inclusive cross section

Precise measurement of top pair production cross section is important to establish a good understanding of SM. First measurements of the inclusive $t\bar{t}$ cross section were published by the ATLAS and the CMS collaboration in 2010 [20, 21]. Subsequently , with more data, an unprecedented level of statistical and systematic precision is reached in measurements of inclusive $t\bar{t}$ cross sections. Figure 1.4 shows a summary of CMS measurements of inclusive cross sections at various center-of-mass energies.

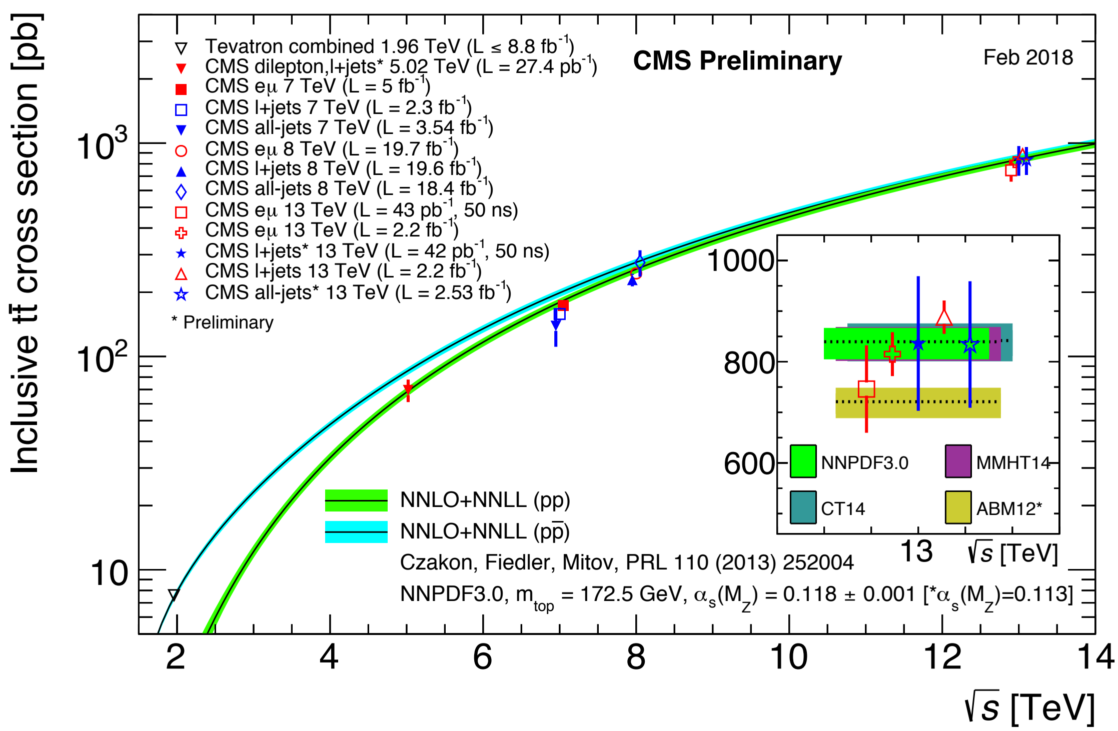


**Figure 1.4:** Top quark pair cross section summary of CMS measurements in the di-leptonic and semi-leptonic channel, compared to theory calcultions at NNLO+NNLL accuracy.

### 1.7.3 Differential cross section

Additional information can be gained from measurements of differential distributions. Comparisons between theoretical calculations with experimental results can improve choices of low energy QCD models and scale parameters. Furthermore, PDF of gluons, especially in the regime with large momentum fractions, can be constrained with top pair differential cross section measurements. New physics decaying into top pairs can be visible in various top pair distributions.

Various top pair kinematic distributions have been measured in the LHC [22], with unprecedented statistical precisions. In general, data are very well described by various theoretical calculations up to an energy scale of about 1 TeV. In this thesis, we explore the region near the threshold of the top pair production since it is particularly sensitive to Electroweak corrections resulting from virtual Higgs exchange.

# Chapter 2

# Constraining the top-Higgs Yukawa coupling with $t\bar{t}$ production

## 2.1 Sensitivity in the top pair production

Precise measurements of the top quark pair ($t\bar{t}$) production cross section and the decay properties of the top quark provide crucial information for testing the standard model (SM) and searching for new phenomena. The current state of the art calculations provide next-to-next-to-leading order (NNLO) calculations in perturbative QCD [23] for the $t\bar{t}$ production cross section. Weak corrections affect the cross section at the loop-induced order $\alpha_s^2 \alpha_{weak}$ (Fig. 2.1) so they make only a small contribution to the total cross section and are not implemented in the Monte Carlo (MC) event generators. However, in kinematic regions with large momentum transfer, weak corrections become significant and may lead to strong distortions of differential distributions. In addition, it has been shown that in the threshold region, which corresponds to kinematic regions with small relative velocity between the top and antitop quarks, the $t\bar{t}$ cross section is sensitive to the top-quark Yukawa coupling ($Y_t$) [24]. For example, doubling the Yukawa coupling would lead to a change in the cross section of about 9%, which is larger than the current experimental sensitivity of around 6% [25]. Thus, it is feasible to constrain the top-quark Yukawa coupling using $t\bar{t}$ differential cross section at the LHC.

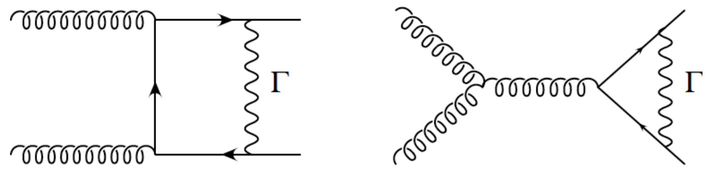


**Figure 2.1:** Example diagrams for gluon-induced production of $t\bar{t}$ with the virtual electroweak corrections. Γ stands for all contributions from gauge, Goldstone and Higgs bosons exchange.

A detailed study of the differential $t\bar{t}$ kinematics close to threshold could, therefore, determine the strength of $Y_t$, or at least provide an upper limit. In this analysis, we calculate the weak correction factors using HATHOR [26] and apply them at the parton level on existing $t\bar{t}$ simulated samples, fully reconstruct them, and use them as templates to fit the differential distributions in semi-leptonic $t\bar{t}$ decays to determine the most likely $Y_t$ value. The most sensitive distributions to Yukawa coupling are the distributions in the invariant mass of the top quark pair $M_{t\bar{t}}$ and the rapidity difference between the top quark and antiquark $\Delta y_t = y_t - y_{\bar{t}}$. The analysis covers the phase space from threshold (the lowest reconstruction events from ≈200 GeV) to 2 TeV in $M_{t\bar{t}}$ and from 0 to 6 in $|\Delta y_t|$. Low $M_{t\bar{t}}$ and small $|\Delta y_t|$ regions are most sensitive to $Y_t$. Due to the differences in the reconstruction quality and background contamination it beneficial to treat events with different number of jets separately.

Top quarks decay almost exclusively through $t \to Wb$ and the final topology depends on the $W$ boson decays. When one $W$ boson decays leptonically and the other hadronically, $t\bar{t} \to W^+b\,W^-\bar{b} \to \ell^+\nu b\,q\bar{q}'\bar{b}$, the final state at leading order consists of an isolated lepton (electron or muon in this analysis), missing energy (due to neutrino), and four jets (two from hadronization of b-quarks and two from light quarks). This final state has a sizable branching ratio (34%), small backgrounds, and allows for the full kinematic reconstruction of the original top candidates. For events containing four or more jets in the final state this analysis follows the same methodology employed in the measurement of the top quark pair differential production cross section with an integrated luminosity of 35.8 $\text{fb}^{-1}$ at 13 TeV [25]. In addition, we introduce a novel algorithm to reconstruct the $t\bar{t}$ pair when only three jets are reconstructed. In this analysis we do not attempt to determine the parton level kinematics of top and anti-top quarks. Instead, once the MC samples are reweighted with different $Y_t$ values and reconstructed, they are compared to background-subtracted data distributions at the reconstruction level.

## 2.2 Electroweak correction

The electroweak (EW) corrections to the $t\bar{t}$ production cross section were calculated [27] before the top quark discovery, and are found to have a very small effect on the total cross section. They can, however, have a sizable impact on differential distributions and the top-quark charge asymmetry. There is no interference term of order $\alpha_s\alpha$ between lowest order strong and neutral current amplitudes in the quark induced processes. Weak corrections start entering the cross section at loop-induced order $\alpha_s^2\alpha$, as shown in Fig. 2.1. These corrections are negative for both quark- and gluon induced diagrams when the parton energies are large. The corrections for quark induced top production are about twice of those of the gluon induced processes, due to the non-zero weak charge of the initial state quarks. Top quark production at the LHC occurs mainly through gluon fusion, but if we require large top quark transverse momenta, the phase space is dominated by quark-antiquark annihilation. Thus, weak corrections will be enhanced in the large top $p_{\mathrm{T}}$ regime due to the relative increase of the $q\bar{q}$-initiated corrections. The majority of weak corrections does not depend on $Y_t$. The contributions linear in $Y_t$ arise from the production of an intermediate s-channel Higgs boson through a closed b-quark loop, which we can ignore due to the small b-quark mass. Finally, we are only left with a quadratic dependence on the top-quark Yukawa coupling arising from the Higgs boson exchange between the final state top quark and antiquark, which involves two $Ht\bar{t}$ vertices, as shown in Fig. 2.1.

In HATHOR [26], we calculate the partonic cross section value including the EW corrections at leading order $\mathcal{O}(\alpha_s^2\alpha)$ for given values of top mass $m_t$, partonic center-of-mass energy squared $s$ for the $t\bar{t}$ system, and the cosine of the scattering angle $\theta$ for top in the center-of-mass frame. HATHOR supplies such values in the form of $\sigma(m_t, s, cos\theta)$. In our case, $\sqrt{s}$ is the invariant mass of $t\bar{t}$ system and $cos\theta$ can be recast in terms of the invariant $\Delta y_t$. The $m_t$=172.5 GeV is assumed fixed, and will be treated as a source of systematic uncertainty by reconstructing events with $\Delta m_t = \pm 1$ GeV. We therefore use HATHOR to extract a two-dimensional correction factor that is defined to be the ratio of the $t\bar{t}$ production cross section with electroweak corrections to that without corrections. This is done for different values of $Y_t$ = 0, 1, 2, or 4 in bins of $M_{t\bar{t}}$ and $\Delta y_t$. We then apply this correction factor on each $t\bar{t}$ POWHEG MC event at the parton level.

Figure 2.2 shows the leading order $t\bar{t}$ production cross section, $t\bar{t}$ production cross section with weak correction assuming $Y_t = 2$, and their ratio in bins of $M_{t\bar{t}}$ and $\Delta y_t$. We apply this ratio to reweigh each $t\bar{t}$ MC event according to its value of $M_{t\bar{t}}$ and $\Delta y_t$. Fig. 2.3 shows the relative EW correction on the NLO QCD production cross section as a function of $M_{t\bar{t}}$ and $\Delta y_t$. Several $Y_t$ scenarios are shown. It is important to note that while non-SM $Y_t$ values can lead to large distortions in the $M_{t\bar{t}}$ and $\Delta y_t$ spectra, especially in the low $M_{t\bar{t}}$ region, experimental resolution can smear out such dramatic distortions. The experimental effects are discussed in the following Chapters.

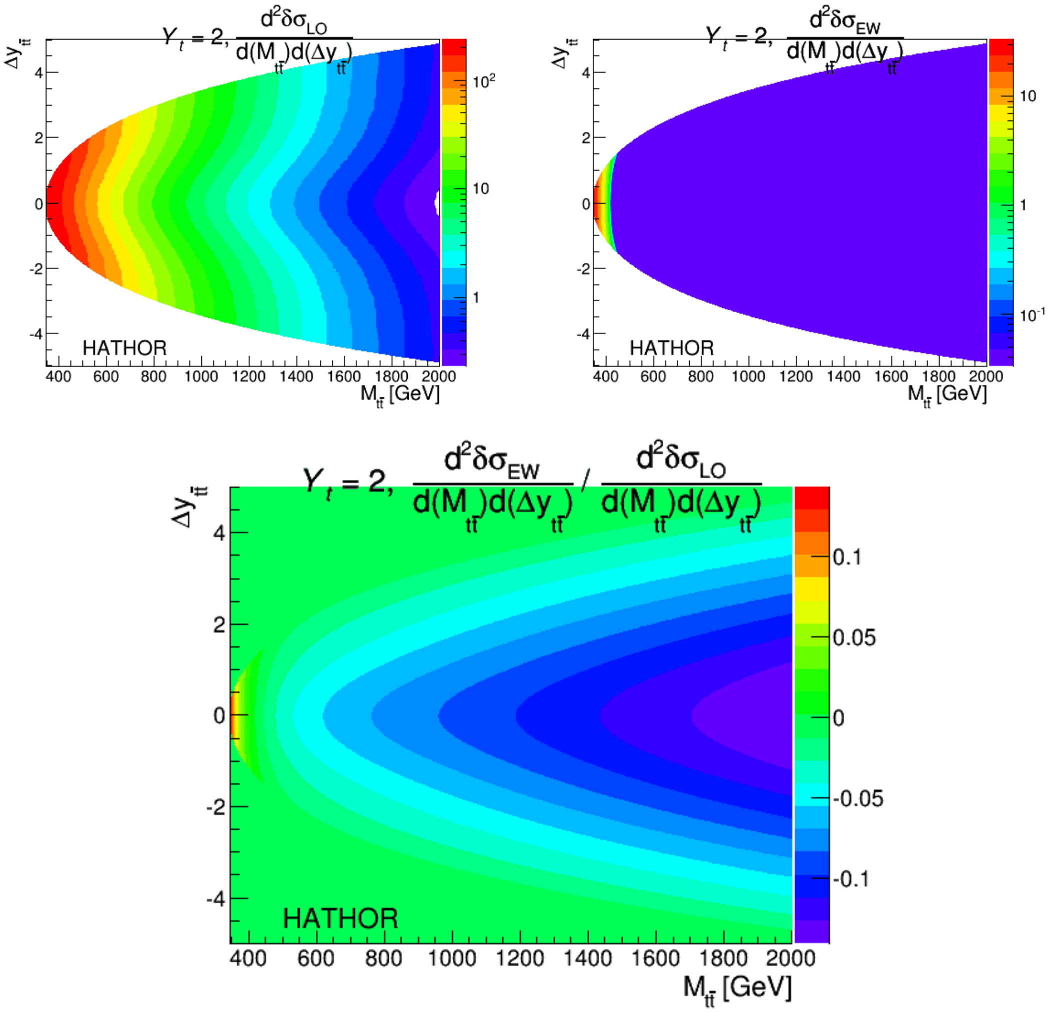


**Figure 2.2:** The $M_{t\bar{t}}$ and $\Delta y_t$ distributions for the leading order $t\bar{t}$ production cross section in pb/GeV$^2$ (top left) and the electroweak correction calculated in HATHOR (top right) for $Y_t$ =2. The reweighting factors for implementing electroweak corrections to the MC samples are given by the ratio of the weak correction over the leading order production cross section (shown in the bottom plot).

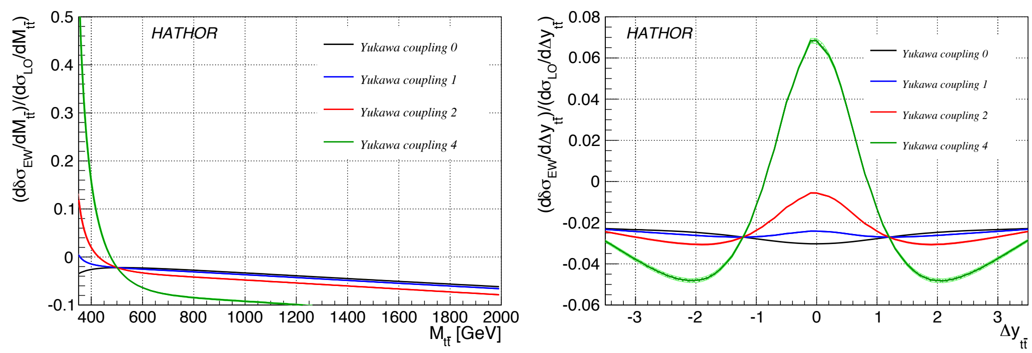


**Figure 2.3:** The ratio of electroweak correction over the leading order production cross section for the sensitive kinematic variables for different values of Yukawa coupling. The lines contain an uncertainty band derived from the dependence of the EW correction on the top mass up and down by 1 GeV. We can see that the effect of the top mass is very small at the parton level.

# Chapter 3

# Large Hadron Collider and CMS experiment

## 3.1 The LHC

The Large Hadron Collider (LHC) is the world's largest accelerator facility located at the border of Switzerland and France. It is operated by the European Centre for Nuclear Research (CERN). The accelerator counter-circulates beams of protons or heavy ions. The purpose of the LHC is to extend the understanding of some of the fundamental open questions in Physics to new frontiers in energy and luminosity. Its design center of mass collision energy of $\sqrt{s}$ =14 TeVand instantaneous luminosity of $10^{34}$ $cm^{-2}$ $s^{-1}$ are orders of magnitude greater than those of the previous state of the art collider, Fermilab's Tevatron.

Figure 3.1 shows a schematic view of this accelerator complex. LHC proton beams are first accelerated to 50 MeV by the Linear Accelerator 2 (LINAC2). In successive stages through the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), the protons are accelerated to 1.4 GeV, 26 GeV, and 450 GeV, respectively, before being injected into the LHC. Inside the LHC, the bunches are steered by 1200 superconducting dipole magnets and accelerated further to the desired energy 6.5 TeV(current running energy). The LHC ring is housed by a 27 km long tunnel located as far as 175 meters under ground together with four detectors (ALICE, ATLAS, LHCb and CMS) built to record the results of the high energy collisions.

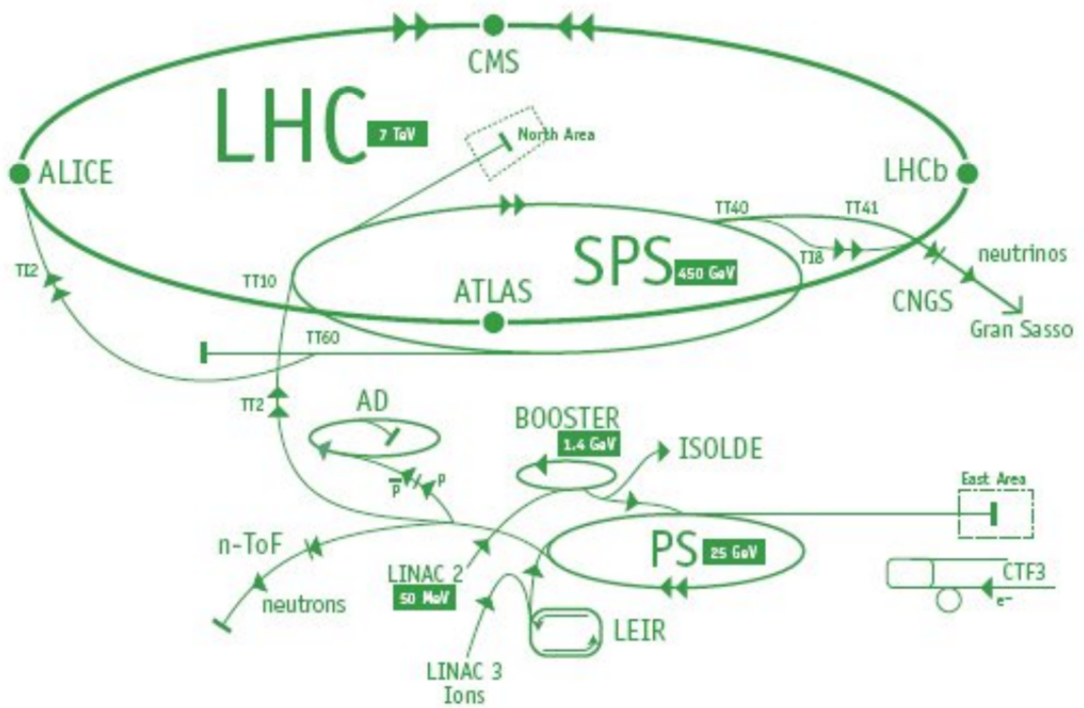


**Figure 3.1:** A schematic view of the CERN accelerator complex for the LHC. From [28].

## 3.2 The CMS detector

The CMS detector is a cylindrically symmetric construction around the LHC beamline, centered on the beam intersection point, to precisely measure the energy and momenta of debris particles from the collisions. The detector consists of four main sub-detectors and a superconducting solenoid. The magnetic field defects charged particles along spiral paths with their axis parallel to the field and beamline, each with a radius proportional to its transverse momentum of the particle. Figure 3.2 shows a schematic view of the CMS detector and its sub-detectors.

The first layer of the CMS detector is the silicon tracker, which records the charged particles trajectories. The second layer is the electromagnetic calorimeter (ECAL), which absorbs and measures the energy of photons and electrons produced through electromagnetic showering. The third layer is the hadron calorimeter (HCAL), which specializes in absorbing and measuring energies of hadrons from hadronic showering. The tracker, ECAL and the inner part of HCAL are located inside a 3.8 T superconducting solenoid. Outside the solenoid, are the iron return yoke and the final detector layer, the muon chambers designed for muon detection since muons are the only particles (apart from neutrinos) that are not stopped by the calorimeters. Figure 3.3 shows the CMS slice and how particles interact with each of sub-detectors. Details of the sub-detectors and triggers are discussed in the following subsections.

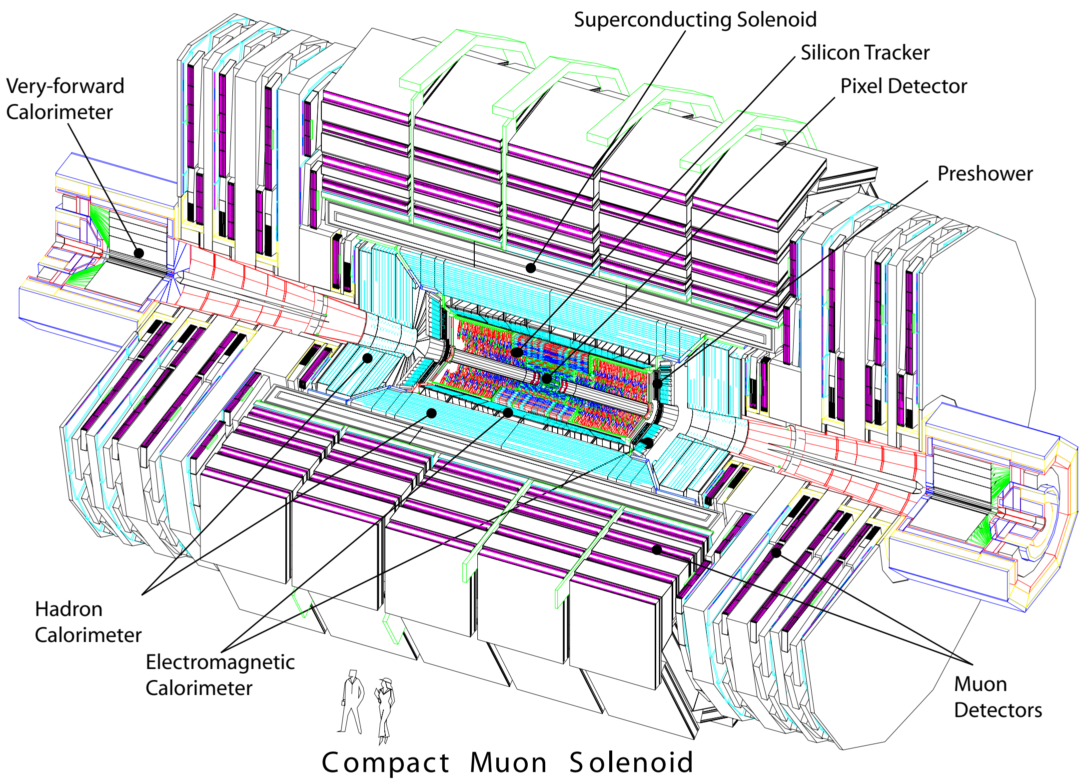


**Figure 3.2:** A representation of the CMS detector and its associated sub-detectors. From [29].

### 3.2.1 CMS coordinate

The CMS coordinate system is oriented such that the $x$-axis points to the center of the LHC ring, the $y$-axis points vertically upward and the $z$-axis is in the direction

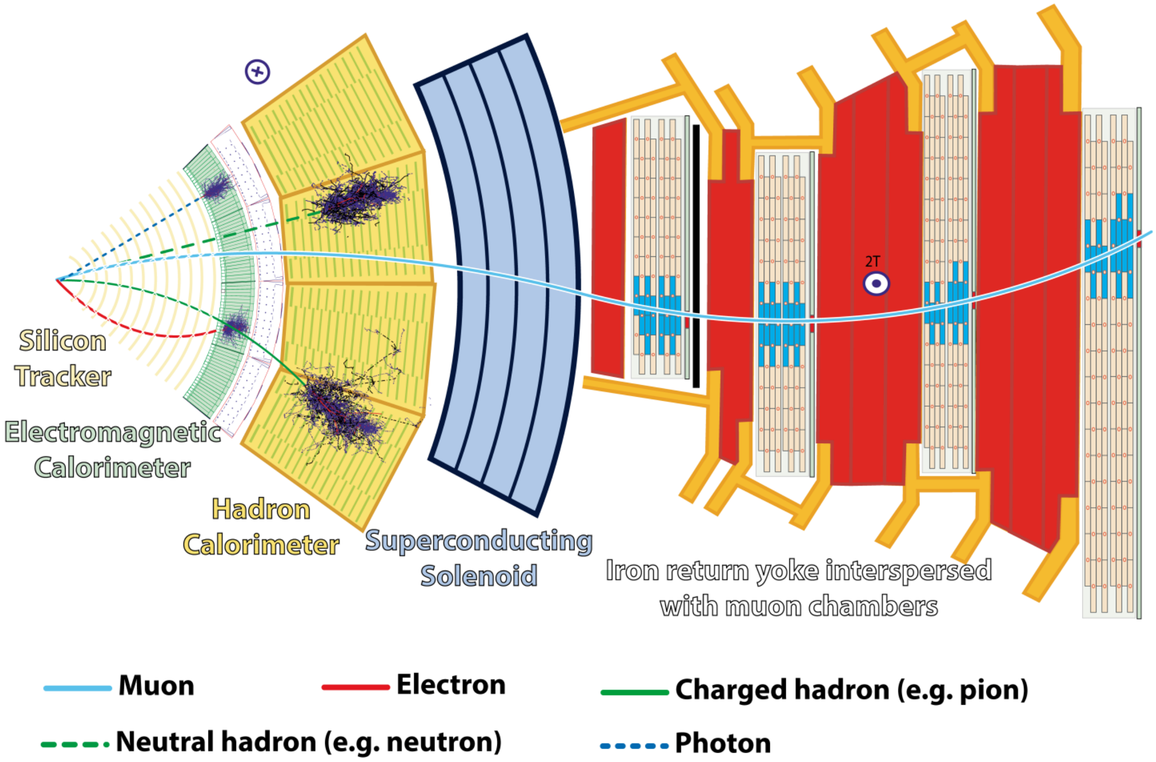


**Figure 3.3:** Slice showing CMS sub-detectors and how particles interact with them. From [29].

of the beam. The azimuthal angle $\phi$ is measured from the $x$-axis in the $x - y$ plane and the radial coordinate in this plane is denoted by $r$. The polar angle $\theta$ is defined in the $r - z$ plane and the pseudorapidity is defined as $\eta = -ln \tan(\frac{\theta}{2})$. Rapidity is used to define a measure of angular separation between particles commonly used in particle physics $\Delta R \equiv \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$, which is Lorentz invariant under a boost along the longitudinal (beam) direction. Often, the rapidity term in this expression is replaced by pseudorapidity, yielding a definition with purely angular quantities $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, which is Lorentz invariant if the involved particles are massless. The momentum component transverse to the beam direction, denoted by $p_T$, is computed from the $x$- and $y$-components.

### 3.2.2 The tracker

CMS features an all silicon tracker with a total active area of 200 $m^2$. The tracking detector is divided into a pixel detector and a strip detector. The innermost detector is the Pixel Silicon Tracker, one meter long with three barrel layers and two endcap layers on each end, extending to $r < 20$cm, $|\eta| < 2.5$. The Pixel detector consists of 66 million $100 \times 150$ $\mu$m pixels. It is used to precisely pinpoint the primary and secondary collision vertices of particle tracks near the beamline. An position occupancy of about 15 $\mu$m can be achieved.

The Silicon Strip Tracker with a length of 5.8 m and a diameter of 2.4 m complements the Pixel detector. It detect the trajectories of charged particles in the region between $r > 20$ cm and $r > 120$ cm. The Strip Tracker is composed of four subsystems: the four-layer Tracker Inner Barrel (TIB), the six-layer tracker outer barrel (TOB) and on each side three-disk Tracker Inner Disks (TID) and nine-disk Tracker Endcaps (TEC). An $r - z$ view of the tracker geometry is shown in Figure 3.4. The Strip tracker, with 9.6 million silicon strips covers the rapidity region to $|\eta| < 2.5$. It detects charged particles with 100 $\mu$m scale azimuthal resolution.

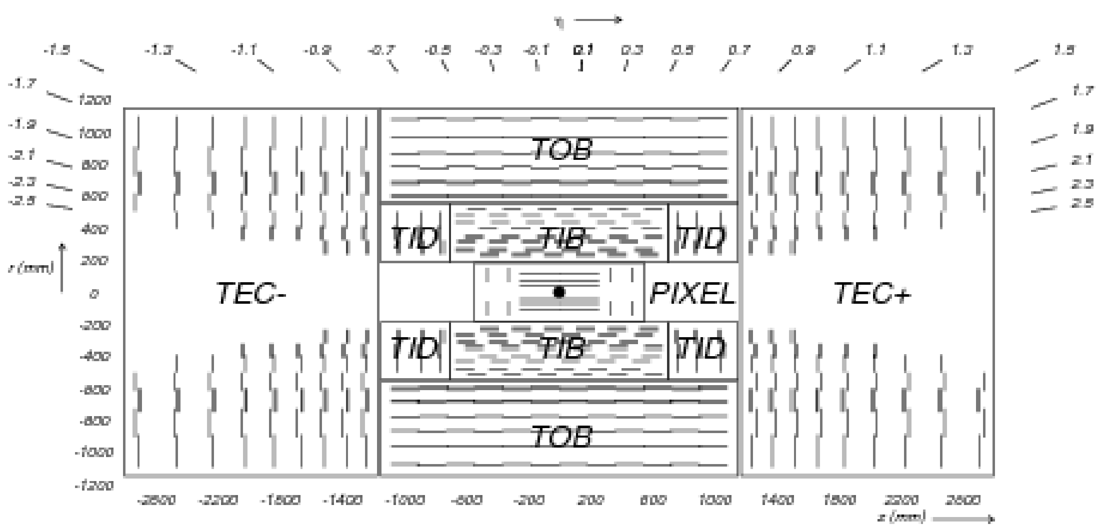


**Figure 3.4:** The cross section of the CMS tracker. Detector modules are represented by black lines. From [30].

### 3.2.3 The electromagnetic calorimeter

The CMS ECAL surrounds the silicon tracker and is used to measure the energy of electromagnetically interacting particles, mainly electrons and photons. The ECAL with a large pseudorapidity coverage up to $|\eta| < 3$ is divided into barrel detector, endocap detector, and Preshower as illustrated in Figure 3.5. The ECAL barrel and endcap are situated at a radial distance of 1.29 m and longitudinal distance of 3.15 m from the interaction point, respectively. The Preshower detector is placed in front of the endcaps to guarantee a reliable discrimination of single photons and photons produced in pairs in neutral pion decays.

The ECAL is constructed of scintillation crystals made from lead tungstate ($PbWO_4$), and the choice of the material is governed by its fast response time and high radiation resistance. The crystals are transparent, and when electrons or photons are stopped, light from the shower propagates outward where it is measured by photo diodes and photo-triodes. The ECAL energy resolution as a function of the electron energy, $E$, measured in $GeV$ is 0.5%/$\sqrt(E)$.

### 3.2.4 The hadron calorimeter

The HCAL, as illustrated in Figure 3.6, is divided into a barrel part (HB and HO) at $|\eta| < 1.3$, an endcap (HE) on each side at $1.3 < |\eta| < 3$ and a forward calorimeter (HF) extending up to $|\eta| < 5.2$ to achieve a most hermetic detector coverage. The barrel part includes the inner part HB, inside the solenoid magnet, and the outer part HO, which located between the solenoid magnet and the muon detectors. The HB is located between the outer extent of the ECAL $r = 1.77$ m and the inner extent of the magnetic coil $r = 2.95$ m. Thus, the HB is supplemented by the HO in order to absorbs any late-starting or highly-penetrating showers. The HF is positioned at a longitudinal distance of 11.2 m from the interaction point outside the solenoid magnet.

The HO, HB and HE consist of layers of brass absorber interleaved with scintillating plastic tiles. Brass absorbers, as they are non-magnetic and have a short interaction length ($\sim$ 16 cm), ensure a full embedding of hadron showers. The HF

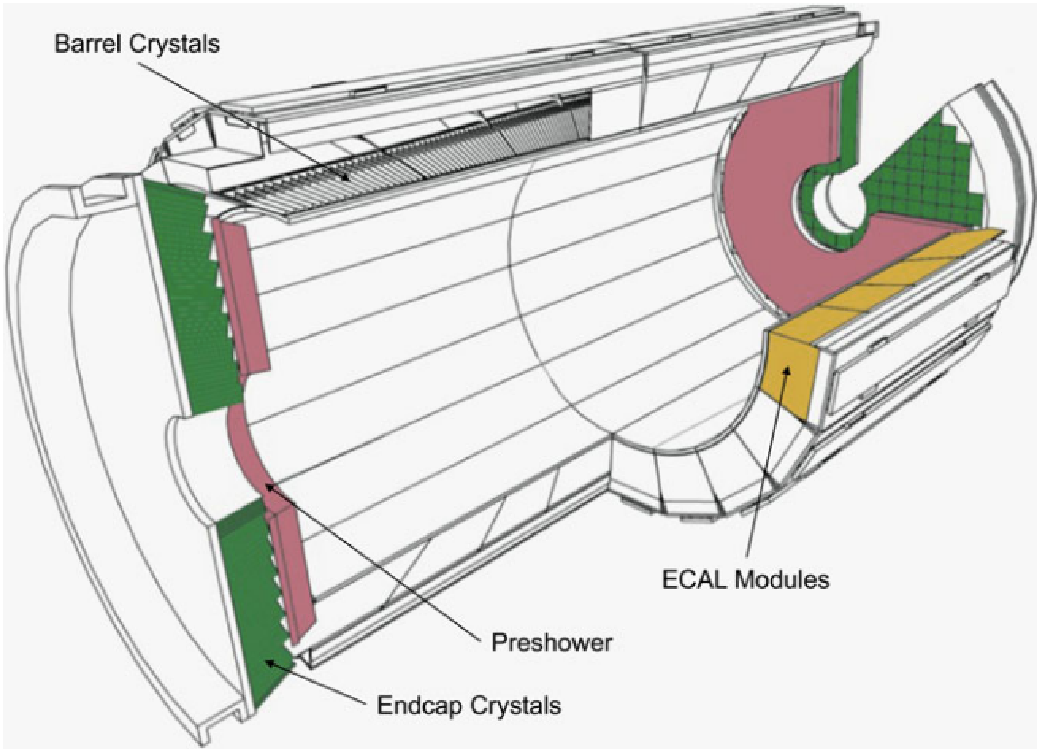


**Figure 3.5:** Layout of the CMS electromagnetic calorimeter. From [31].

is made from steel absorber plates composed of 5 mm thick grooved plates with quartz fibers inserted as active medium.

### 3.2.5 The muon system

The muon system located outside the solenoid magnet is the outermost part interleaved with an iron return yoke for the magnetic field. The muon system is equipped with gaseous detector chambers for muon identification and momentum measurement. There are three different types of gaseous detectors integrated into the system. In the barrel part, where both the muon rate and the neutron induced background are small and the magnetic field is very low, drift tube (DT) chambers are used. However, in the endcaps the muon and the background flux is much higher. A cathode strip chambers (CSCs), which provide a faster response, a higher granularity, and a better resistance against radiation are used. The muon system is also instrumented throughout with resistive plate chambers (RPC), which provide

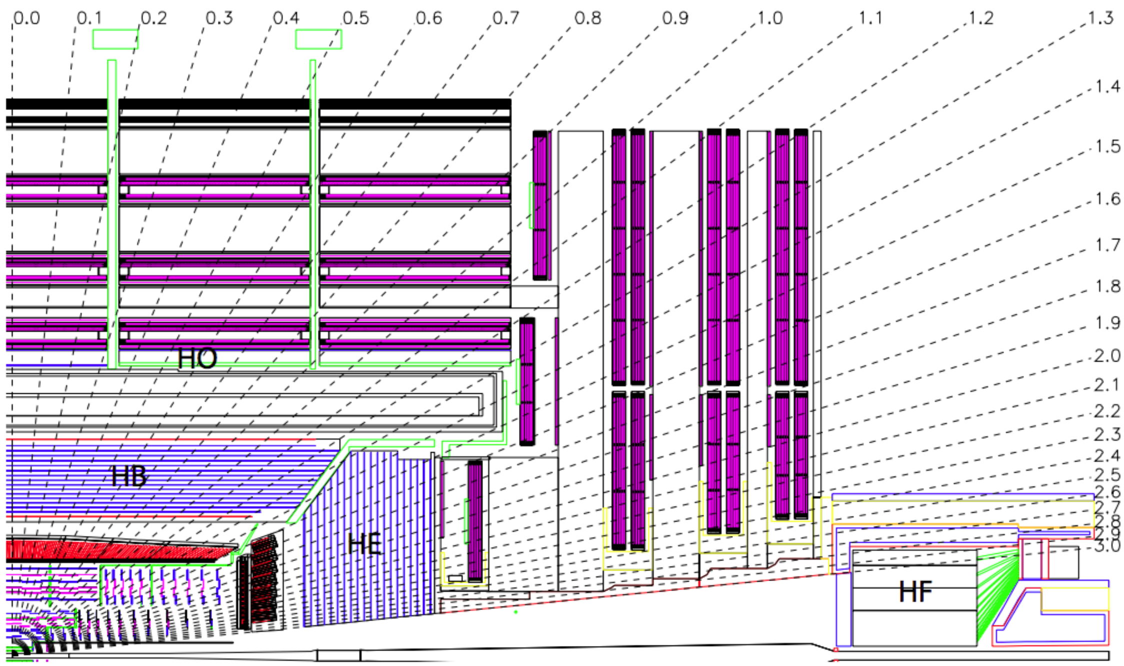


**Figure 3.6:** Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. From [31].

a coarser spatial resolution, yet, a fast response of a time resolution that is used in the muon trigger system. The layout of the muon system can be seen in Figure 3.7.

### 3.2.6 The Trigger system and data acquisition system

Data from all sub-detectors are read out and passed to the trigger system, which decides whether to record the event for permanent storage or ignore it. Trigger must ensures a maximum data recording efficiency after application of real-time (online) selective requirements to reduce the event rate. The trigger is a multi-layer system of selective online algorithms performing both locally on the sub-detector level or globally, when outputs from local algorithms are merged. Up to 128 algorithms can be executed in parallel in order to cope with the unprecedented high event rate of 40 MHz .

If an event is accepted by the online trigger, the full detector information is read out at a rate of up to 100 kHz as the input of subsequent further discriminations, i.e.

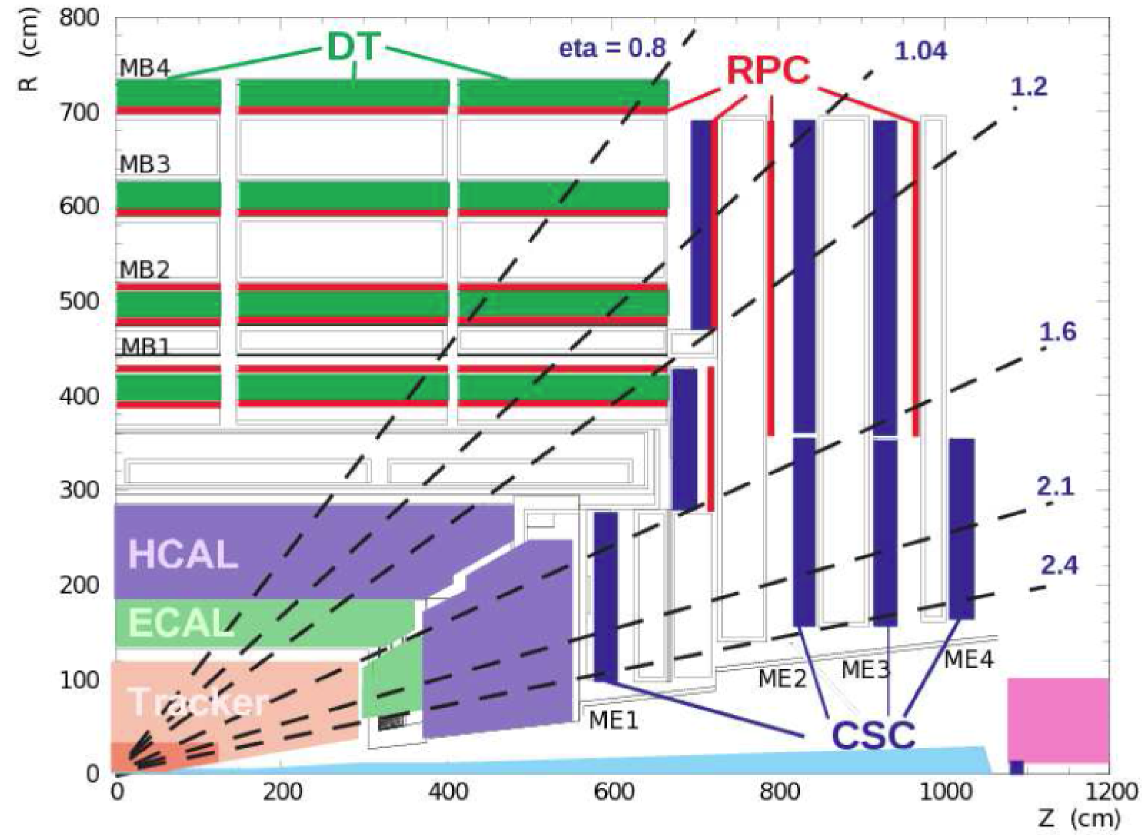


**Figure 3.7:** A schematic view of a quadrant of the CMS muon system. The locations of the Drift Tube (DT), Resistive Plate Chamber (RPC) and Cathode Strip Chamber (CSC) subsystems are displayed. From [31].

high level trigger (HLT). The HLT menu is composed by a set of triggers addressing specific selections of physics objects and event topologies, and it further reduces the event rate to 1 kHz. If an event is accepted both by the online trigger and the HLT, it is stored for further data analysis. The data are also transferred to each data storage site around the world, so that the CMS collaboration members in different regions can access data for physics analyses.

# Chapter 4

# Physics object reconstruction and event selections

## 4.1 Particle flow algorithm

The particle flow (PF) algorithm [32] is used in CMS experiment to combine the information from all sub-detectors to identify muons, electrons, photons, charged hadrons, and neutral hadrons. Charged particles with a track pointing towards a primary vertex other than the main vertex, defined as the vertex with the highest sum of ${p_{\mathrm{T}}}^2$, are considered to be products of a pile-up (parasitic) interaction and are removed from consideration.

### 4.1.1 Muon

Muons are reconstructed from tracks in the silicon tracker and measurements in the muon system. In this analysis we require the reconstruction of a track in the muon system is compatible with a silicon tracker track. Several selection criteria, are further applied to increase the purity of the muons and to suppress background contribution, e.g., from non-isolated muons from decays of pions etc. Lepton isolation is a measure of the hadronic activity around the lepton. It is defined as the ratio of the $p_{\mathrm{T}}$ sum of particles (leptons, charged hadrons, neutral hadrons and photons) within a defined $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ around the lepton exclude the contributions from the physics object itself and pileup events to the $p_T$ of the lepton. Muon is selected to satisfy the isolation requirement. Events with additional muons are removed from this analysis.

### 4.1.2 Electron

Electrons are reconstructed from a cluster in the electromagnetic calorimeter ECAL(with the gap between the barrel and endcap region ($1.44 < |\eta_{sc}| < 1.57$) being excluded [1]) and a single track pointing towards this cluster. To take into account the effect of bremsstrahlung due to interaction with the tracker material energy of photons spatially compatible with originating from the electron track are added using the Gaussian-sum filter algorithm [33]. As discussed in the muon selection, electron is also required to satisfy the isolation selection. Events with additional electrons are removed from this analysis.

The energy and resolution of leptons is calibrated using Z boson dilepton decays. The efficiencies of lepton identifications and selections are derived using tag-and-probe method [34] in $p_T$ and $\eta$ regions.

### 4.1.3 Jets

Jets are reconstructed from PF particle candidates, clustered using the anti-k$_T$algorithm [35, 36] with radius $\Delta R$ of 0.4. Jet momentum is determined as the vectorial sum of all PF candidate momenta in the jet. In this analysis jets are selected within the tracker acceptance of $|\eta| < 2.4$ with the minimum $p_T$ of 30 GeV. Isolated muons and electrons are removed from the jet collection if $\Delta R(\ell, jet) < 0.4$.

[1] $\eta_{sc}$ is the detector pseudo-rapidity defined with respect to the origin of the coordinate system at the center of CMS in contrast to the primary vertex that is used to describe the direction of a particle.

### 4.1.4 *b*-tagging

Due to the relatively long lifetime of b-hadrons, they typically travel a measurable distance before decaying. The high-resolution tracker can resolve displaced vertices. Secondary vertices and the possible presence of muon or electrons from semileptonic decays of b-hadrons are used to identify jets from hadronized b-quarks (b-jets). CMS uses various algorithms to exploit these features for robust identification of b-jets (b-tagging) [37]. In this thesis the Combined Secondary Vertex version 2 (CSVv2) algorithm is used.

The probability of correctly identifying jets as originating from b quarks (b-tagging efficiency), and the probability of misidentifying jets originating from light-flavour partons (u, d, s quarks or gluons) or a charm quark as a b-tagged jet (the light-flavour and charm mistag probabilities) are measured in control samples [38]. CSVv2 algorithm provides a discriminant between light-flavor and b-jets based on the combined information of secondary vertices and the impact parameter of tracks with respect to the primary vertex. Several working points based on the value of this discriminate are calibrated on data. In this analysis we choose a "medium" working point that corresponds to b-tagging efficiency of 63%, and charm and light-flavour mistag probabilities of approximately 12% and 2%, respectively. At least two jets in each event are required to be b-tagged using this working point.

### 4.1.5 Missing transverse energy

The missing transverse momentum $\vec{E}_T$ is calculated as the negative vector sum of the transverse momenta of all PF candidates in an event. Its magnitude is referred to as $E_T^{miss}$ (MET). We do not impose any requirements on $E_T^{miss}$ at the event selection stage. Instead, this variable is used to fully reconstruct the kinematics of the $t\bar{t}$ events.

## 4.2 $t\bar{t}$ event selection

Candidate signal events are selected to have an isolated muon or electron, and at least three jets.

To reduce the contribution from background processes and optimize the $t\bar{t}$ reconstruction additional, more stringent, requirements on the events are imposed as summarized in Table 4.1.

| Preselection | |
|---|---|
| exactly one $\mu(e)$ | $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ |
| veto additional $\mu(e)$ | $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ |
| number of jets | $n_{jets} \geq 3$ |
| jets | $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ |
| Selection during $t\bar{t}$ reconstruction | |
| $b_l$,$b_h$ | $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ |
| $b_l$,$b_h$ | CSVv2 medium working point |
| $j_W$ | $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ |
| $W$ boson | $M_T^{\ell,\nu} < 140\,\mathrm{GeV}$ |
| Additional selection if $n_{jets} = 3$ | |
| leading jet | $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ |
| likelihood | $-\ln\lambda_3 < 13$ |

**Table 4.1:** Event selection requirements. $b_{l(h)}$ is the b-jet on leptonic (hadronic) side of $t\bar{t}$ event, $j_W$ is the light jet assigned to the W boson decay. The W boson transverse mass is defined as $M_T^{\ell,\nu} = (E_T^\ell E_T^\nu - \vec{p}_T^{\,\ell} \cdot \vec{p}_T^{\,\nu})^{1/2}$, where $E_T^\ell$ and $\vec{p}_T^{\,\ell}$ refer to the transverse energy and transverse momentum for the lepton and $E_T^\nu$ and $\vec{p}_T^{\,\nu}$ refer to the transverse energy and transverse momentum for the neutrino as found by the neutrino solver. The likelihood $\lambda_3$ for the three jet category is defined in Sec .

In the three jets category, in order to lower the contribution from QCD multijet production, additional selection criteria are imposed: the event likelihood $-\ln(\lambda_3) < 13$ and leading jet $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. The likelihood selection rejects QCD events by 18% while it keeps 95% of $t\bar{t}$ events. The $p_{\mathrm{T}}$ selection rejects QCD events by 12% and it still keeps 98.5% of $t\bar{t}$ events. Furthermore, the likelihood and the $p_{\mathrm{T}}$ selection reject 31% of single top events and 40% of V+jets events.

# Chapter 5

# Data and simulated samples

## 5.1 Data

Events are taken from certified runs, for which a good performance and full functionality of the CMS detector is ensured. These runs correspond to an integrated luminosity 35.8 $fb^{-1}$. For the muon channel events are selected based on a logical OR between the triggers HLT_IsoMu24 and HLT_IsoTkMu24. The electron events are triggered with HLT_Ele27_WPTight_Gsf. The $p_T$ thresholds of 24 GeV for the muon and 27 GeV for the electron triggers are below the analysis selection requirement of 30 GeV. Further details about trigger efficiencies and parametrization can be found in [39].

## 5.2 Simulated samples

The Monte Carlo(MC) program POWHEG [40–43] (v2) is used to simulate $t\bar{t}$ events, which includes next-to-leading order (NLO) QCD matrix element calculation. It is followed with the parton shower simulation using PYTHIA [44, 45] (v8.205) (PYTHIA8) with the CUETP8M2T4 tune. All processes with up to two additional partons are calculated at NLO and combined with the PYTHIA8 parton shower simulation using the FxFx [46] algorithm. The default parametrization of the Parton Density Functions (PDF) used in all simulations is NNPDF30_nlo_as_0118 [47]. Top quark mass $m_t = 172.5$GeVis used. All Monte Carlo samples are listed in Table 5.1.

When compared to data, the simulation is normalized to an inclusive $t\bar{t}$ production cross section of $832^{+40}_{-46}$ pb [23]. This value is calculated with NNLO precision including the resummation of next-to-next-to-leading-logarithmic (NNLL) soft gluon terms. Its given uncertainty is due to the choice of hadronization/factorization scales and PDF.

In all simulated processes event weights are calculated that represent the uncertainty eigenvectors of the PDF. There are also event weights available that represent the changes of factorization and renormalization scales by a factor of two or one half. These additional weights allow for the calculation of systematic uncertainties due to the PDF and the scale choices. For additional uncertainty estimations we use POWHEG+PYTHIA8 simulated samples with top quark masses of 171.5 and 173.5 GeV, with initial and final state parton shower scales varied up and down by a factor of two. Simulation based on POWHEG combined with HERWIGPP [48] (v2.7.1) with the EE5C tune [49] is used for further comparison.

The main background processes are simulated using the same techniques. The MG5_aMC@NLO generator is used for the simulation of W boson production in association with jets, single top quark production (the dominant $t$-channel, and the smaller $tW$ and $s$-channel), and Drell–Yan (DY) production in association with jets. The POWHEG generator is used for the simulation of single top quark associated production with a W boson (tW) and PYTHIA8 is used for multijet production. In all cases the parton shower and the hadronization are described by PYTHIA8 with the CUETP8M1T4 tune [50]. The W boson and DY backgrounds are normalized to their NNLO cross sections [51]. The single top quark processes are normalized to NLO calculations [52, 53], and the multijet simulation is normalized to the LO calculation [45].

The detector response is always simulated using GEANT4 [54]. The same reconstruction algorithms that are applied to the data are used.

| RunIISummer16 | Evts.(weights) [$\times 10^6$] | $\sigma$[pb] |
|---|---|---|
| Signal | | |
| TT_TuneCUETP8M2T4_13TeV-powheg-pythia8 | 154.64(154.64) | 832 |
| TT_TuneCUETP8M2T4_13TeV-powheg-fsrdown-pythia8 | 59.1(59.1) | |
| TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8 | 59.2(59.2) | |
| TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8 | 59.0(59.0) | |
| TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8 | 58.84(58.84) | |
| TT_TuneCUETP8M2T4_mtop1715_13TeV-powheg-pythia8 | 19.58(19.58) | |
| TT_TuneCUETP8M2T4_mtop1735_13TeV-powheg-pythia8 | 19.42(19.42) | |
| TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8 | 58.34(58.34) | |
| TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8 | 58.95(58.95) | |
| TT_hdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8 | 58.16(58.16) | |
| TT_hdampUP_TuneCUETP8M2T4_13TeV-powheg-pythia8 | 58.86(58.86) | |
| TT_TuneCUETP8M2T4_erdON_13TeV-powheg-pythia8 | 59.88(59.88) | |
| TT_TuneEE5C_13TeV-powheg-herwigpp | 59.17(59.17) | |
| TTJets_TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8 | 43.56(15.07) | |
| Backgrounds | | |
| W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 45.17(45.17) | 11917.5 |
| W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 59.52(59.52) | 3850.4 |
| W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 58.07(58.07) | 1124 |
| W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 30.00(30.00) | 579.3 |
| DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 144.61(144.61) | 5765 |
| WW_TuneCUETP8M1_13TeV-pythia8 | 7.98(7.98) | 118.7 |
| WZ_TuneCUETP8M1_13TeV-pythia8 | 4.00(4.00) | 47.13 |
| ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 | 38.81(38.81) | 81 |
| ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 | 67.24(67.24) | 136 |
| ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1 | 5.43(5.43) | 19.3 |
| ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1 | 5.37(5.37) | 19.3 |
| ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1 | 1.0(0.62) | 3.4 |
| QCD_Pt-30to50_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 29.56(29.56) | |
| QCD_Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 19.81(19.81) | |
| QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 23.18(23.18) | |
| QCD_Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 19.98(19.98) | |
| QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 36.96(36.96) | |
| QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 49.0(49.0) | |
| QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 19.36(19.36) | |
| QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 19.74(19.74) | |
| QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 19.77(19.77) | |
| QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8 | 13.2(13.2) | |
| QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8 | 23.47(23.47) | |
| QCD_Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8 | 11.5(11.5) | |
| QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8 | 77.7(77.7) | |
| QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8 | 77.77(77.77) | |
| QCD_Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8 | 11.54(11.54) | |
| QCD_Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8 | 7.37(7.37) | |

**Table 5.1:** Simulation samples

## 5.3 Pile-up correction to simulated samples

The distribution of the number of pileup (PU) interactions in MC is different from that in data, which varies over time. This difference between MC and data is corrected using the pileup reweighting factors. These factors are defined as the ratios of the distribution of the number of interactions in data and MC. The distribution for data is derived by measuring the instantaneous luminosity of colliding bunches and multiplying by the total inelastic proton-proton cross section. Figure 5.1 shows the distribution of the mean number of interactions per crossings.

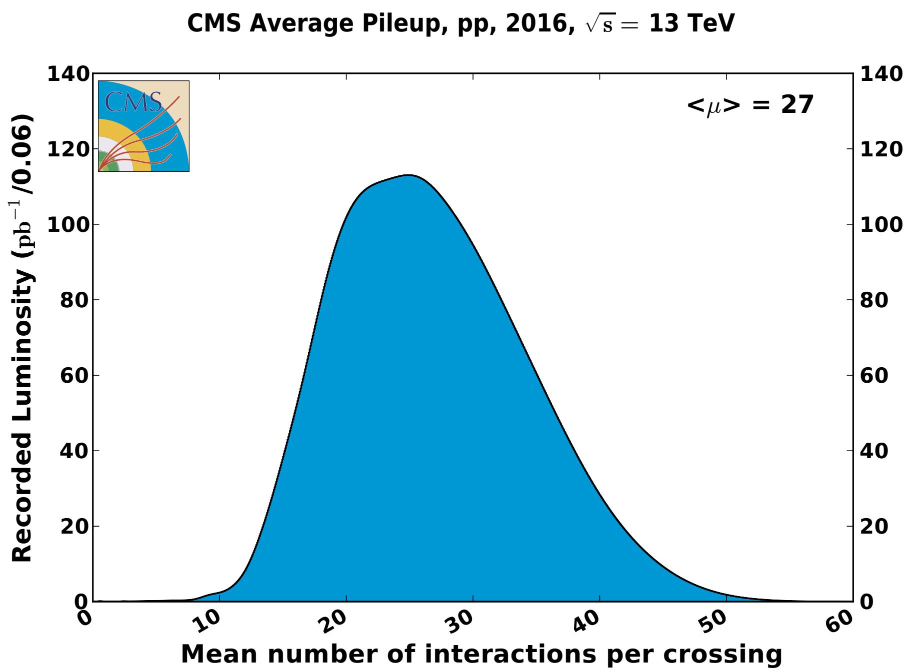


**Figure 5.1:** Mean number of interactions per bunch crossing for the 2016 proton-proton run at 13 TeV. The cross section is taken to be 80 mb.

# Chapter 6

# Reconstruction of the top quark-antiquark system

## 6.1 Reconstruction of the top quark-antiquark system

The goal of reconstructing $t\bar{t}$ events is to determine the full kinematics of top and antitop quarks from their decay products. For this it is important match the observed objects (leptons and jets) to the top and antitop quark decay products. We always assume that the two b-tagged jets with the highest CSV values are associated with the two b-quarks from $t\bar{t}$ decays. We test all possible assignments of jets to quarks from the $t\bar{t}$ decay and build a likelihood that each assignment is correct. For each event, the jet to quark assignment with the highest probability is selected.

The first step in building a likelihood is to reconstruct the neutrino four-momentum $p_\nu$ based on the measured $\vec{\not{E}}_T$. The Neutrino Solver algorithm [55], finds all possible solutions for the neutrino momentum using two mass constraints $(p_\nu + p_\ell)^2 = m_{\mathrm{W}}^2$ and $(p_\nu + p_\ell + p_{\mathrm{b}_\ell})^2 = m_{\mathrm{t}}^2$. Each equation describes an ellipsoid in the three-dimensional neutrino momentum space. The intersection of these two ellipsoids is usually an ellipse. We select $p_\nu$ as the point on the ellipse for which the distance $D_{\nu,\mathrm{min}}$ between the ellipse projection onto the transverse plane ($p_{\nu x}$,$p_{\nu y}$) and $\vec{\not{E}}_T$ is minimal. The algorithm leads to a unique solution for the longitudinal neutrino momentum and an improved resolution for the transverse component. The minimum distance $D_{\nu,\mathrm{min}}$ is also used to identify the correct b jet in the leptonic top decay, $\mathrm{b}_\ell$.

When the invariant mass of the lepton and $b_\ell$ candidate is above $m_t$, no solution can be found in the mass constraint and we continue with the next jet assignment.

### 6.1.1 Reconstruction of events with at least four jets

The likelihood to find correct jet to quark assignment for events with at least four reconstructed jets is constructed using the W boson mass and the top quark mass constraints on the hadronic side and the calculated $D_{\nu,\mathrm{min}}$ on the leptonic side. It is defined as $\lambda_4$ (index 4 refers to the requirement of at least four jets):

$$-\ln(\lambda_4) = -\ln(P_m(m_2, m_3)) - \ln(P_\nu(D_{\nu,\mathrm{min}})), \tag{6.1}$$

where $P_m$ is a two-dimensional probability distribution of the invariant masses of correctly reconstructed W bosons and top quarks. This probability is calculated based on the invariant mass of the two jets $m_2$ hypothesized to be the W boson decay products, and the invariant mass of the three jets $m_3$ hypothesized to be the decay products of the hadronically decaying top quark. The distributions for the correct jet assignments, normalized to unity, are shown in Fig. 6.1 (upper left). This part of the likelihood is sensitive to the correct reconstruction of the hadronically decaying top quark, modulo a permutation of the two jets from the W boson, but none of the measured kinematic variables will be affected by this ambiguity.

The probability $P_\nu$ describes the distribution of $D_{\nu,\mathrm{min}}$ for a correctly selected $b_\ell$. In Fig. 6.1 upper right, the normalized distributions of $D_{\nu,\mathrm{min}}$ for $b_\ell$ and for other jets are shown. On average, the distance $D_{\nu,\mathrm{min}}$ for correctly selected $b_\ell$ is smaller and has a lower tail compared to the distance obtained for the wrongly assigned jets. Jet assignments with values of $D_{\nu,\mathrm{min}} > 150\,\mathrm{GeV}$are rejected since they are very unlikely to originate from a correct $b_\ell$ association. This part of the likelihood is sensitive to the correct reconstruction of the leptonically decaying top quark.

The likelihood $\lambda_4$ is plotted in lower plot of Fig. 6.1 for the best permutation. Here, the signal simulation is divided into the following categories: correctly reconstructed $t\bar{t}$ systems ($t\bar{t}$ right reco), events where all decay products are available, but the algorithm failed to identify the correct jet assignments ($t\bar{t}$ wrong reco), $\ell$+jets tt̄ events where at least one decay product is missing (tt̄ not reconstructible), and tt̄ events from dileptonic or fully hadronic decays (tt̄ background).

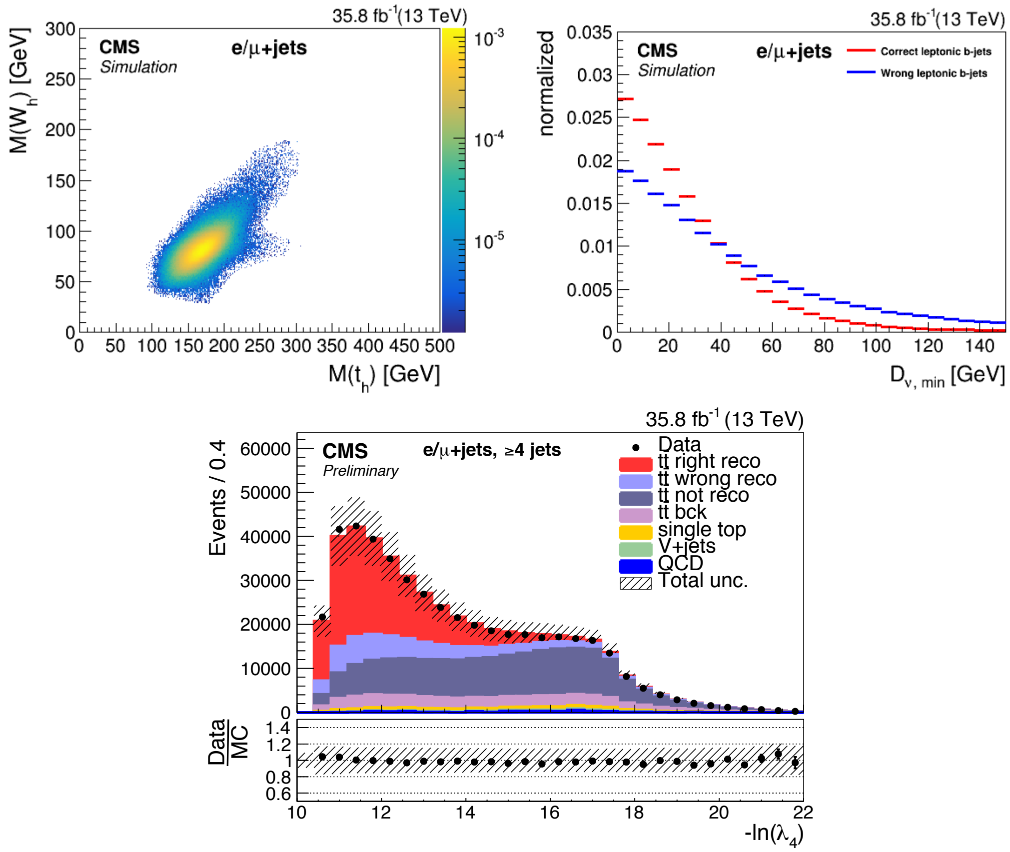


**Figure 6.1:** At least four jets reconstruction. Upper left: normalized two-dimensional mass distribution of the correct reconstructed hadronically decaying W bosons $M(\mathrm{W})$ and the correct reconstructed top quarks $M(\mathrm{t_h})$. Upper right: normalized distributions of the distance $D_{\nu,\mathrm{min}}$ for correctly and wrongly selected b jets from the leptonically decaying top quarks. Lower plot: Distribution of the negative log-likelihood for the selected best jet assignment for the tt̄ MC and the backgrounds. The possible categories from the tt̄ reconstruction algorithm are explained in the text.

### 6.1.2 Reconstruction of events with one missing jet

The most sensitive to top Yukawa coupling region of the phase space is the threshold of tt̄ production. At the same time the efficiency of selecting tt̄ events in this region is rather low since one or more quarks from tt̄ decay is likely to have $p_\mathrm{T}$ or $\eta$ below the selection thresholds resulting in a missing jet. To mitigate this effect we developed an algorithm of $t\bar{t}$ reconstruction with one missing jet. The idea is inspired by [56].

In 93% of the 3-jet events the missing jet is from $W$ boson decay. Starting with events that have two medium b-tagged jets and one light quark jet, and assuming that the two highest b-tagged jets are associated with b-quarks from $t\bar{t}$ decays (94% of the time the true b-quarks from tops are the two highest b-tagged jets), the only ambiguity is in the assignment of the b-tagged jets: which one originates from the hadronic top decay and which one from the leptonic top decay. To sort out this ambiguity the algorithm proceeds in the following way:

- We run the Neutrino Solver using each of the b-jets as a leptonic b-jet candidate to calculate the corresponding minimum distance $D_{\nu,\mathrm{min}}$.

- If one of the b-jet candidates does not yield a solution (48% of the overall cases), that b-jet is considered to be originating from the hadronic top decay and the other b-jet which has a correct neutrino solution is considered to be coming from the leptonic top decay (90% true).

- If instead both b-jets candidates yield solutions for neutrino momentum (52% of the overall cases), a combined likelihood $\lambda_3$ (index 3 refers to three jets), is constructed using the minimum distance $D_{\nu,\mathrm{min}}$ and the invariant mass $m_{t_h}$ of the two jets hypothesized to belong to the hadronic top decay:

$$-\ln(\lambda_3) = -\ln(P_{m_{t_h}}) - \ln(P_\nu(D_{\nu,\mathrm{min}})), \tag{6.2}$$

  where $P_\nu(D_{\nu,\mathrm{min}})$, shown in Fig. 6.2 upper left, is the probability to identify $b_\ell$; $P_{m_{t_h}}$, shown in Fig. 6.2 upper right, is the probability of the invariant mass of the hypothesized $b_h$ and the jet from W boson decay. The distribution in $\lambda_3$ is shown in the lower plot of Fig. 6.2. We choose the b-jet with the lowest $\lambda_3$ as the leptonic b-jet $b_\ell$ (72% true), and the other b-jet as the hadronic b-jet $b_h$.

When the top candidate is off-shell, no solution can be found to satisfy the mass constraints and we continue with the next b-jet candidate. Jets assignment with values of $D_{\nu,\min} > 150\,\text{GeV}$ and hypothesized invariant mass $m_{t_h} > 500\,\text{GeV}$ are rejected since they are very unlikely to originate from a correct b-jets association. Overall, the described algorithm identifies the correct b-jets in 80% of three jet events, where all three jets are coming from $\text{t}\bar{\text{t}}$ decay.

Leptonic top decays are fully reconstructed, regardless of whether we have three or four jets. The hadronic-top candidate in the missing jet category is approximated by the system of the b-jet identified to be associated with the hadronic top decay and the only light-quark jet. Figure 6.3 shows the resolutions of the invariant mass of the $\text{t}\bar{\text{t}}$ system and the difference in rapidity for three jet events compared to those with at least four jets. It is worth noting that the widths of these distributions are more important than the central value, since the respective templates are derived separately for different jet multiplicities. The resolution shows that the three jet reconstruction is competitive with the one achieved in the four jet category.

To summarize, the newly developed three jets reconstruction allows to increase the statistics in the sensitive low $M_{\text{t}\bar{\text{t}}}$ region. As will be shown in Section 9.2, the addition of three jets events also helps reducing the systematic uncertainty due to sources that cause migration between jet multiplicity bins, e.g. jet energy scale variation and hadronization model.

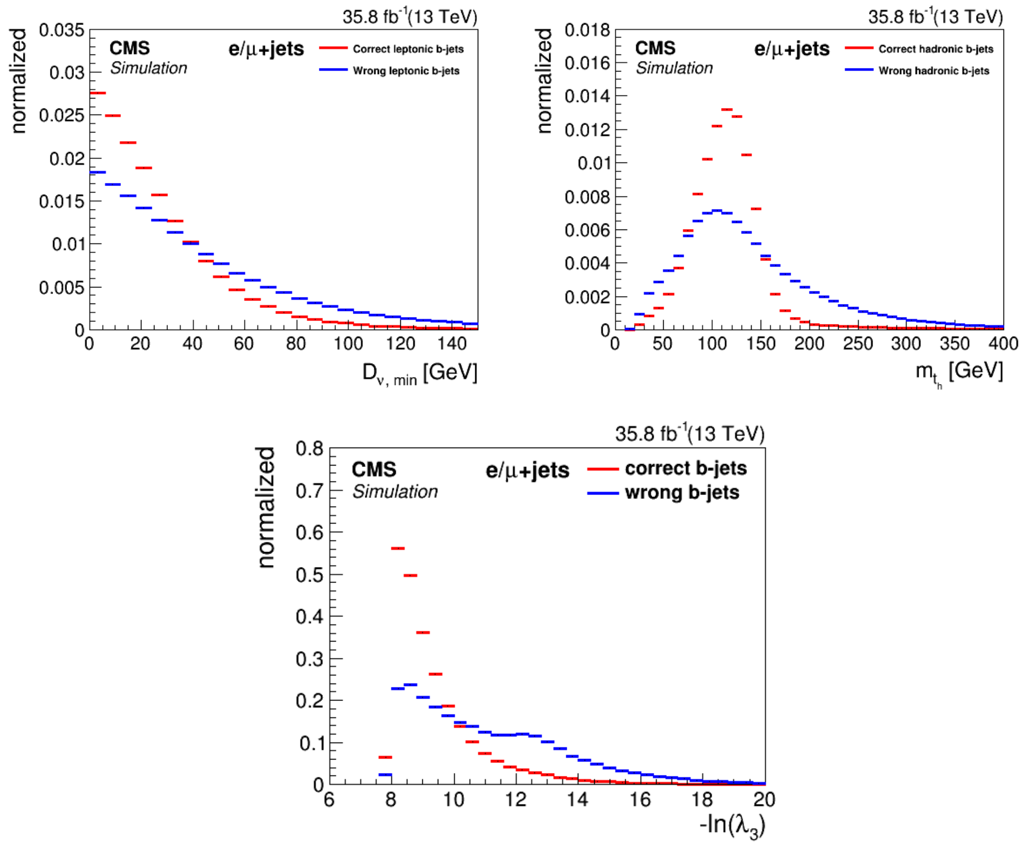


**Figure 6.2:** Three jets reconstruction. Upper left: normalized distributions of the distance $D_{\nu,\min}$ for correctly and wrongly selected $b_\ell$ candidates. Upper right: normalized mass distribution of the correctly and wrongly selected $b_h$ and the jet from W boson decay. Lower plot: Distribution of the negative combined log-likelihood.

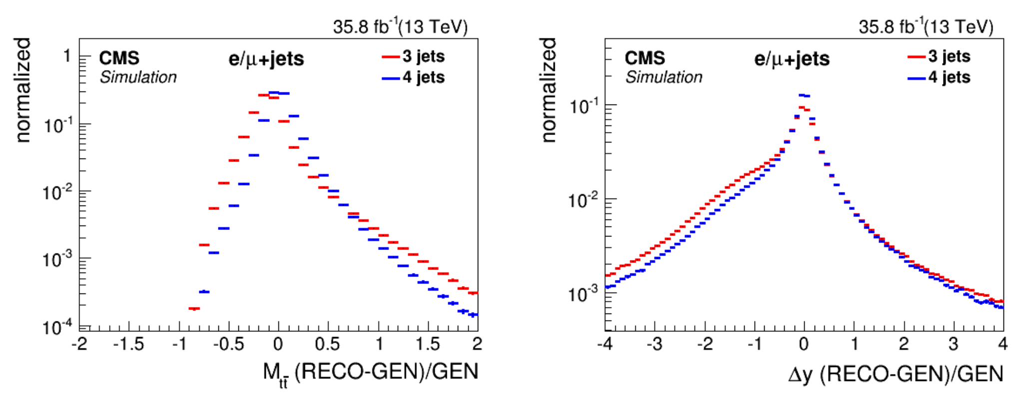


**Figure 6.3:** Resolution comparison between three jets reconstruction and at least four jets reconstruction.

# Chapter 7

# Background estimation

Backgrounds in this analysis arise from QCD multijet production, single top quark production, and vector boson production in association with jets (V+jets). The expected number of events from $WW$ and $WZ$ production as shown in Fig. 7.2 and Fig. 7.3 is negligible and we ignore it in the signal region.

The contribution from single top is estimated from the MG5_aMC@NLO and POWHEG simulated samples for $t$-channel and $Wt$-production, respectively. The MC statistical uncertainty and an inclusive 15% uncertainty (estimated by the latest LHCTOPWG combined cross section both in $t$-channel and $Wt$-production [57]) in the single top cross section are applied. The variations from all the other sources of systematics such as scale factors are included as well. The contribution from V+jets is estimated from the simulated samples and normalized according to the cross sections given in Table 5.1. The systematic uncertainty of the normalization of V+jets is conservatively set at 30%, based on the uncertainty of the $W$+heavy flavor production cross section [58].

Due to the low statistics of QCD simulated events after the selection, we derive shape templates in the $M_{t\bar{t}}$ and $\Delta y_t$ from a control region in order to get smooth distributions. One of the common control region for QCD estimation is in the region of inverted lepton isolation. Another control region is the low CSV region, enriched with V+jets and QCD. The CSV control region follows the same event selections and is defined by requiring that the maximum CSV value of jets in each event is less than 0.6.

The lepton $p_T$ in these two control regions are shown, shown in Fig. 7.1 top plots. The kinematics is observed to be softer in the region of inverted isolation. Shapes in both control regions are compared to the QCD simulation in the signal region, and shown in Fig. 7.1 bottom plot. It indicates that the CSV control region is in better agreement with the kinematics in signal region. For this reason, we use the CSV data control region to derive the shape of QCD.

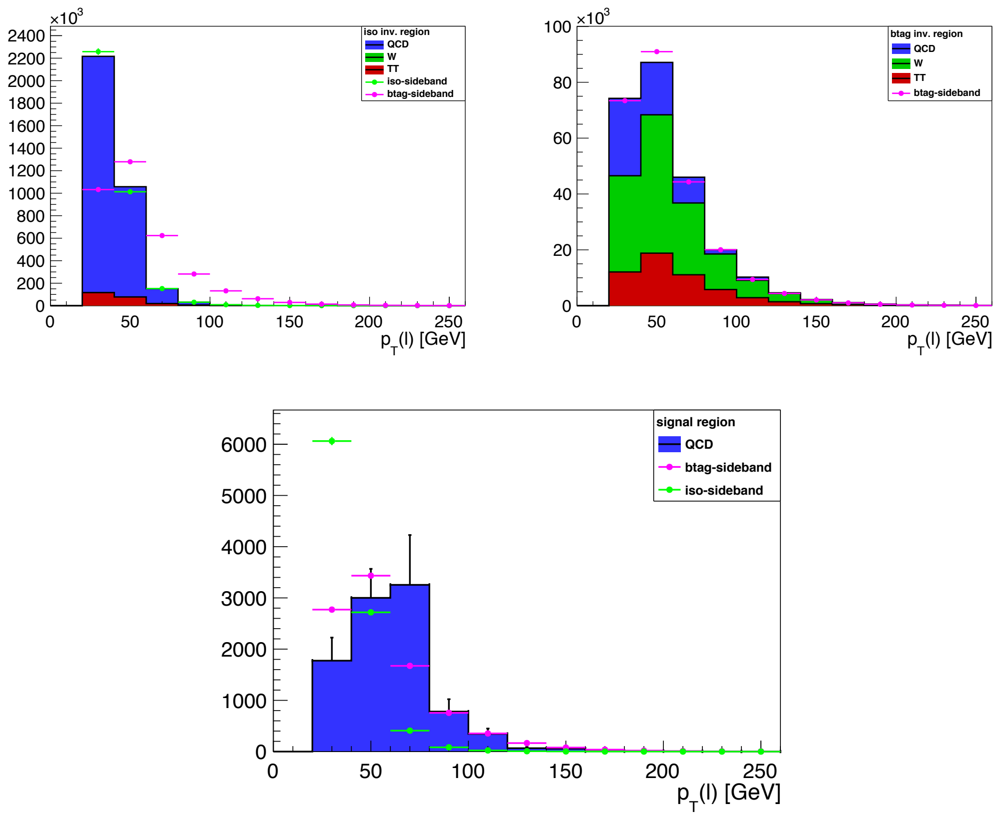


**Figure 7.1:** Top: The lepton $p_T$ distribution in the isolation (left) and the CSV (right) control regions. It shows good agreement between data and MC. In the left plot, the data distribution in the CSV control region is shown together with the isolation control region distributions for comparison purpose. Bottom: The QCD simulation for lepton $p_T$ distribution in the signal region compared to the data distributions in both control regions. The CSV control region shows a better agrement.

The distribution of the CSV b-tagging discriminant of the highest tagged jet is shown in Fig. 7.2, and it shows overall good agreement between data and MC. We further compare the $M_{t\bar{t}}$ and $\Delta y_t$ distributions in this sideband region in Fig. 7.3, which shows consistency of the data and MC within statistical uncertainties.

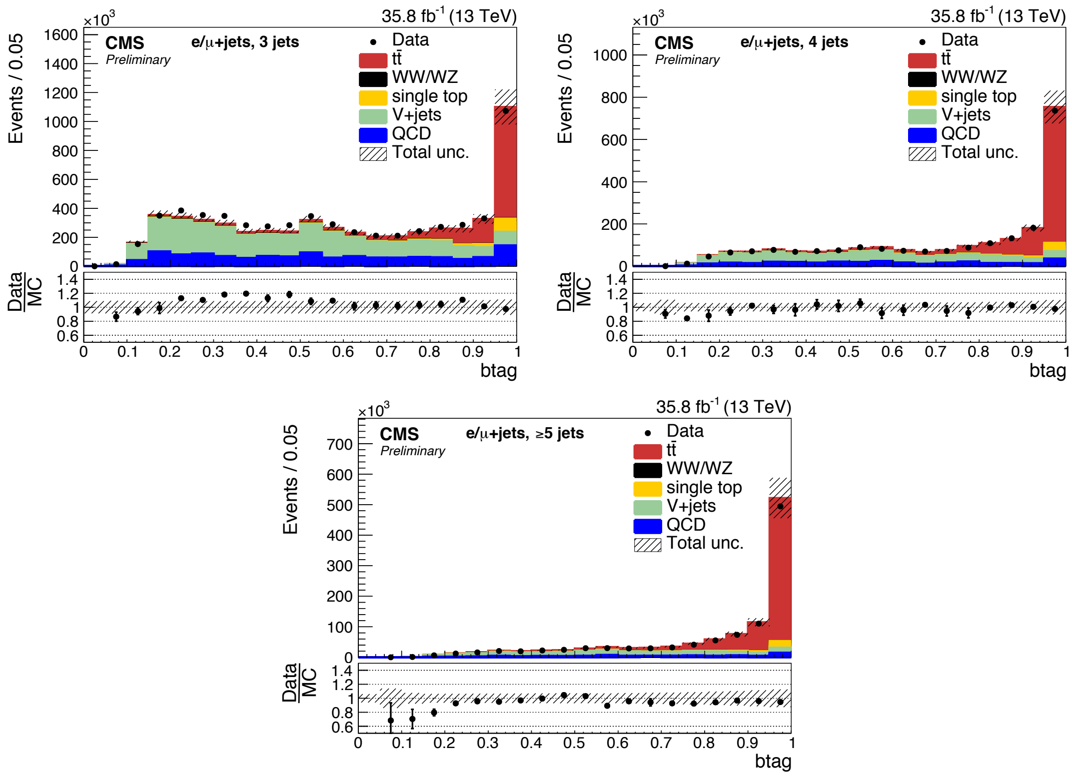


**Figure 7.2:** The CSV b-tagging discriminant distribution for the highest *b*-tagged jet in events with three, four, and five or more jets. The CSV control region is defined by CSV<0.6. This plot illustrates the control region is dominated by V+jets and QCD (92%).

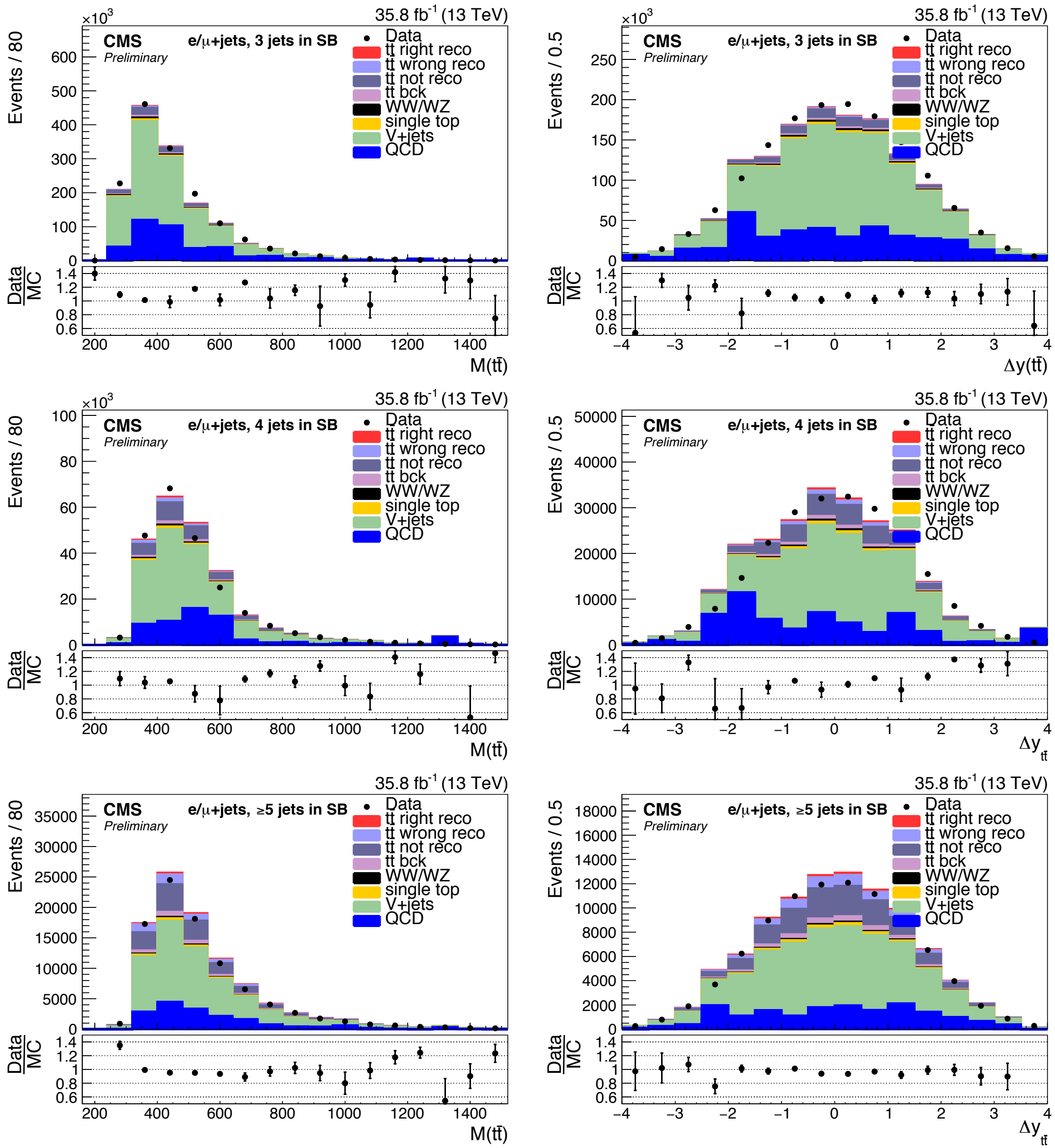


**Figure 7.3:** $M_{t\bar{t}}$ and $\Delta y_t$ distributions for events with three, four, and five or more jets in the control region (CSV<0.6).

We take the shape distributions for $M_{t\bar{t}}$ and $\Delta y_t$ from the data in the CSV sideband (minus the residual single top, $t\bar{t}$ and $WW$ and $WZ$ contributions), and normalize it as follows:

$$N^{SR}_{QCD} = N^{CR}_{resDATA} \times \frac{N^{SR}_{QCDMC}}{N^{CR}_{QCDMC}}, \tag{7.1}$$

where SR is the signal region, CR is the CSV sideband region, $N_{resDATA}$ is the residual data (data minus the non-QCD components), and $N_{QCDMC}$ is the MC estimate of the QCD background. The normalization uncertainty of the QCD background is 30%.

The QCD shape uncertainty due to the b-tagging reversal is evaluated in the non-isolated data. We consider the difference between tagged and anti-tagged events in this sample. Figure 7.4 demonstrates the sideband regions used for the QCD shape evaluation. Region I indicates the signal region. Region II is where we derived the QCD shape. Region III and IV are regions with inverted lepton isolation. For region III and IV, the single lepton trigger requirement is removed. The difference between III and IV is taken as the QCD shape uncertainty.

Figure 7.5 and Figure 7.6 show the shape comparisons in $M_{t\bar{t}}$ and $\Delta y_t$ for these four regions. The plots confirm that the shape derived from region II is closer to the QCD simulation in the signal region. Also, the shapes derived from region III and IV are closer to each other but more different from that in the region II. This means the shape does not change much by going from region III to IV, but it changes more by going from region III to region II, i.e., it has a smaller change by reversing the b-tagging than inverting the isolation selections.

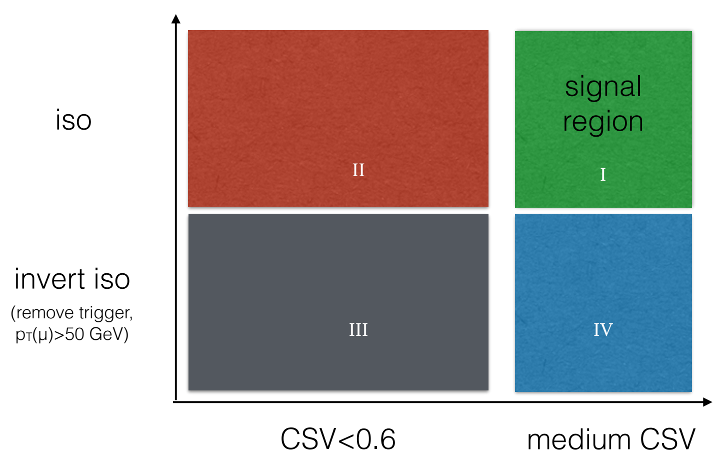


**Figure 7.4:** The QCD shape is derived in the region of low CSV (data control region II), and the shape uncertainty is estimated from the difference between region III and IV.

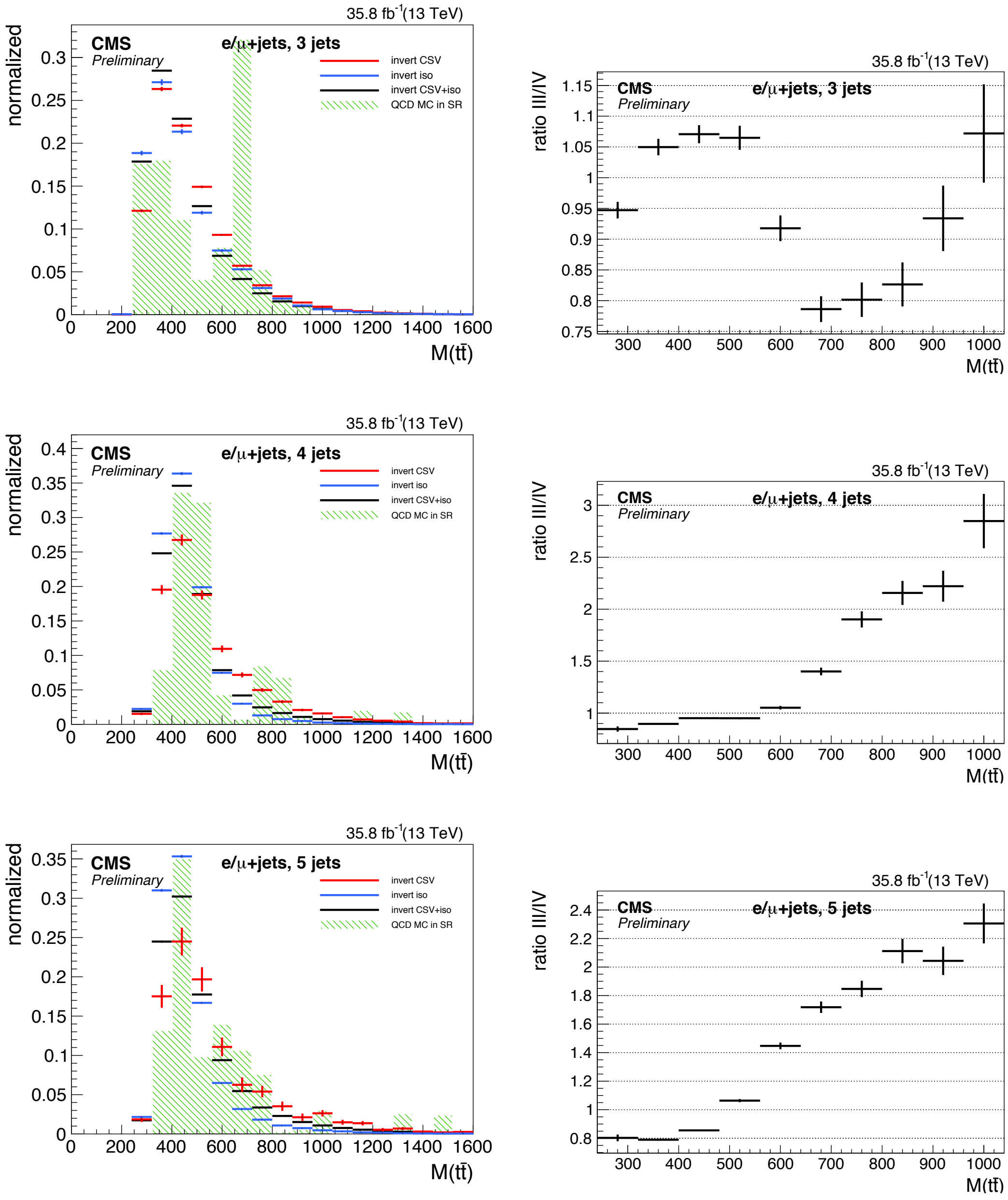


**Figure 7.5:** Left: Comparison of the QCD multijet background shape in $M_{t\bar{t}}$ derived from different data control regions (red: region II, black: region III, and blue: region IV) and MC in the signal region (green block) for events with three, four, and five or more jets. The lepton $p_T$ in region I and II are further required to be larger than 50 GeV in order to get a consistent selection with inverted isolation samples. Right: The ratio between region III and region IV, which is used as QCD shape uncertainty.

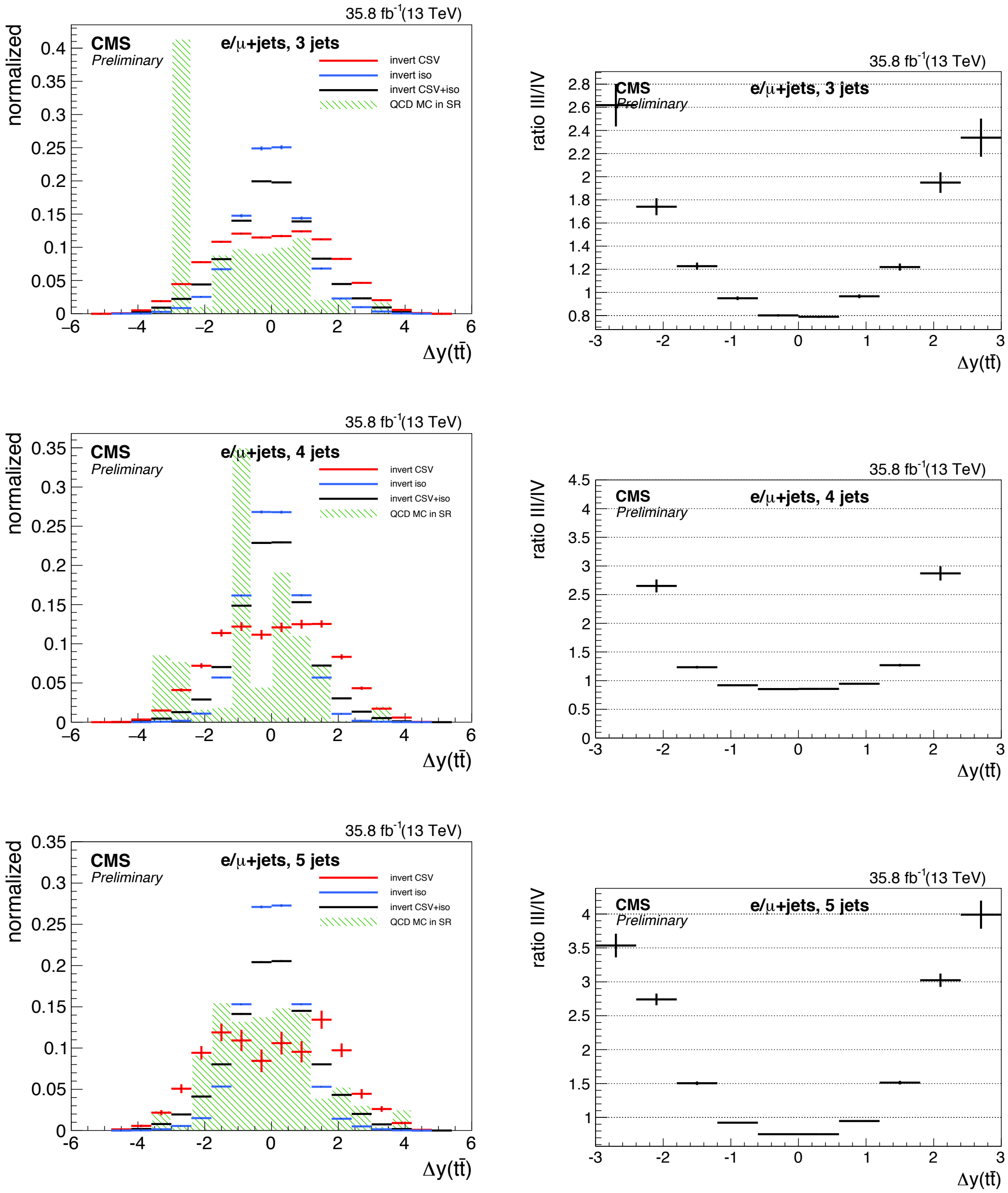


**Figure 7.6:** Left: Comparison of the QCD multijet background shape in $\Delta y_t$ derived from different data control regions (red : region II, black: region III, and blue: region IV) and MC in the signal region (green block) for events with three, four, and five or more jets. The lepton $p_{\mathrm{T}}$ in region I and II are further required to be larger than 50 GeV in order to get a consistent selection with inverted isolation samples. Right: The ratio between region III and region IV, which is used as QCD shape uncertainty.

# Chapter 8

# Event yields and control plots

The efficiency of the reconstruction algorithm is defined as the probability that the most likely assignment, as identified by the likelihood $\lambda$, is the correct one, given that all decay products from the $t\bar{t}$ decay are reconstructed and selected. These efficiencies as a function of the jet multiplicity are shown in Fig. 8.1 (right). Since the number of possible jet-to-quark assignments increases drastically with the number of jets, it is more likely to select a wrong assignment if there are additional jets. Thus, low number of jets has a higher reconstruction efficiency and it also means a higher fraction of correctly reconstructed $t\bar{t}$ events, which is the most sensitive category to Yukawa coupling. This is also one of advantages of reconstructing three-jets events. However, all the $t\bar{t}$ components carry the signature of Yukawa coupling in $t\bar{t}$ production, so all of the categories are considered as signal in the final likelihood. Table 8.1 shows the event yields with statistical uncertainties after event selection and $t\bar{t}$ reconstruction.

In Fig. 8.2 to Fig. 8.4, we display kinematic variables: MET, lepton $\eta$, $p_{\mathrm{T}}$ and absolute rapidity $y$ for hadronically and leptonically decaying top and $t\bar{t}$ system, reconstructed by the algorithm for three, four, five and more jets events. The V+jets and QCD background distributions in these plots come from a data sideband described in Sec. 8. In general, good agreement between data and prediction is observed.

| Source | 3 jets | 4 jets | ≥5 jets |
|---|---|---|---|
| $t\bar{t}$ right | 130523.1 ± 153.8 | 92895.1 ± 129.5 | 71642.6 ± 114.1 |
| $t\bar{t}$ wrong | 29297.9 ± 73.1 | 17355.6 ± 56.6 | 43073.0 ± 89.3 |
| $t\bar{t}$ semi | 50695.4 ± 96.2 | 88762.8 ± 127.1 | 80960.1 ± 122.1 |
| $t\bar{t}$ background | 53464.6 ± 99.0 | 26084.5 ± 69.1 | 25043.6 ± 68.1 |
| single t | 17847.8 ± 40.0 | 6922.3 ± 27.1 | 6294.3 ± 25.9 |
| V+jets | 8989.7 ± 100.3 | 2823.7 ± 51.6 | 2477.6 ± 48.9 |
| QCD multijet | 19834.8 ± 6247.0 | 2100.4 ± 603.2 | 1082.9 ± 209.6 |
| MC sum | 310653.4 ± 253.8 | 236944.5 ± 211.9 | 230574.0 ± 211.0 |
| Data | 308932.0 ± 555.8 | 237491.0 ± 487.3 | 226788.0 ± 476.2 |

**Table 8.1:** Expected and observed yields with statistical uncertainties after event selection. Events are categorized in the $t\bar{t}$ MC as: correctly identified $t\bar{t}$ systems ($t\bar{t}$ right); events where all decay products are available, but the $t\bar{t}$ reconstruction algorithm does not identify them as a correct $t\bar{t}$ permutation ($t\bar{t}$ wrong); events where at least one decay product is missing ($t\bar{t}$ not reco); events arising from the dileptonic or fully hadronic $t\bar{t}$ channels ($t\bar{t}$ background).

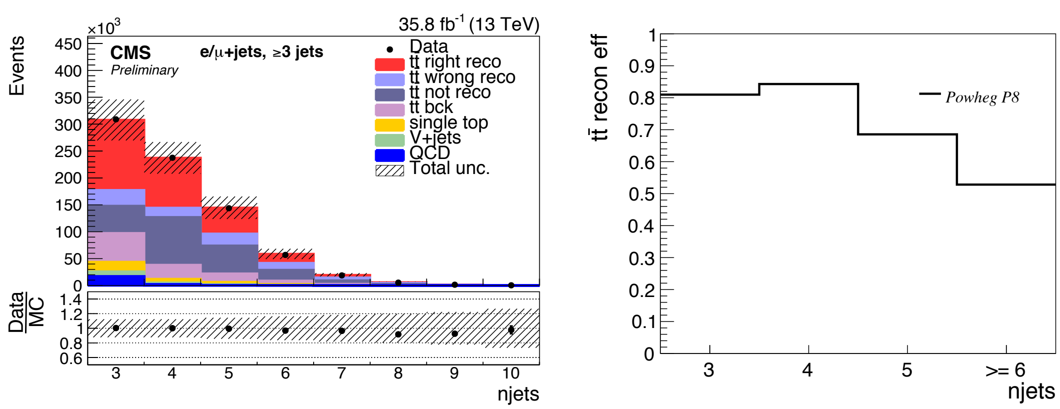


**Figure 8.1:** Left: Number of reconstructed $t\bar{t}$ events as a function of the number of jets. Right: Reconstruction efficiency of the $t\bar{t}$ system as a function of number of jets calculated based on the simulations with POWHEG+PYTHIA8 (P8).

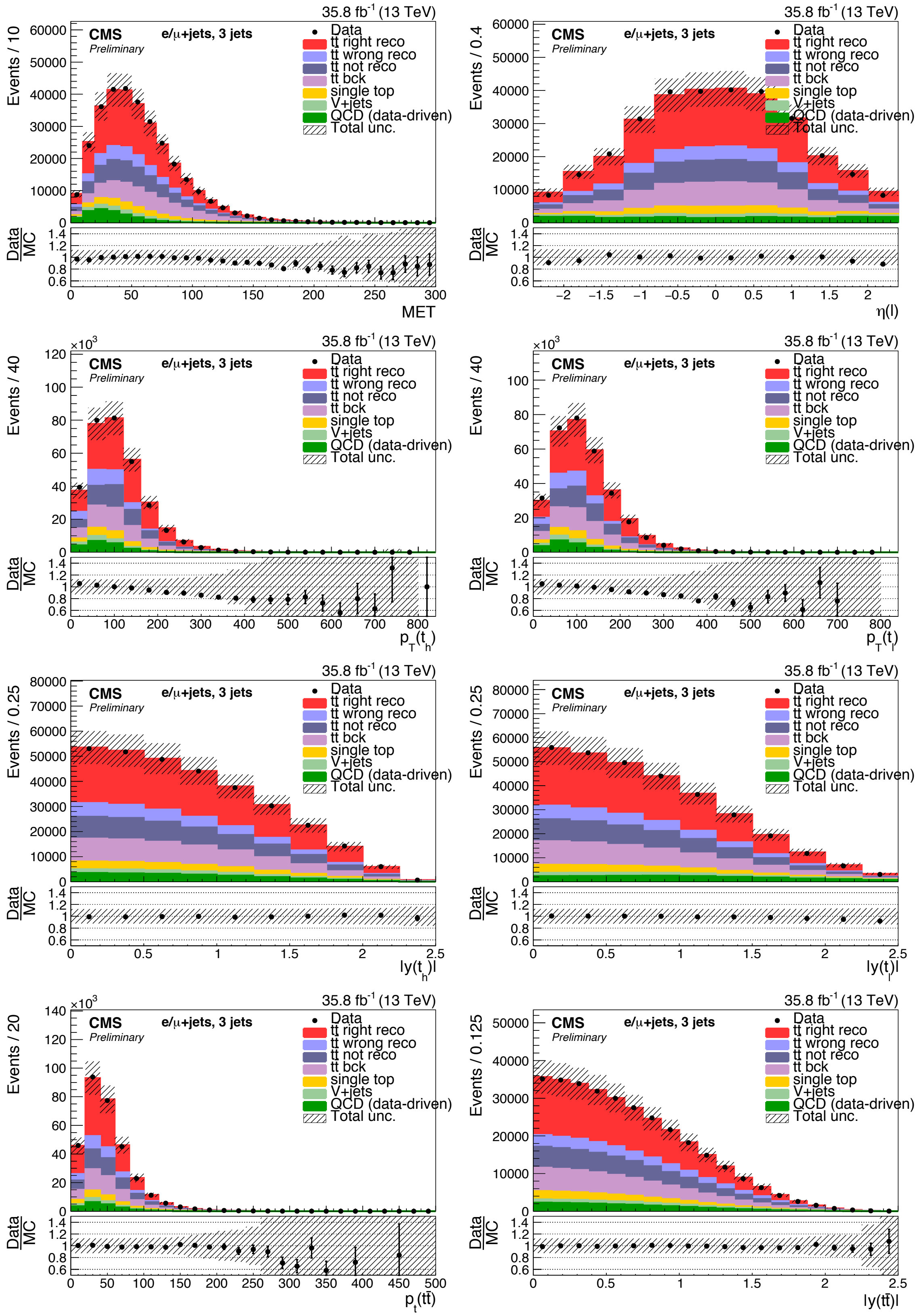


**Figure 8.2:** Kinematic distribution in the $t\bar{t}$ system: MET, lepton rapidity, $p_T$ and absolute rapidity $y$ of hadronic and leptonic top and $t\bar{t}$ system, of the events that pass our selection criteria in three jets $t\bar{t}$ events.

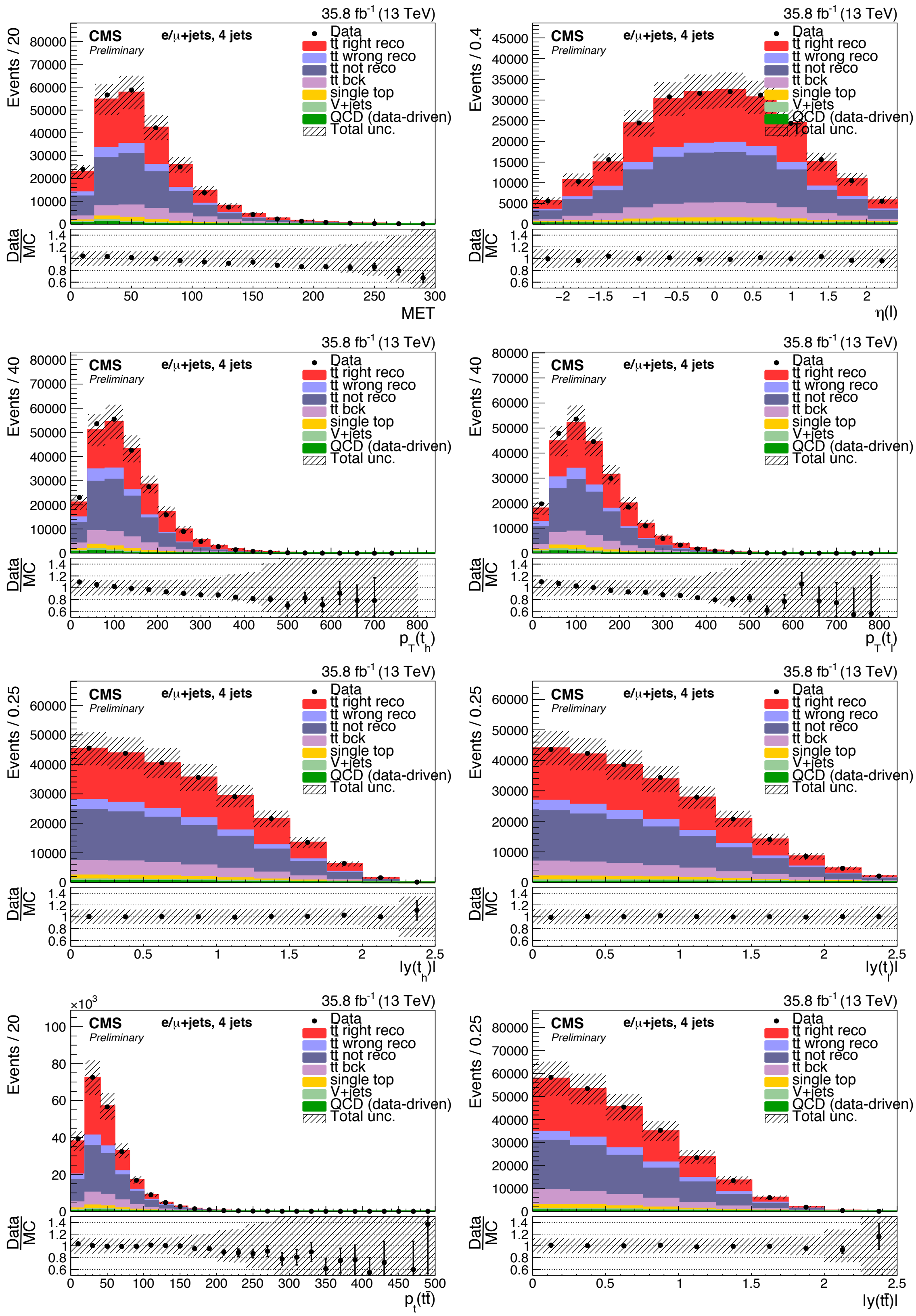


**Figure 8.3:** Kinematic distribution in the $t\bar{t}$ system: MET, lepton rapidity, $p_T$ and absolute rapidity $y$ of hadronic and leptonic top and $t\bar{t}$ system, of the events that pass our selection criteria in four jets $t\bar{t}$ events.

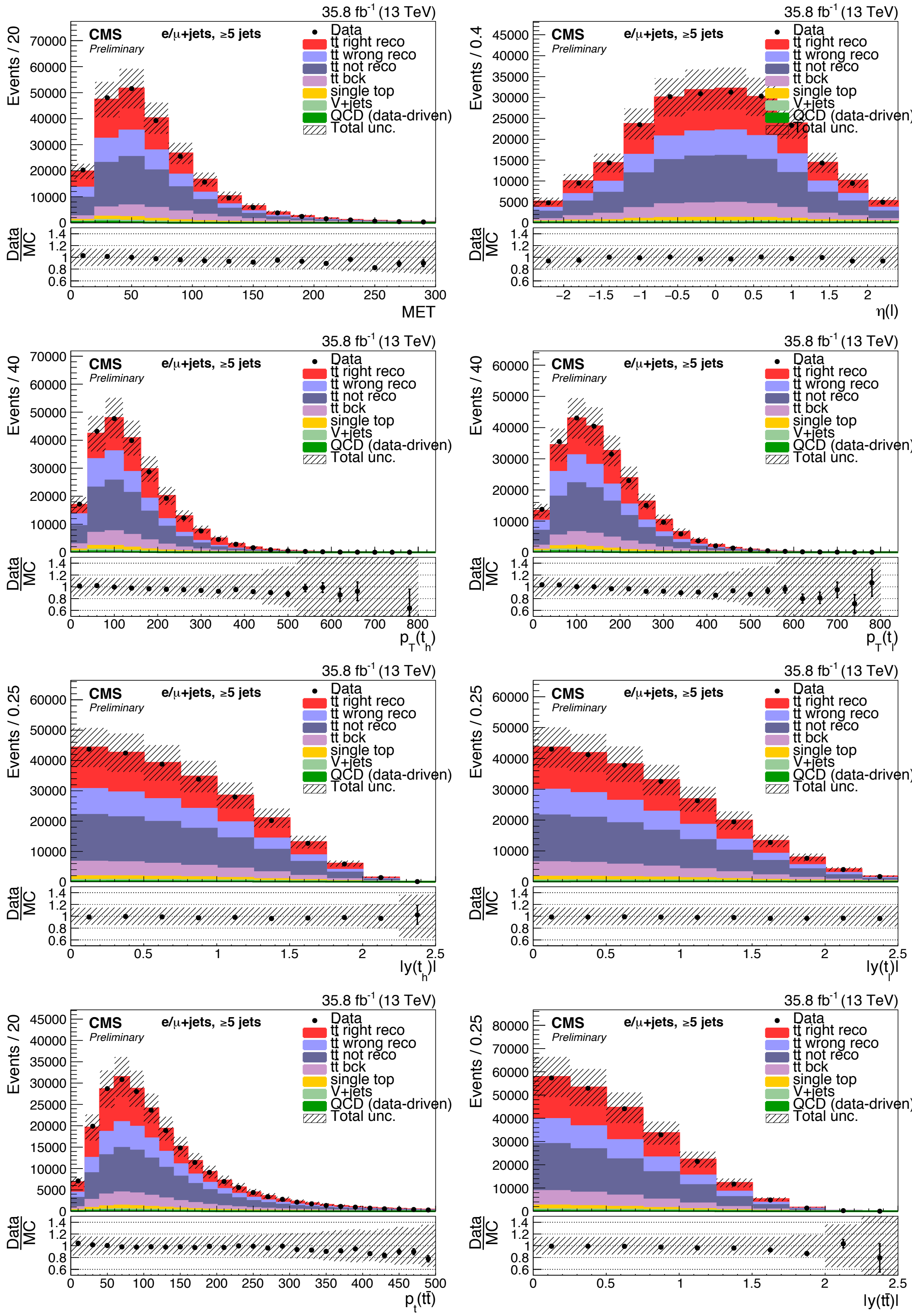


**Figure 8.4:** Kinematic distribution in the $t\bar{t}$ system: MET, lepton rapidity, $p_T$ and absolute rapidity $y$ of hadronic and leptonic top and $t\bar{t}$ system, of the events that pass our selection criteria in five or more jets $t\bar{t}$ events.

# Chapter 9

# Statistical Treatment

## 9.1 Likelihood model

We determine the most probable value of the $Y_t$ and its uncertainty using the likelihood model. It is constructed as a product of Poisson likelihoods for the observation in each ($M_{t\bar{t}}$, $|\Delta y_t|$) bin, as:

$$\mathcal{L} = \prod_{\text{bin } \in (M_{t\bar{t}}, |\Delta y_t|)} \mathcal{L}_{\text{bin}} = \prod_{\text{bin}} \text{Pois}(n_{\text{obs}}^{\text{bin}} | s^{\text{bin}}(\theta) \times R^{\text{bin}}(Y_t) + b^{\text{bin}}(\theta)) \times \rho(\theta|\tilde{\theta}) \quad (9.1)$$

where $\theta$ represents the full suite of nuisance parameters indicating our degree of belief on what the true values might be, $s^{\text{bin}}$ is the $t\bar{t}$ prediction from the POWHEG simulation in a given ($M_{t\bar{t}}$, $|\Delta y_{t\bar{t}}|$) bin, $b^{\text{bin}}$ is the prediction from simulation for various backgrounds, and $R^{\text{bin}}(Y_t) = s^{\text{bin}}(Y_t)/s^{\text{bin}}$ encodes the effect of different $Y_t$ coupling scenarios, formulated as quadratic functions of $Y_t$ in each bin. $R^{\text{bin}}(Y_t)$ provides electrweak corrections to the predictions on $t\bar{t}$, depending on the value of $Y_t$ assumed.

Finally, $\rho(\theta|\tilde{\theta})$ denotes constraint terms for each nuisance parameters, where $\tilde{\theta}$ is the default value of the nuisance parameter on $\theta$. Different sources of systematics are explained in detail in Section 9.2.

To extract $Y_t$ from data, a simultaneous likelihood fit is performed in all bins, defined by $M_{t\bar{t}}$ and $|\Delta y_{t\bar{t}}|$. The binning is chosen to capture different bahavior of electroweak corrections quantitatively. Figure 9.6 shows $R^{\text{bin}}(Y_t)$ and its quadratic dependence on $Y_t$. Three bins are defined in $|\Delta y_{t\bar{t}}|$: 0.0-0.6, 0.6-1.2, and 1.2 or above. Within each $|\Delta y_{t\bar{t}}|$ bin, each $M_{t\bar{t}}$ bin is formed by requiring at least 10000 simulated events. With this criterion, 21, 17 and 17 bins are defined for the three-jet, four-jet and five-or-more jet event category respectively.

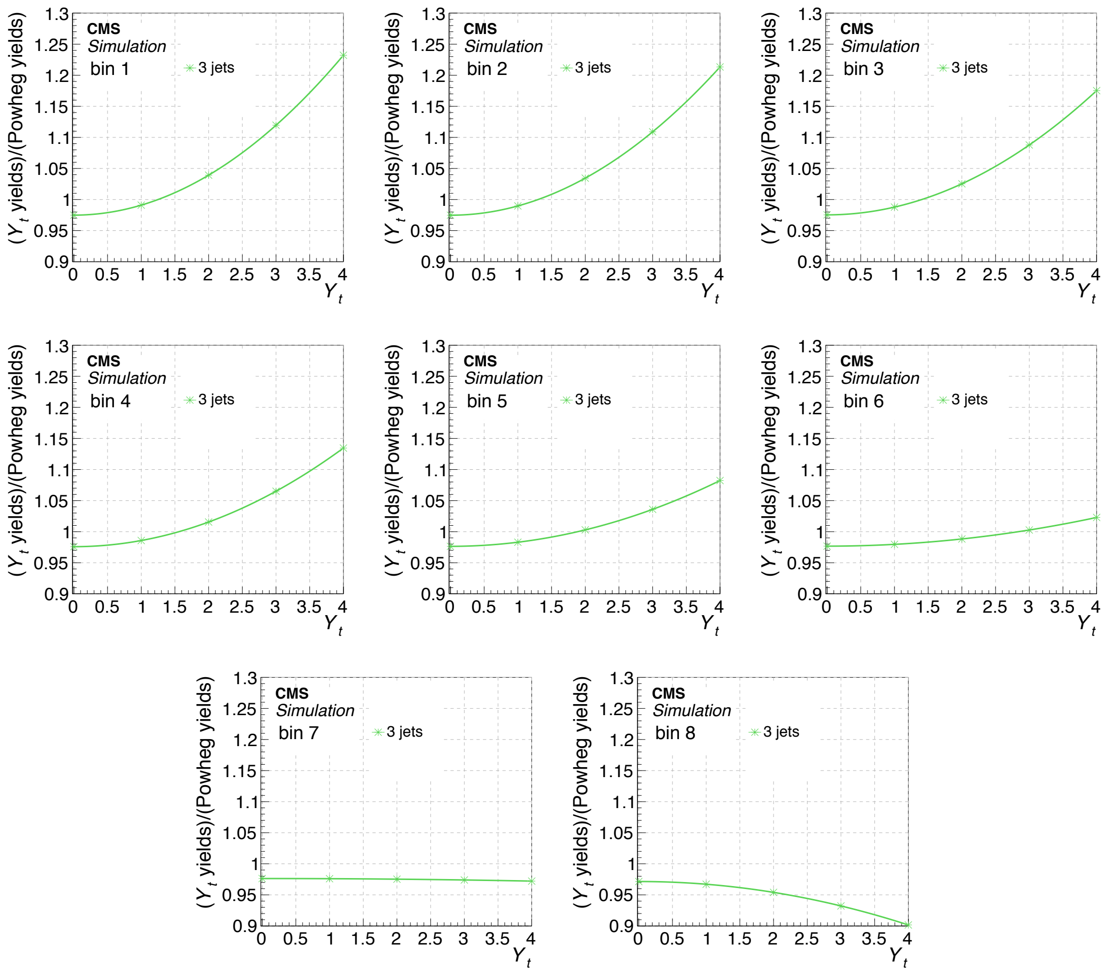


**Figure 9.1:** This plot shows $R^{\text{bin}}(Y_t)$ in their corresponding $|\Delta y_t| < 0.6$ bins along with $M_{t\bar{t}}$ in 3 jets. We apply a quadratic fit, in each bin, on for $Y_t$ couplings equal to 0, 1, 2, 3, 4, and 5. The plot shows the event yield with respect to the nominal that can be expected in a specific $M_{t\bar{t}}$ and $|\Delta y_t|$ region for each $Y_t$ scenario.

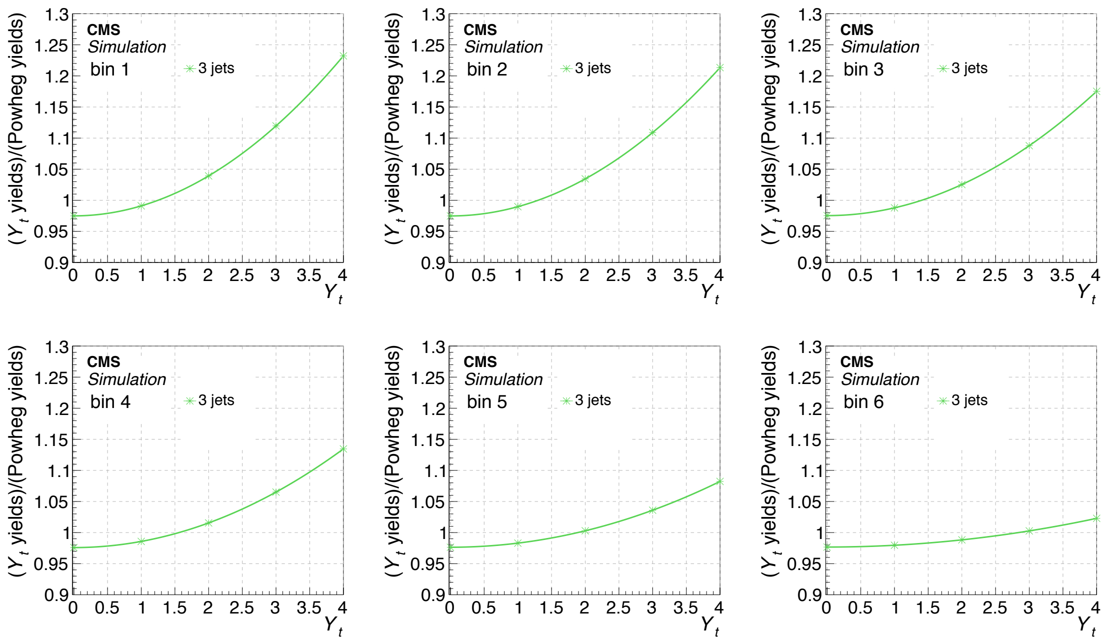


**Figure 9.2:** This plot shows $R^{\text{bin}}(Y_t)$ in their corresponding $|\Delta y_t| < 1.2$ bins along with $M_{t\bar{t}}$ in 3 jets. We apply a quadratic fit, in each bin, on for $Y_t$ couplings equal to 0, 1, 2, 3, 4, and 5. The plot shows the event yield with respect to the nominal that can be expected in a specific $M_{t\bar{t}}$ and $|\Delta y_t|$ region for each $Y_t$ scenario.

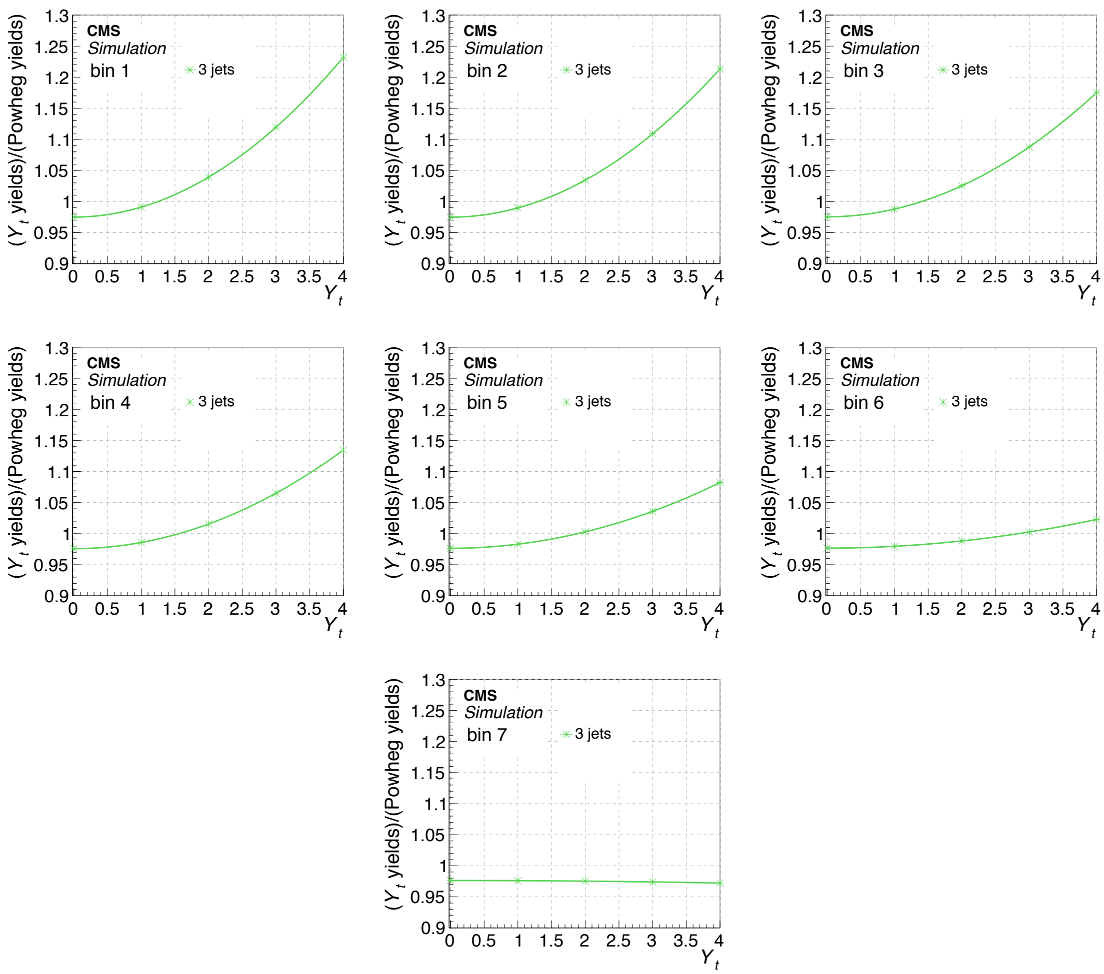


**Figure 9.3:** This plot shows $R^{\text{bin}}(Y_t)$ in their corresponding $|\Delta y_t| > 1.2$ bins along with $M_{t\bar{t}}$ in 3 jets. We apply a quadratic fit, in each bin, on for $Y_t$ couplings equal to 0, 1, 2, 3, 4, and 5. The plot shows the event yield with respect to the nominal that can be expected in a specific $M_{t\bar{t}}$ and $|\Delta y_t|$ region for each $Y_t$ scenario.

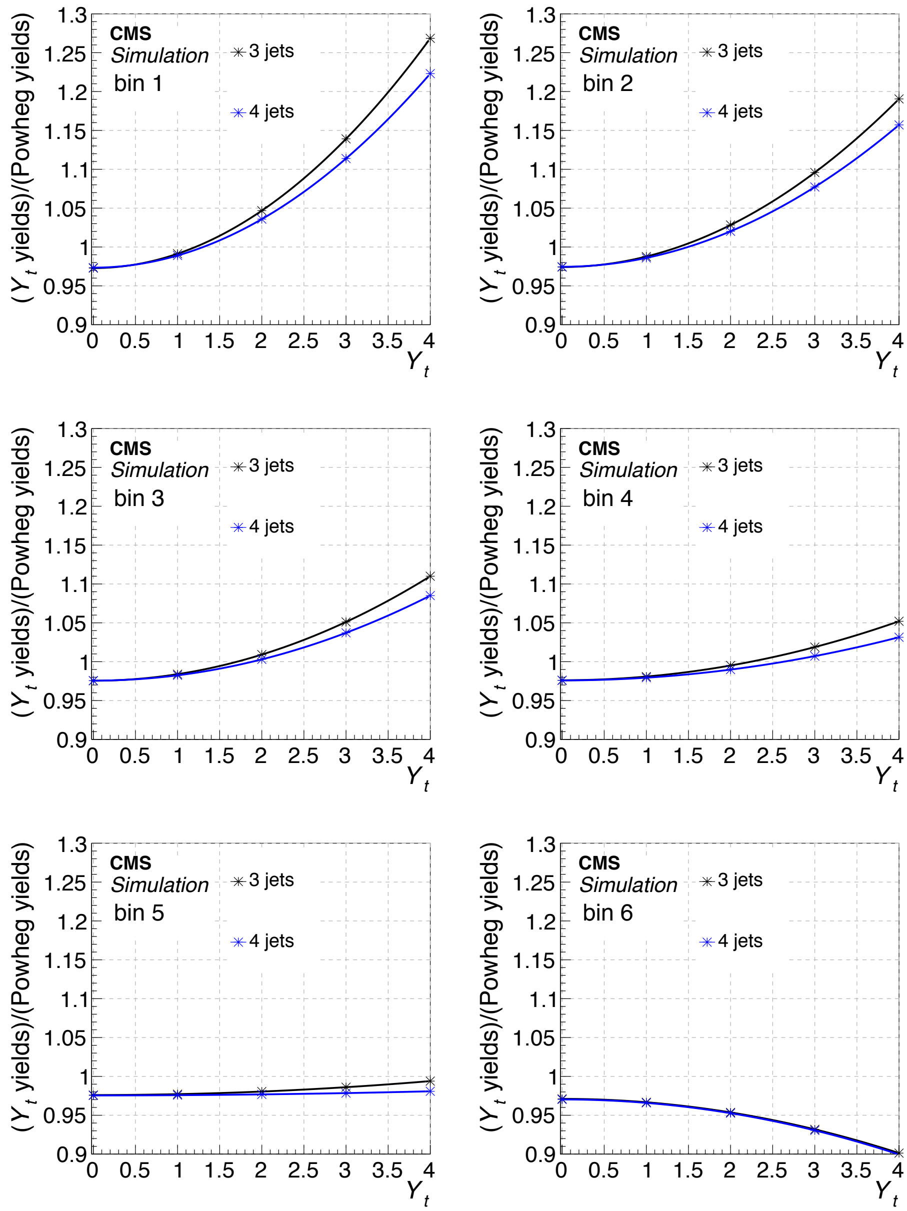


**Figure 9.4:** This plot shows $R^{\mathrm{bin}}(Y_t)$ in their corresponding $|\Delta y_t| < 0.6$ bins along with $M_{t\bar{t}}$ in 4 and $\geq$5 jets (with same binning). Colors indicate different jets multiplicities. We apply a quadratic fit, in each bin, on for $Y_t$ couplings equal to 0, 1, 2, 3, 4, and 5. The plot shows the event yield with respect to the nominal that can be expected in a specific $M_{t\bar{t}}$ and $|\Delta y_t|$ region for each $Y_t$ scenario.

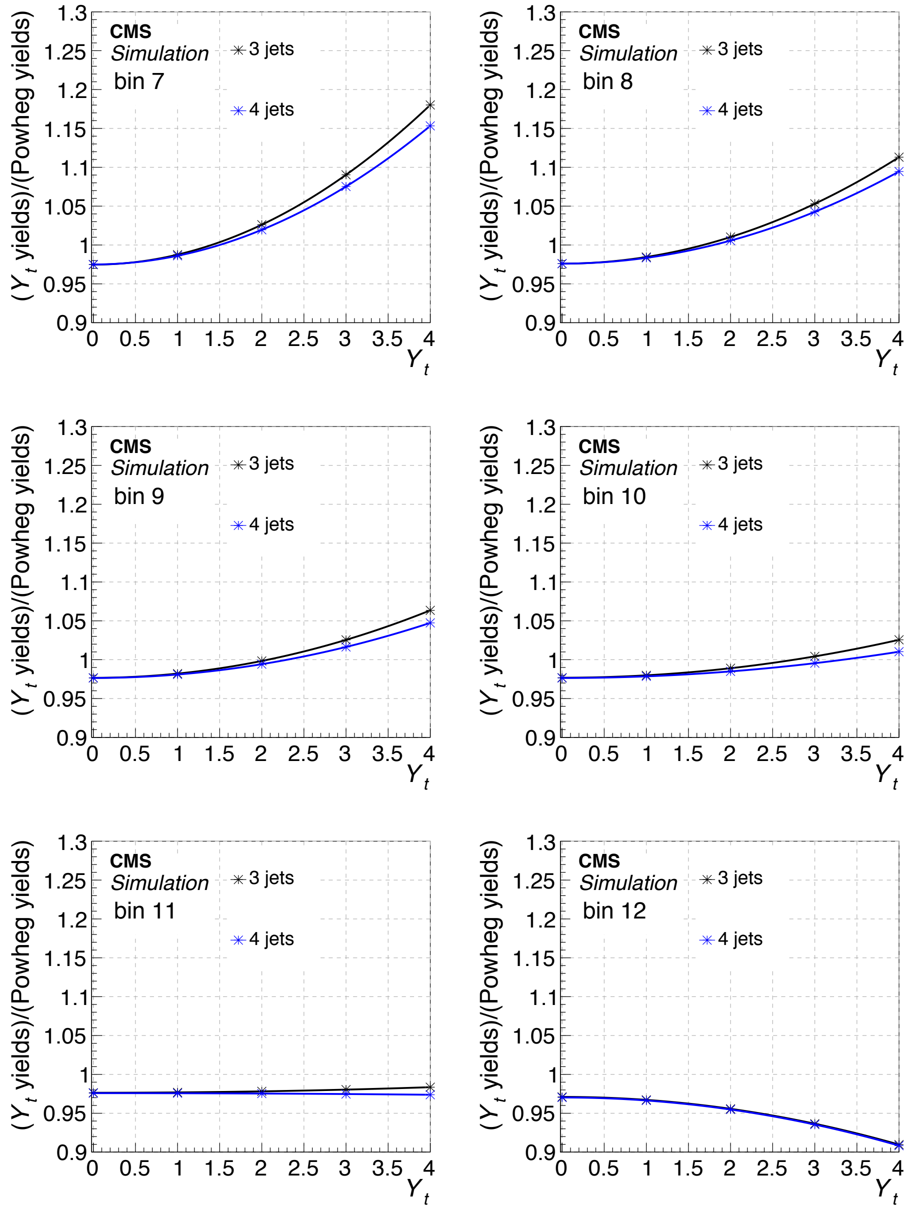


**Figure 9.5:** This plot shows $R^{\text{bin}}(Y_t)$ in their corresponding $|\Delta y_t| < 1.2$ bins along with $M_{t\bar{t}}$ in 4 and $\geq$5 jets (with same binning). Colors indicate different jets multiplicities. We apply a quadratic fit, in each bin, on for $Y_t$ couplings equal to 0, 1, 2, 3, 4, and 5. The plot shows the event yield with respect to the nominal that can be expected in a specific $M_{t\bar{t}}$ and $|\Delta y_t|$ region for each $Y_t$ scenario.

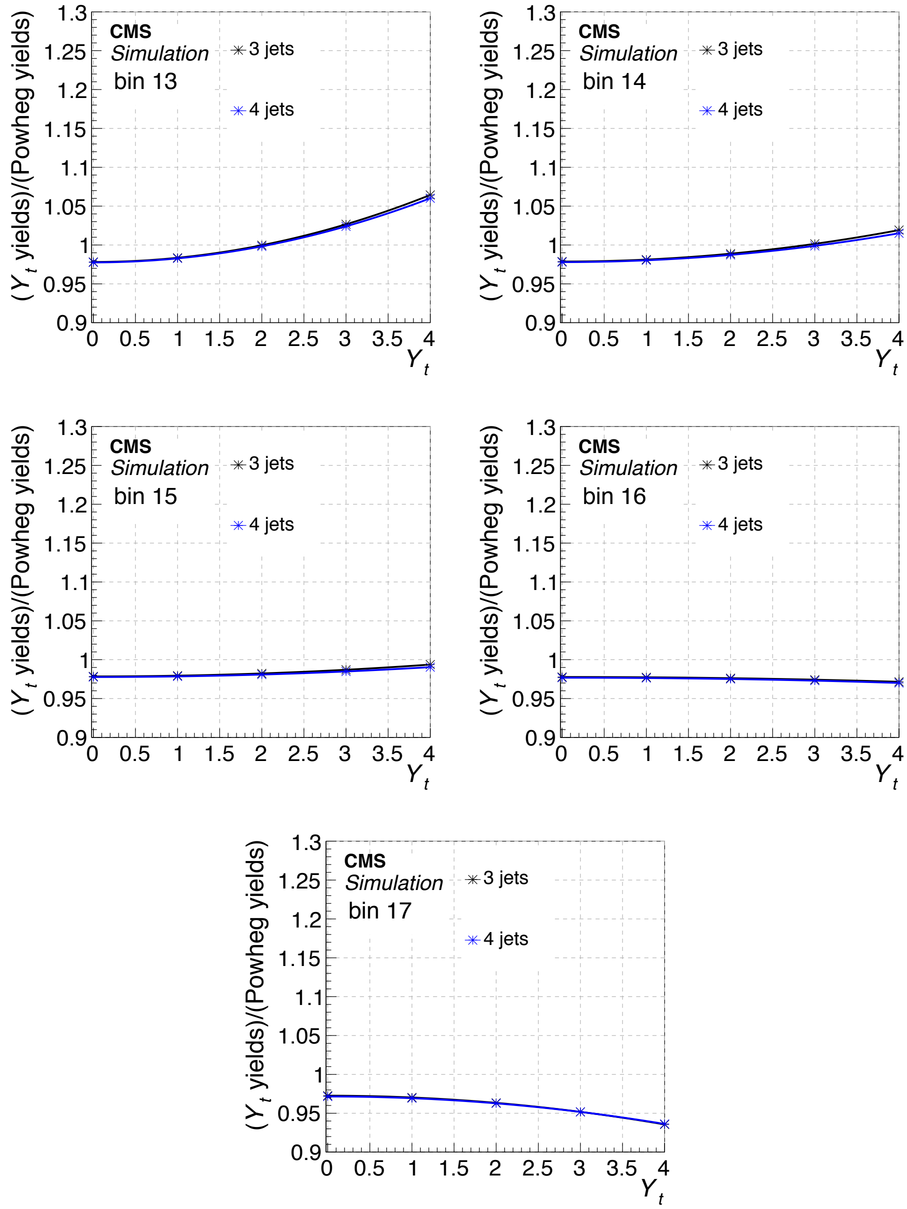


**Figure 9.6:** This plot shows $R^{\text{bin}}(Y_t)$ in their corresponding $|\Delta y_t| > 1.2$ bins along with $M_{t\bar{t}}$ in 4 and $\geq$5 jets (with same binning). Colors indicate different jets multiplicities. We apply a quadratic fit, in each bin, on for $Y_t$ couplings equal to 0, 1, 2, 3, 4, and 5. The plot shows the event yield with respect to the nominal that can be expected in a specific $M_{t\bar{t}}$ and $|\Delta y_t|$ region for each $Y_t$ scenario.

## 9.2 Systematic uncertainties

The different sources of experimental and theoretical uncertainties are described in this Section. Depending on the MC statistics in each bin, the systematic variations maybe fluctuating. In order to eliminate the effect from low MC statistics, we adopt a smoothing procedure on the distributions of systematic variations.

### 9.2.1 Experimental uncertainty

The uncertainty on the **integrated luminosity** is 2.5% [59]. The simulation samples are reweighted to match the measured data distribution in the number of collisions per events. The uncertainty on the total inelastic pp cross section, which affects the pile-up estimate, is estimated by varying the average number of pile-up events per bunch crossing by 5%.

Differences in the **pile-up** distribution between data and MC are corrected by reweighting each simulation event to match the pile-up distribution in data. The corresponding uncertainty is estimated by varying the total inelastic proton-proton cross section within its uncertainty. The overall uncertainty is less then 2%.

**Lepton efficiency scale factors**, which account for differences in the trigger, reconstruction, and identification efficiencies between data and simulation, are measured using a tag-and-probe method in $Z \to \ell\ell$ events [60,61]. The efficiencies are measured in bins of lepton $p_{\mathrm{T}}$, $\eta$, and jet multiplicities, and then applied to the simulations to match the data. The overall uncertainty from these lepton scale factors is approximately 2%.

Uncertainties in the **jet energy calibration (JEC)** are evaluated by shifting the energies of jets in the simulation up and down by one standard derivation in bins of $p_{\mathrm{T}}$ and $\eta$. According to the different sources of JEC uncertainties and the jet flavors, eighteen shape variations uncertainties are considered. JEC is the biggest experimental uncertainty in this analysis. The uncertainty in the **jet energy resolution (JES)** is calculated by broadening the resolution in simulation and recomputing the acceptances [62], for which the resulting effect is less than 1% in overall yield.

The **b tagging efficiency** in simulation is corrected using scale factors determined from efficiencies measured in data and simulation [63]. The uncertainty in the measured scale factors ranges between 1 to 20% per jet, leading to an overall effect between 2-3%.

The uncertainty from **MET** due to the electron, muon, and unclustered energy uncertainties, results in a neglectable effect on the acceptance.

As described in **background estimation**, Section 8, the single top events are affected by a 15% normalization uncertainty based on the current experimental measurements. The V+jets normalization uncertainty is 30% and the QCD multijet data-driven estimation includes a 30% normalization uncertainty and a shape difference observed between different control regions.

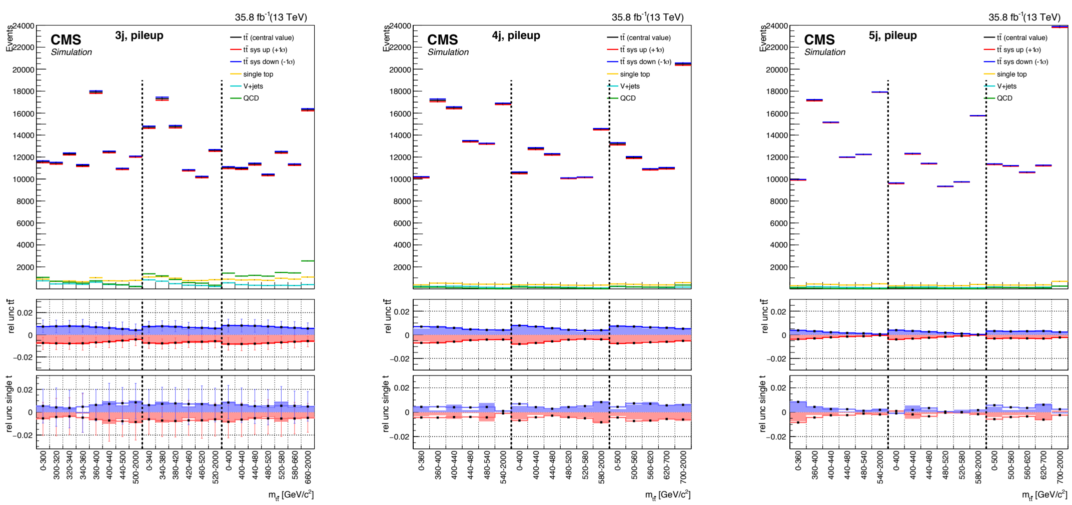


**Figure 9.7:** Shape variation of pile-up uncertainties for three jets, four jets, and five or more jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

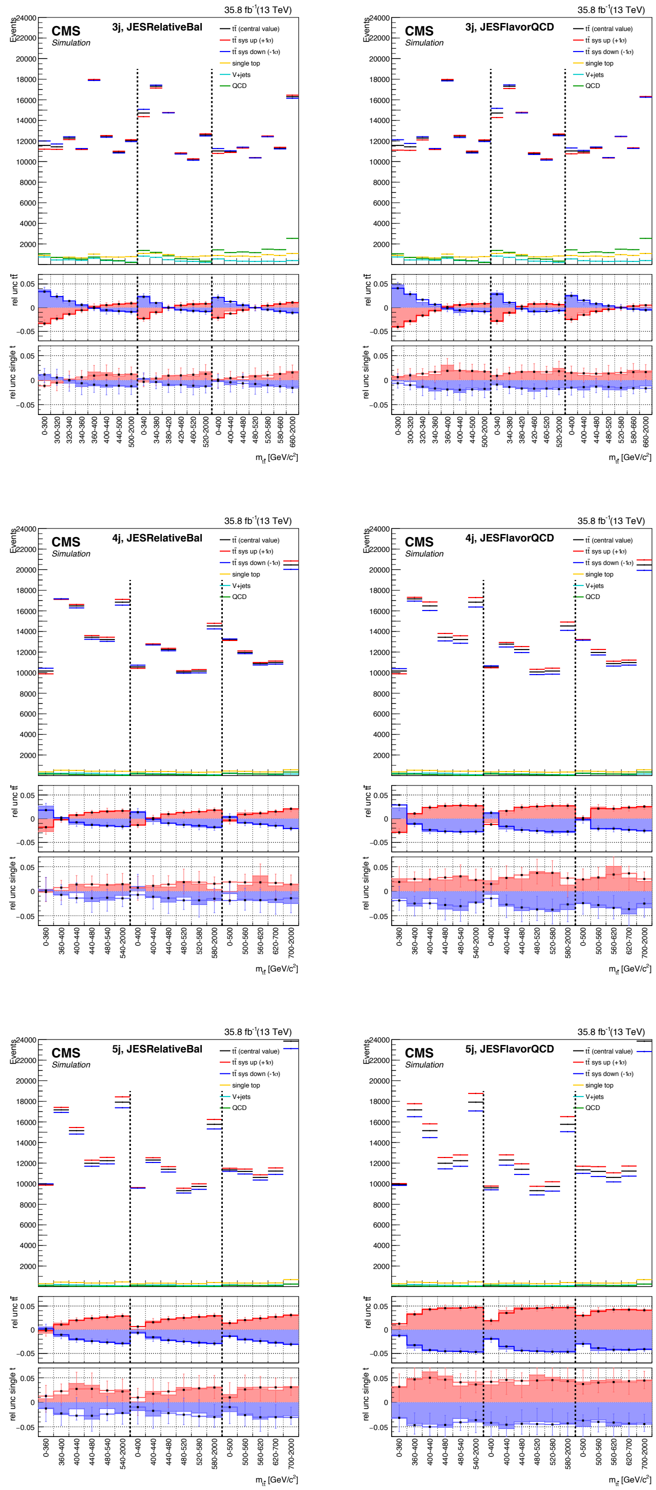


**Figure 9.8:** Shape variation of RelativeBal (difference between MPF and $p_{\mathrm{T}}$ balance methods) and FlavorQCD (difference in jet flavor and gluon responses) in JEC uncertainties for three jets, four jets, and five or more jets for $\mathrm{t\bar{t}}$ and single top background in $M_{\mathrm{t\bar{t}}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure. This is the dominant JEC uncertainty.

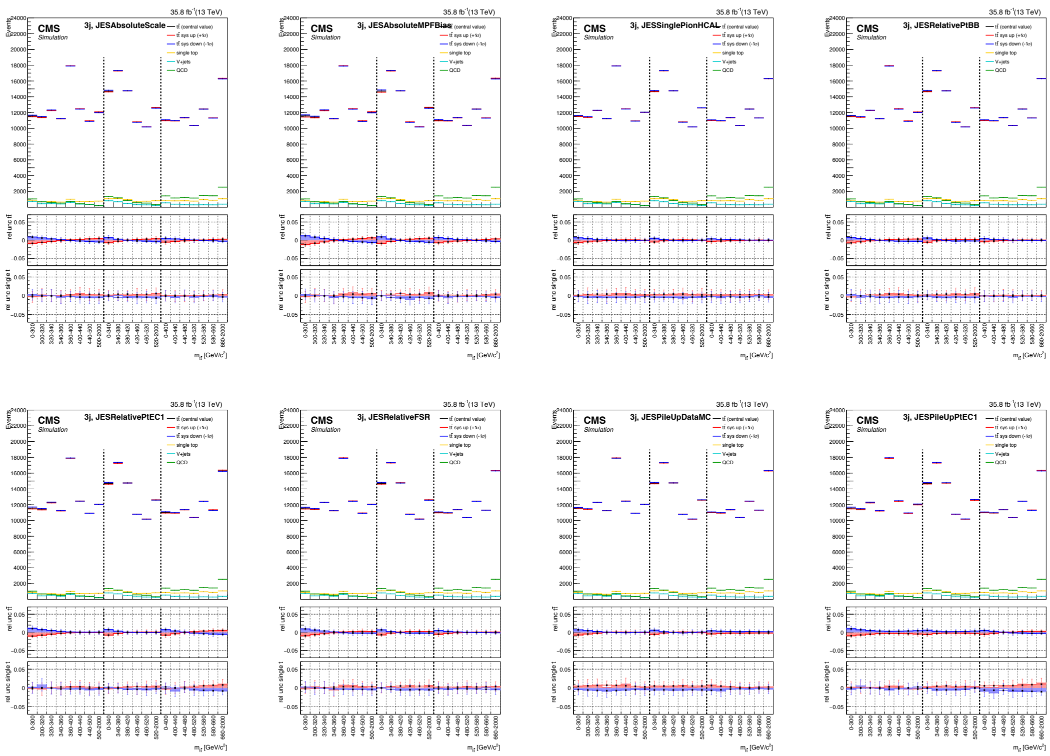


**Figure 9.9:** Shape variation of all the other JEC uncertainties for three jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

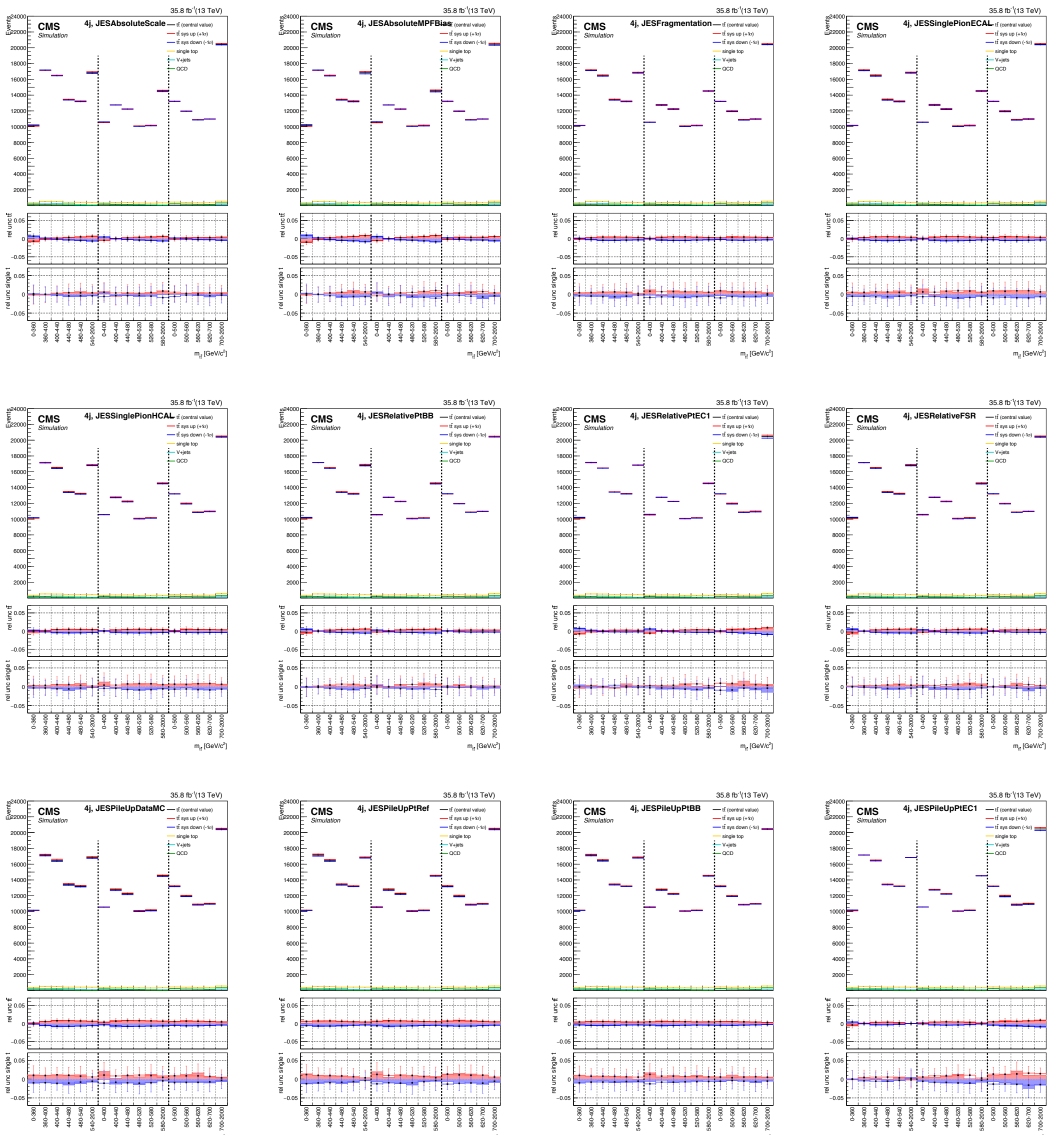

**Figure 9.10:** Shape variation of all the other JEC uncertainties for four jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

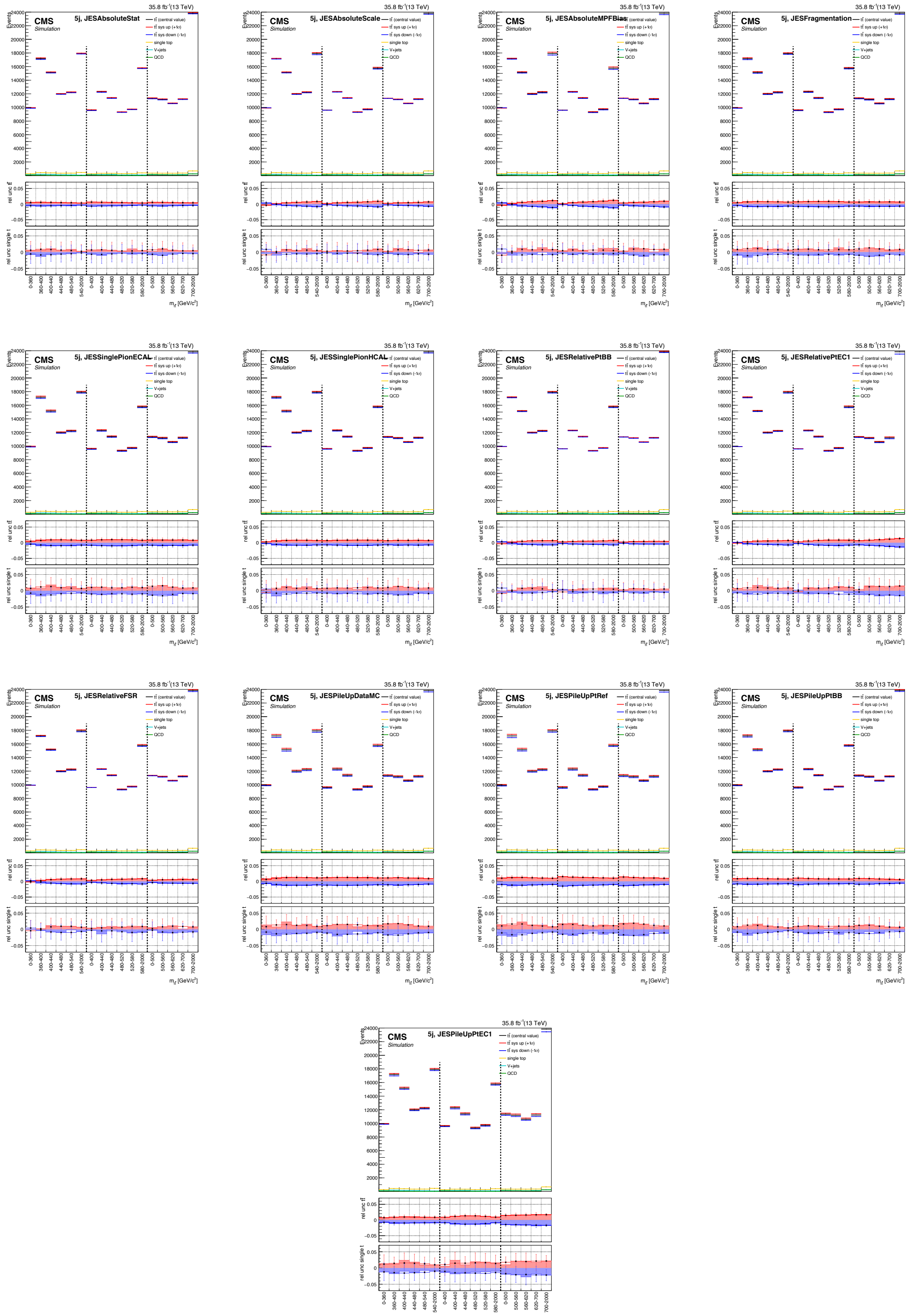


**Figure 9.11:** Shape variation of all the other JEC uncertainties for five or more jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

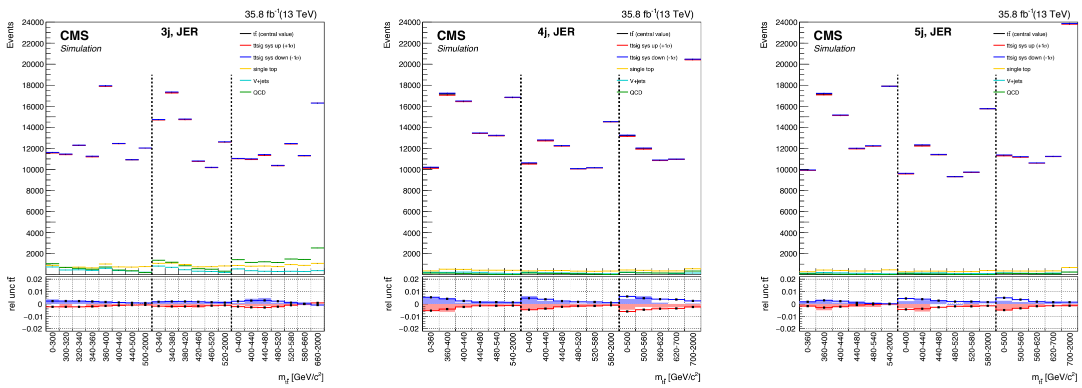


**Figure 9.12:** Shape variation of JER uncertainties for three jets, four jets, and five or more jets for $t\bar{t}$ in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

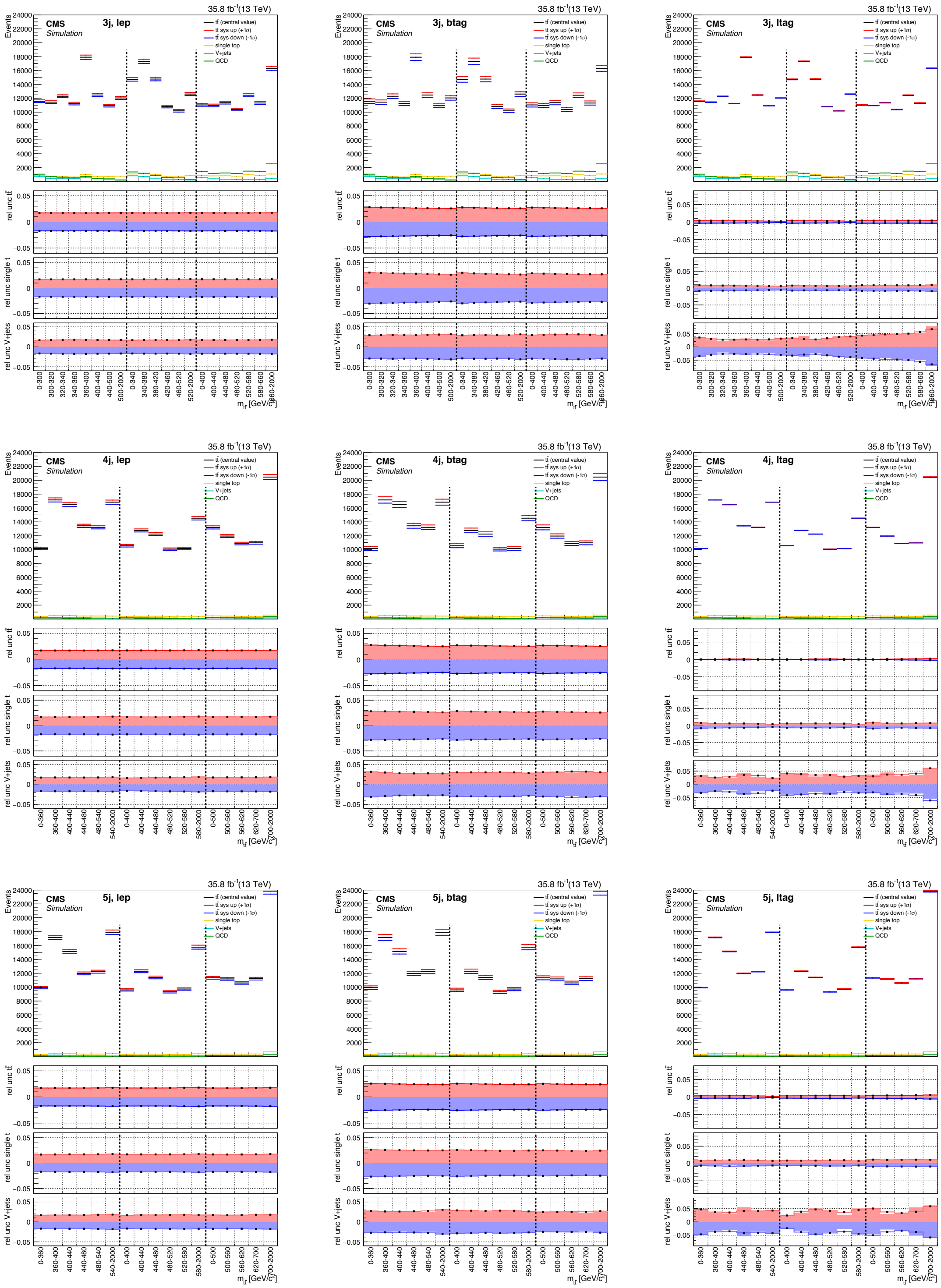


**Figure 9.13:** Shape variation of lepton ID/trigger (left), $b$ tagging scale factor (middle), and $b$ mis-tagging scale factor (right) uncertainties for three jets, four jets, and five or more jets for $t\bar{t}$, single top, and V+jets backgrounds in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

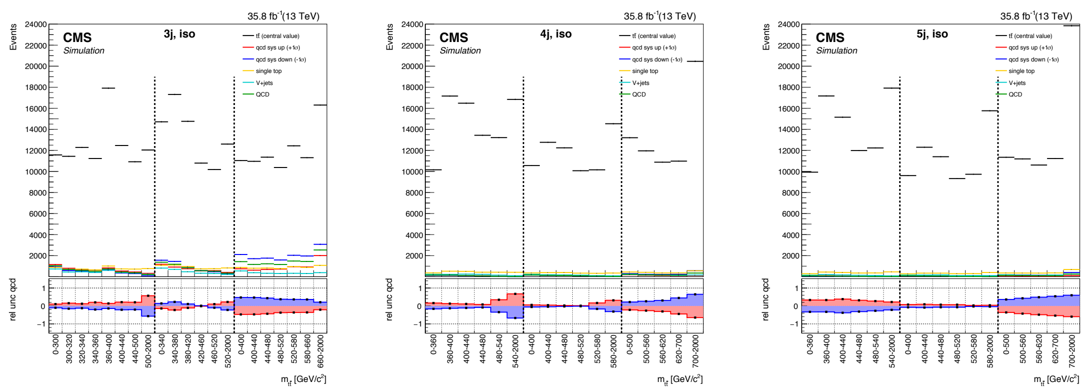


**Figure 9.14:** Shape variation of QCD shape uncertainties for three jets, four jets, and five or more jets for $\mathrm{t\bar{t}}$ in $M_{\mathrm{t\bar{t}}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

### 9.2.2 Theoretical uncertainty

Uncertainties in the **renormalization and factorization scales** affect the number of events expected for $t\bar{t}$ signal. These uncertainties are evaluated by varying each scale independently by a factor of 2 (where the anti-correlated variations are dropped). This results in shape variations as large as 5% in $t\bar{t}$ and single top events, and as large as 12% in the V+jets events.

Different alternative versions (replicas) of NNPDF30 PDF set [47] are used to estimate the corresponding acceptance uncertainty. Eight dominant base variations are used to represent the total **PDF uncertainty** which amount to a shape variation of no more than 1.5%. Different replicas due to the variation of **strong coupling** constant $\alpha_S$ result in around 1% change in the acceptance.

The uncertainty due to the **top quark mass** is estimated by taking the acceptance variations in simulations generated with $m_t$ varied up and down by 1GeV, and it results in a shape variation as large as 5% near the $t\bar{t}$ threshold.

In the following, we describe the sources of uncertainties due to modeling of the **parton shower**. The uncertainty in **matching the matrix element calculation to the parton shower** is estimated by changing the parameter which regulates the damping of real emission in the NLO calculation [64], resulting in an effect of around 1-4% for different event categories. **Color reconnection** reconfigures color strings after parton shower affecting the W decays [64]. This is also considered as a source of uncertainty resulting in an effect of around 1-3%. The scales which determine **initial-state radiation (ISR)** and **final-state radiation (FSR)** are also varied [65], resulting in maximum a 3-10% change of the shape. The uncertainty in **fragmentation**, momentum transfer from b quark to B hadron, is estimated by varying the parameterized function in the simulation. It results in a shape variation as large as 3%. The **branching fraction of B hadron decaying** semi-leptonically may change the b jet energy response and therefore, the uncertainty is calculated by reweighing the simulation with one standard variation of measurements [66]. It makes an effect on the acceptance as large as 3%. Uncertainties in modelling the amount of multiple-parton interactions [64] are found to be negligble.

It was observed in all of the top decaying channels that the available event generators has an overall harder **top $p_{\mathrm{T}}$ spectrum** than observed in data with the discrepancy still being within the analyses uncertainties. The origin of this discrepancy is not entirely understood, and it will likely require a substantial efforts from the theory community. There are indications that the discrepancy is minimized when comparing to NNLO QCD prediction [67]. In order to account for this discrepancy, an empirical reweighting based on the observed top $p_{\mathrm{T}}$ distribution is utilized in many analyses. In our analysis, no empirical reweighting is applied due to the correlation between top $p_{\mathrm{T}}$ and $M_{\mathrm{t\bar{t}}}$ and potential sensitivity to $Y_t$ from it. Instead, we reweight the average top $p_{\mathrm{T}}$ (one half of the scalar sum of the top and anti-top $p_{\mathrm{T}}$ value) by the factor computed with NNLO and we take the difference in the spectrum of $M_{\mathrm{t\bar{t}}}$ and $\Delta y_t$ before and after reweighting as a systematic uncertainty.

Finally, the **weak correction** is implemented by reweighing nominal POWHEG MC samples with the ratio of weak correction over the leading-order cross section calculated by HATHOR. As recommended by the HATHOR authors [68], the associated systematic uncertainty for this procedure can be estimated from the difference between the multiplicative and additive treatments, i.e. $(1+\delta_{QCD})(1+\delta_{EW})$ and $(1+\delta_{QCD}+\delta_{EW})$; where $\delta_{QCD}$ can be estimated by varying the factorization and renomalization scale on the NLO cross-section up and down, and $\delta_{EW}$ is the ratio of weak correction over the leading-order cross-section we got in HATHOR. The difference is $\delta_{QCD} \times \delta_{EW}$, which is also a function of $Y_t$ since the overall effect of weak correction increases for increasing Yukawa couplings.

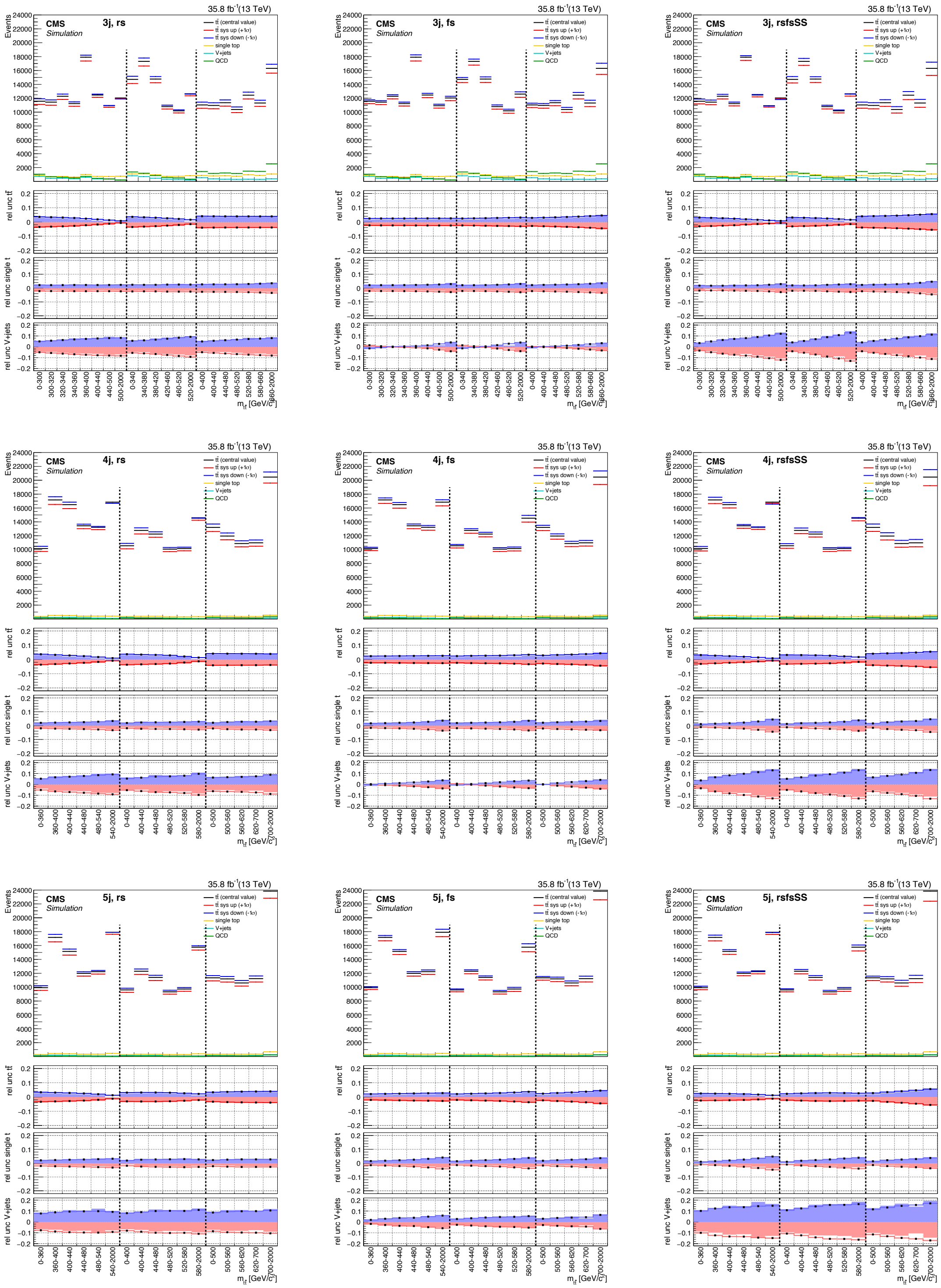


**Figure 9.15:** Shape variation of renormalization scale uncertainties (left), factorization scale uncertainties (middle), and both renormalization and factorization scale varying the same direction uncertainties (right) for three jets, four jets, and five or more jets for $t\bar{t}$, single top, and V+jets backgrounds in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

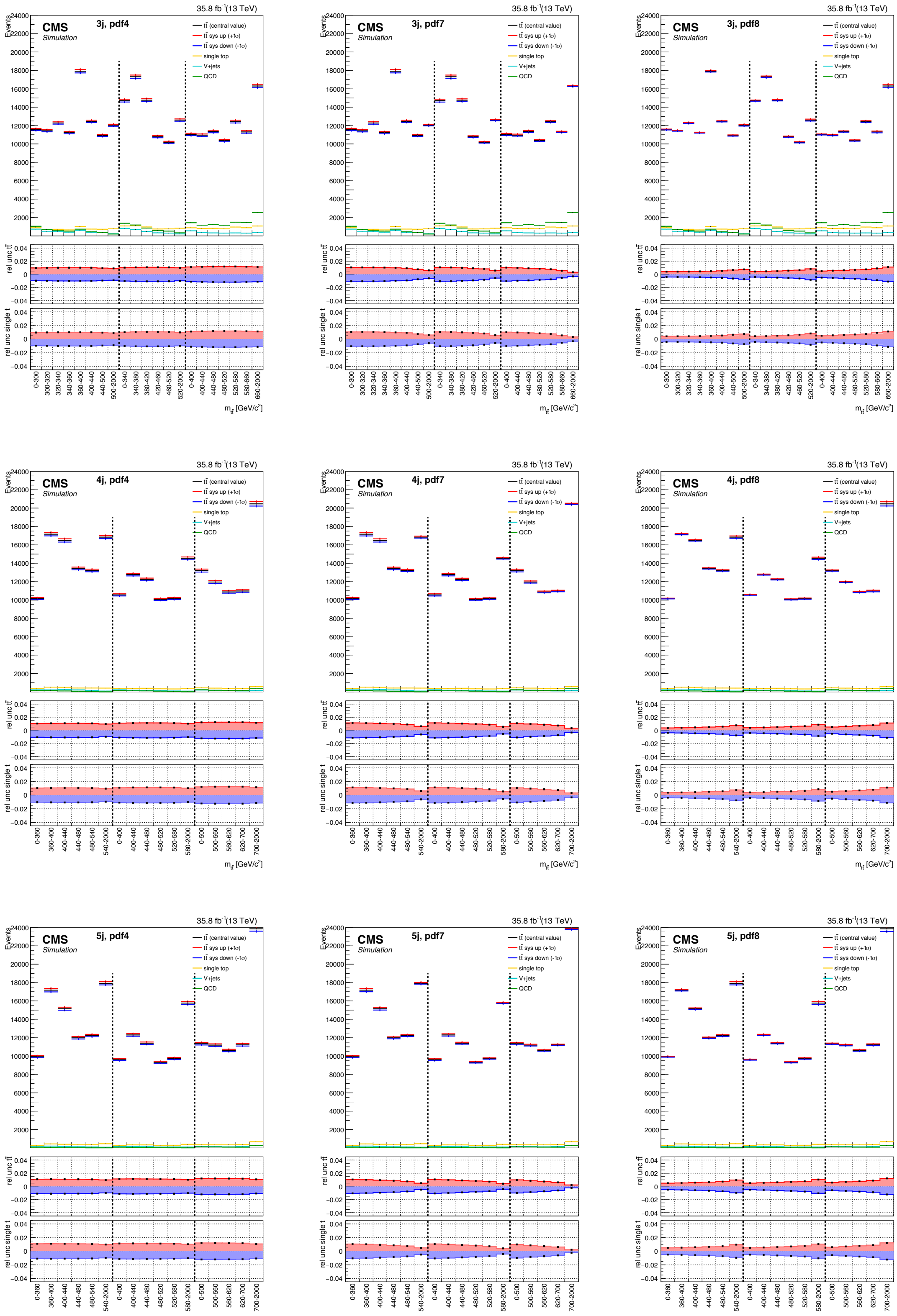


**Figure 9.16:** Shape variation of the biggest three PDF base uncertainties for three jets, four jets, and five or more jets for $t\bar{t}$ and single top backgrounds in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

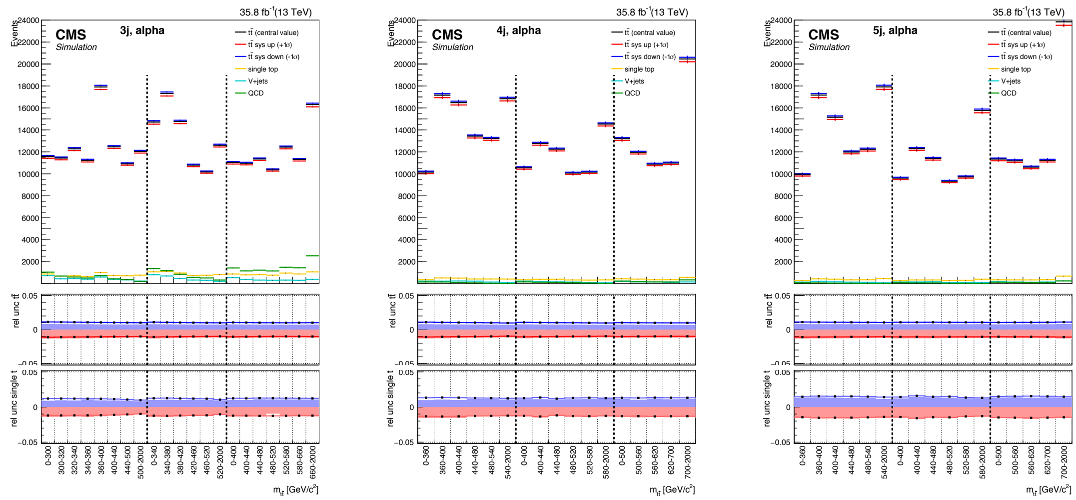


**Figure 9.17:** Shape variation of $\alpha_s(M_Z)$ in PDF uncertainties for three jets, four jets, and five or more jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

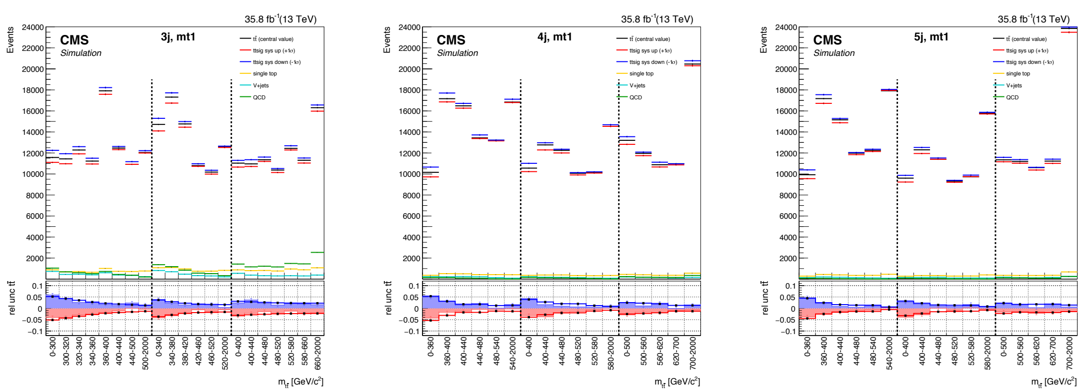


**Figure 9.18:** Shape variation of top mass uncertainties for three jets, four jets, and five or more jets for $t\bar{t}$ in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

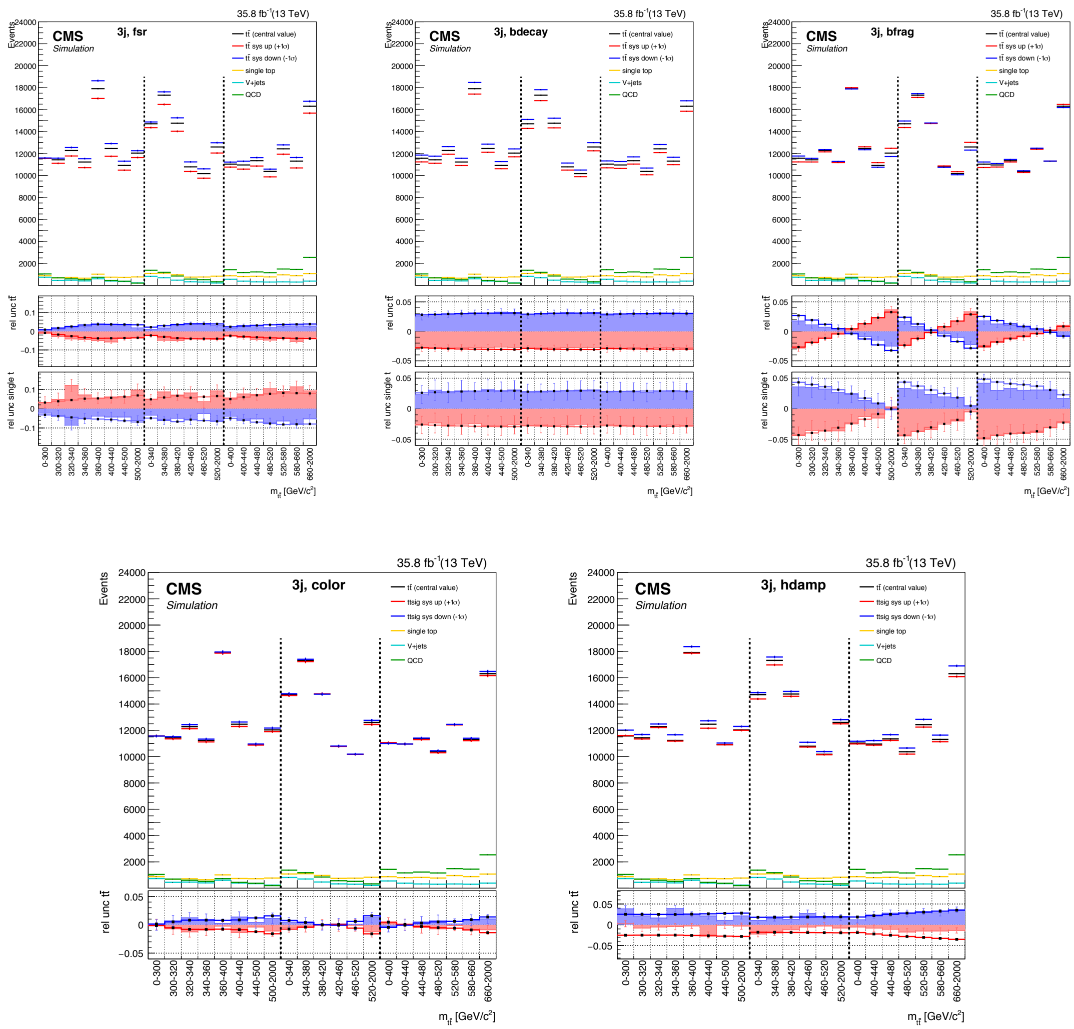


**Figure 9.19:** Shape variation of FSR (upper left), B hadron decaying Br. (upper middle), $b$-jet fragmentation (upper right), color reconnection (lower left), NLO shower matching (lower right) in parton shower uncertainties for three jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

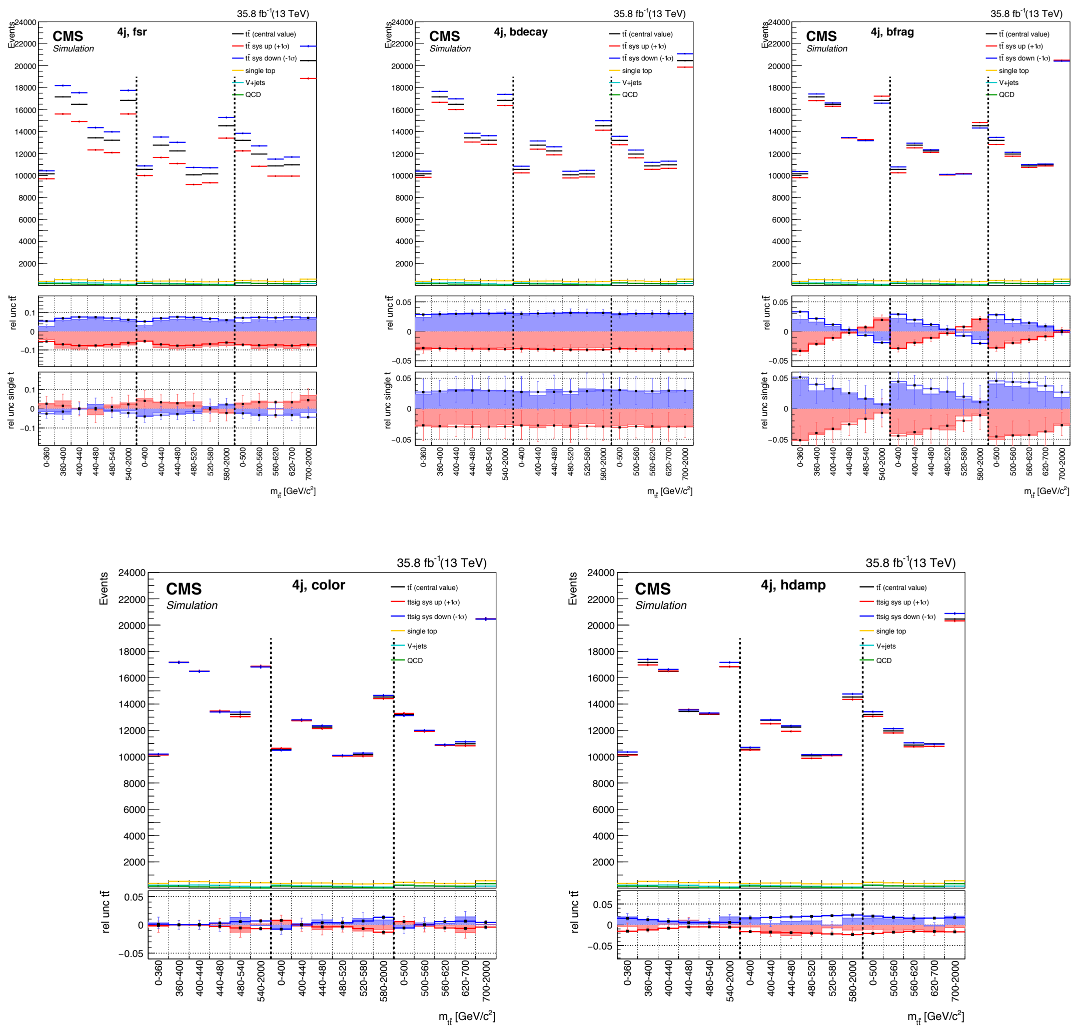


**Figure 9.20:** Shape variation of FSR (upper left), B hadron decaying Br. (upper middle), $b$-jet fragmentation (upper right), color reconnection (lower left), NLO shower matching (lower right) in parton shower uncertainties for four jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

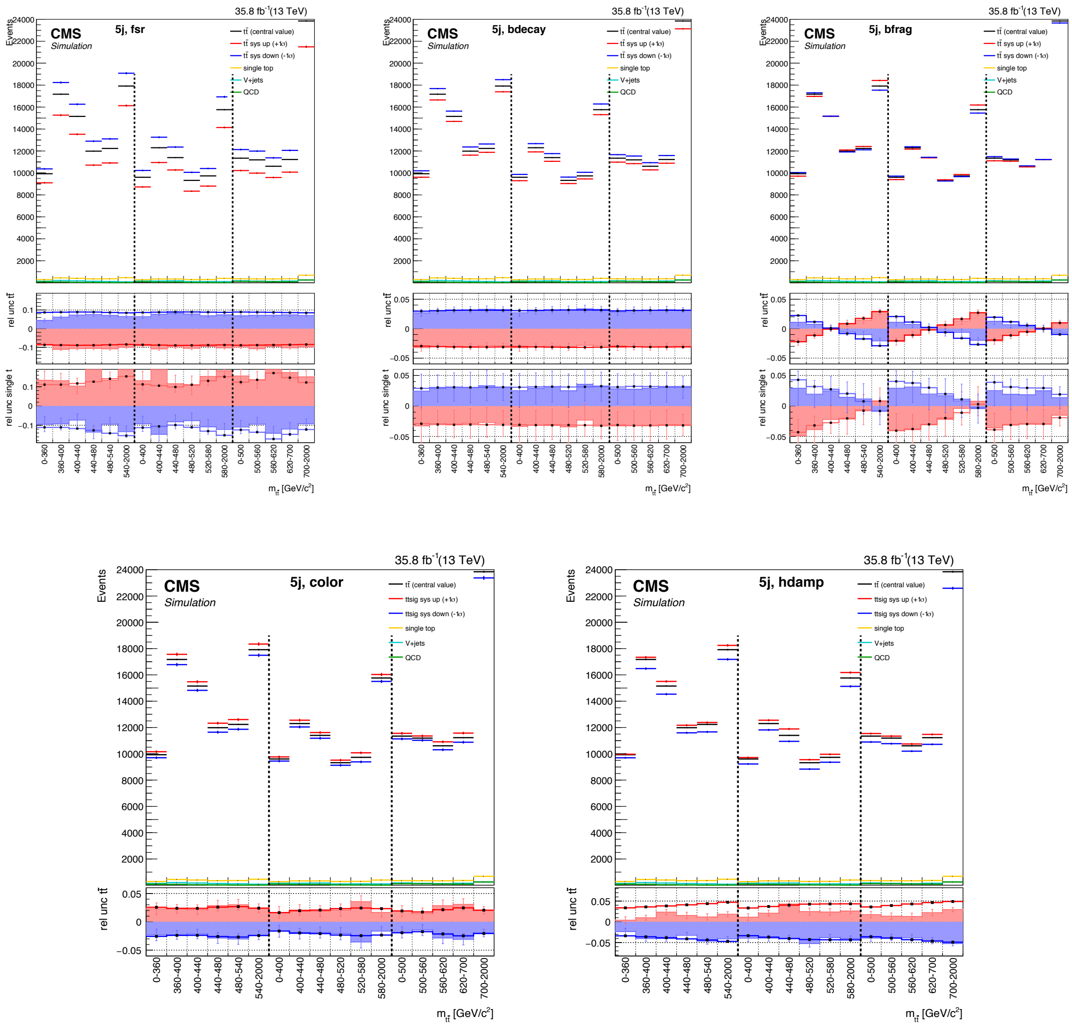


**Figure 9.21:** Shape variation of FSR (upper left), B hadron decaying Br. (upper middle), $b$-jet fragmentation (upper right), color reconnection (lower left), NLO shower matching (lower right) in parton shower uncertainties for five or more jets for $t\bar{t}$ and single top background in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

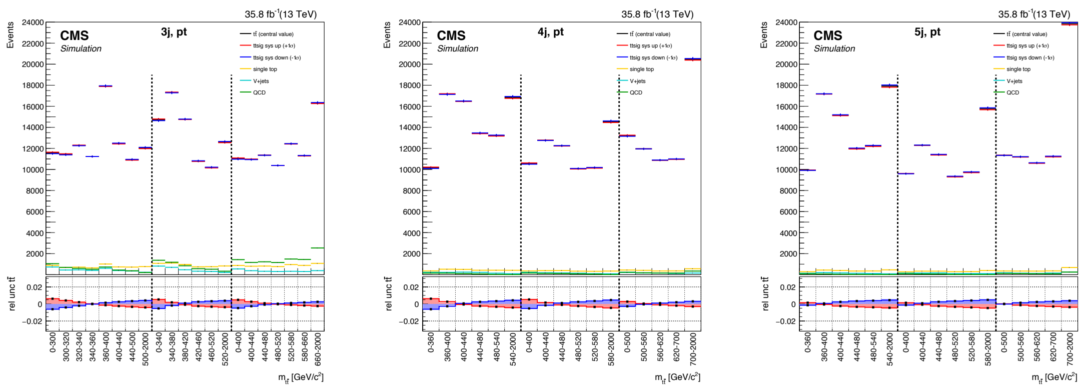


**Figure 9.22:** Shape variation of top $p_{\mathrm{T}}$ uncertainties for three jets, four jets, and five or more jets for $\mathrm{t\bar{t}}$ and single top background in $M_{\mathrm{t\bar{t}}}$ and $|\Delta y_t|$ analysis bins. The blocks show the original up and down variations, and the lines with centre dots is obtained after the smoothing procedure.

### 9.2.3 Summary of the systematic uncertainties

Systematic uncertainties, which have a relatively flat distribution in $M_{\mathrm{t\bar{t}}}$ and $\Delta y_t$ are treated as normalization uncertainties, while the others are treated as shape uncertainties. Table 9.1 lists all the systematic uncertainties. Figure 9.23 to Fig. 9.25 demonstrate uncertainties of the weak correction reweighting effect for some $Y_t$ scenarios.

Experimental uncertainties are treated as 100% correlated among signal and background processes and across the reconstruction channels. The normalization uncertainties for data-driven V+jets and QCD background are considered fully uncorrelated.

| Uncertainty | $t\bar{t}$ | single $t$ | V+jets | QCD |
|---|---|---|---|---|
| Luminosity | 2.5% | 2.5% | 2.5% | 2.5% |
| Pile-up | shape | shape | - | - |
| JEC (19 independent variations) | shape | shape | - | - |
| JER | shape | - | - | - |
| Lepton ID/trigger | shape | shape | shape | - |
| $b$ tagging scale factor | shape | shape | shape | - |
| $b$ mis-tagging scale factor | shape | shape | shape | - |
| Background normalization | - | 15% | 30% | 30% |
| CSV invertsion on QCD template | - | - | - | shape |
| Fact. & reno. scale | shape | shape | shape | - |
| PDF | shape | shape | - | - |
| $\alpha_s(M_Z)$ in PDFs | shape | shape | - | - |
| Top mass | shape | - | - | - |
| Parton Shower | | | | |
| -NLO shower matching | shape | - | - | - |
| -Color reconnection | shape | - | - | - |
| -ISR | 2%/2%/3% | - | - | - |
| -FSR | shape | shape | - | - |
| -b-jet fragmentation | shape | shape | - | - |
| -B hadron decaying Br. | shape | shape | - | - |
| Weak correction $\delta_{QCD}\delta_{EW}$ | shape | - | - | - |

**Table 9.1:** Summary of the sources of systematic uncertainties, their effects and magnitudes on signal and backgrounds. If the uncertainty shows a shape dependency on the $M_{t\bar{t}}$ and $\Delta y_t$ distributions, it is been considered in the likelihood and labeled as "shape" in the table. For columns with several numbers, the numbers refer to the events with three, four, five and more jets.

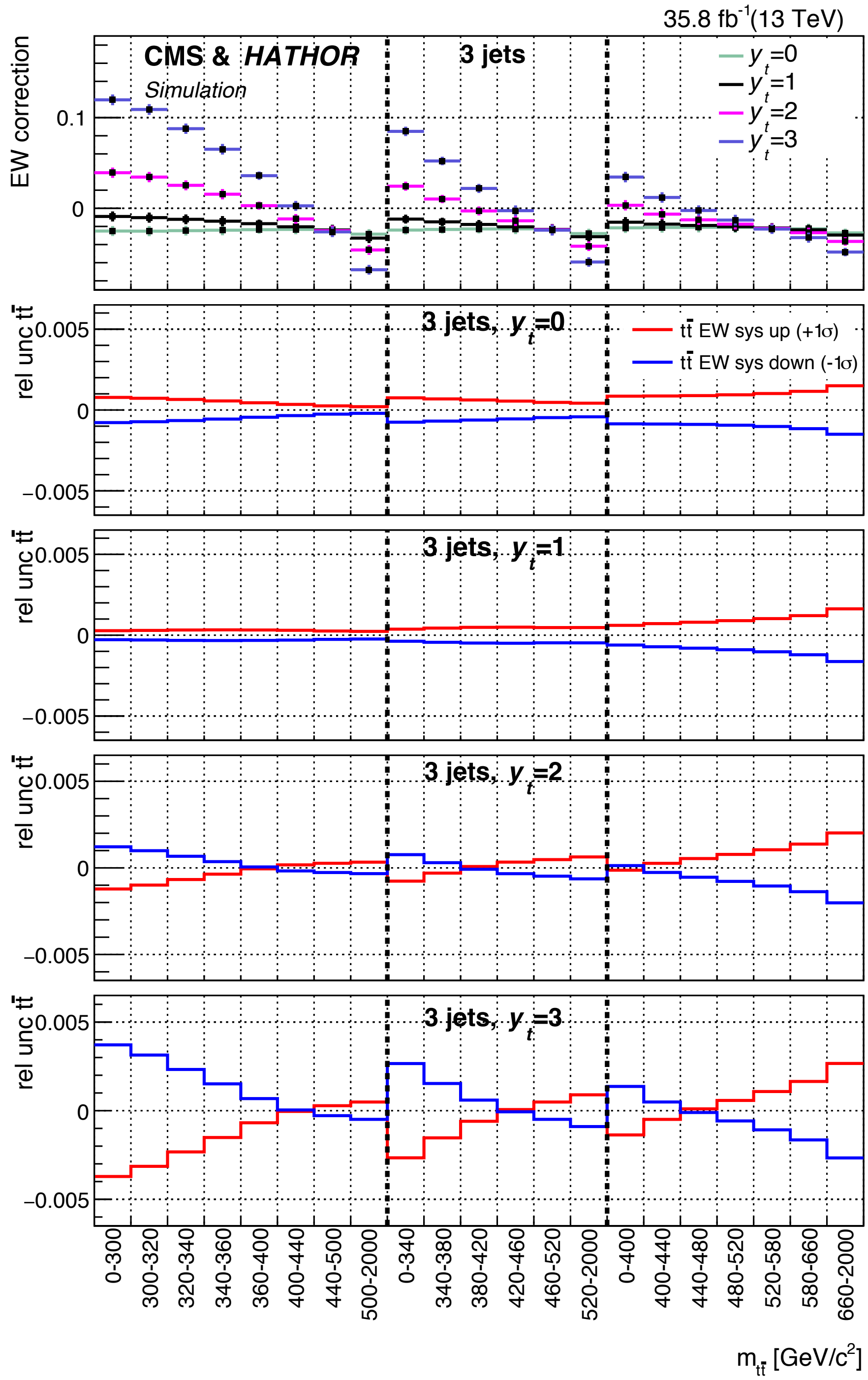


**Figure 9.23:** The effect of the weak correction uncertainties ($\pm 1\sigma$) on the three jets reconstructed $t\bar{t}$ events in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins (The vertical dashed lines separate the three $|\Delta y_t|$ regions). The top panel shows the total electroweak correction from HATHOR for different $Y_t$ values with respect to the POWHEG MC samples. The remaining plots show the relative uncertainties for each $Y_t$ scenario.

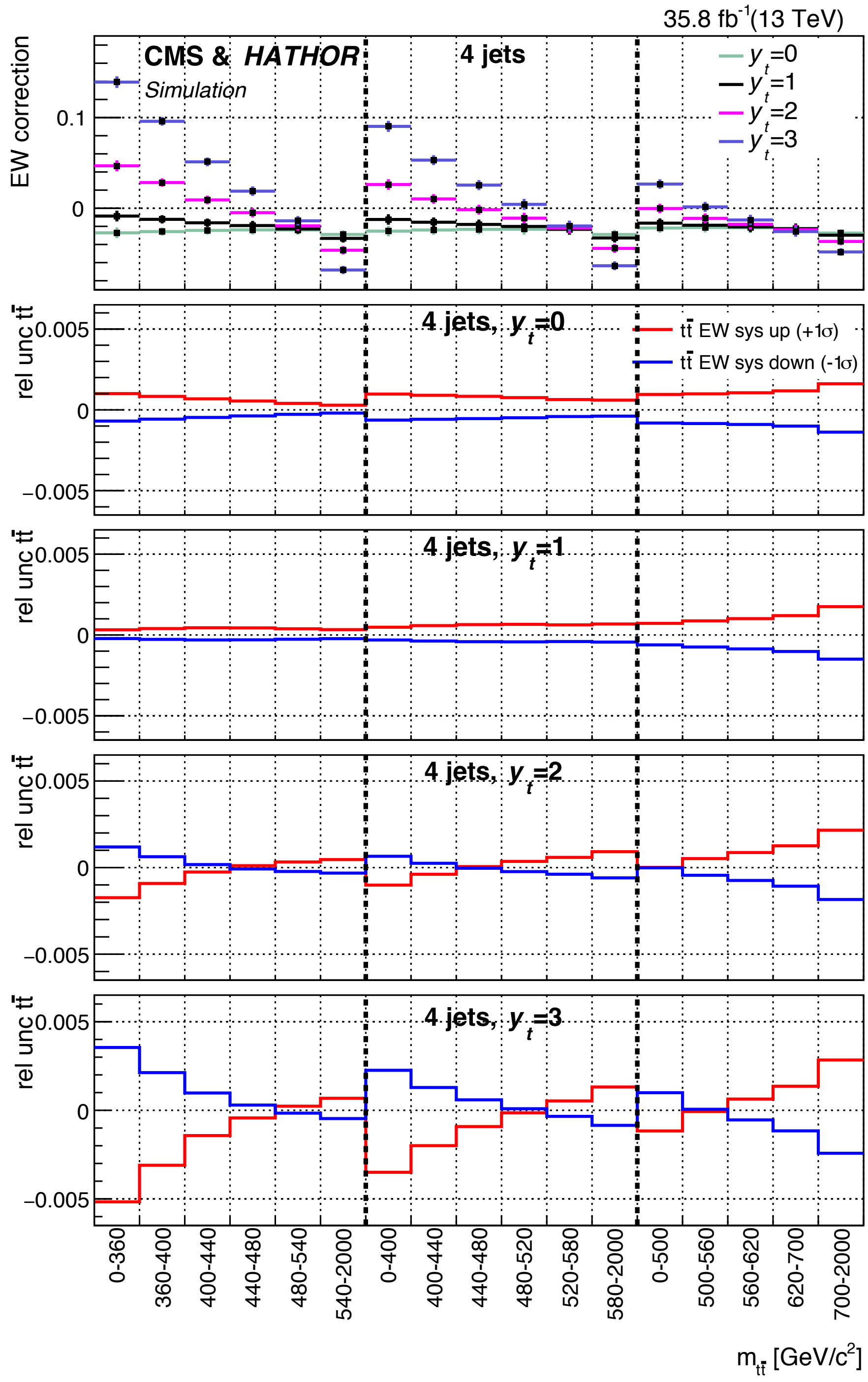


**Figure 9.24:** The effect of the weak correction uncertainties ($\pm 1\sigma$) on the four jets reconstructed $t\bar{t}$ events in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins (The vertical dashed lines separate the three $|\Delta y_t|$ regions). The top panel shows the total electroweak correction from HATHOR for different $Y_t$ values with respect to the POWHEG MC samples. The remaining plots show the relative uncertainties for each $Y_t$ scenario.

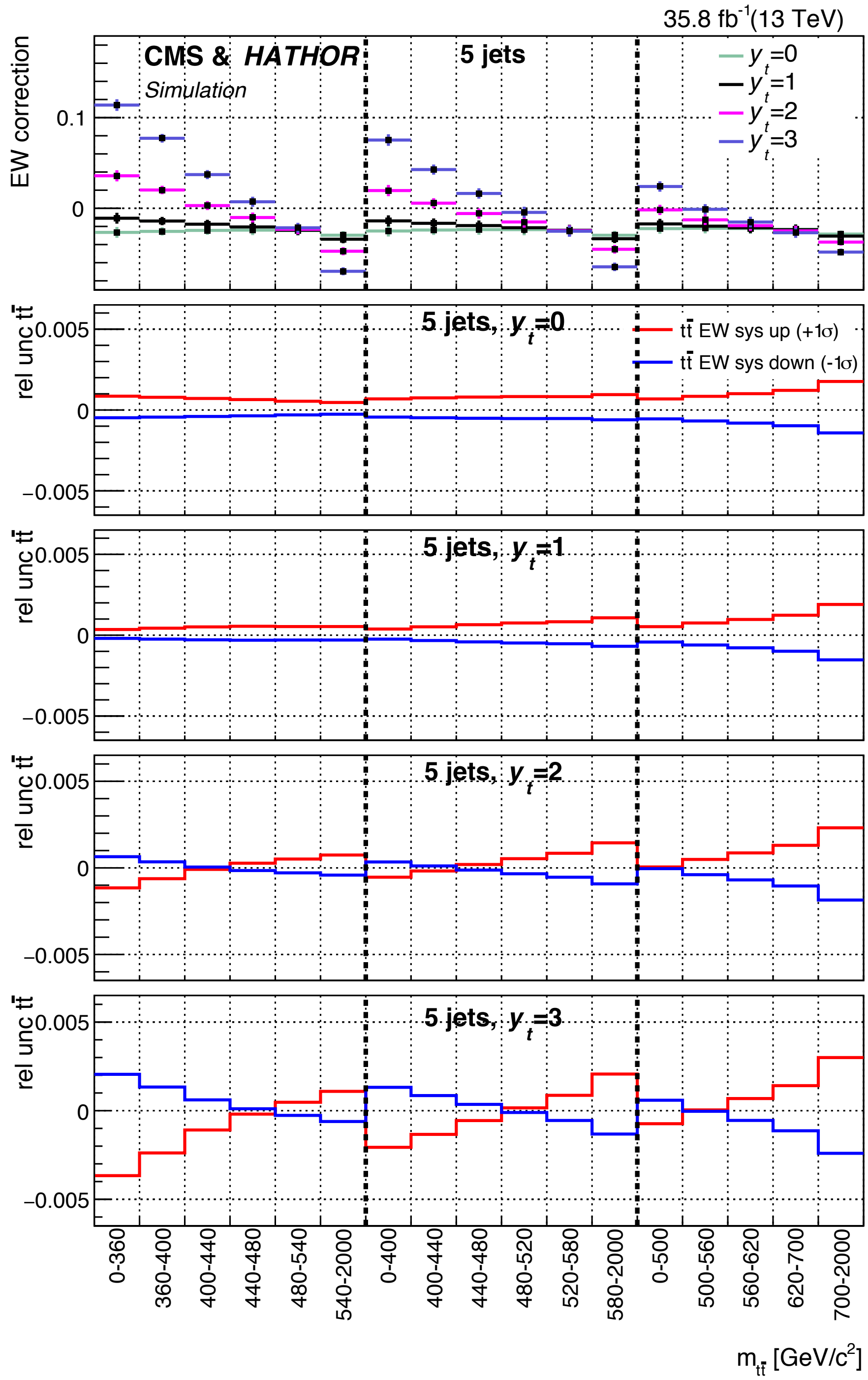


**Figure 9.25:** The effect of the weak correction uncertainties ($\pm 1\sigma$) on the five jets reconstructed $t\bar{t}$ events in $M_{t\bar{t}}$ and $|\Delta y_t|$ analysis bins (The vertical dashed lines separate the three $|\Delta y_t|$ regions). The top panel shows the total electroweak correction from HATHOR for different $Y_t$ values with respect to the POWHEG MC samples. The remaining plots show the relative uncertainties for each $Y_t$ scenario.

# Chapter 10

# Results

Events are analyzed in three exclusive channels according to the number of jets in the final state. The expected signal and background estimation are shown in Tables 8.1. The systematic uncertainties described in Section 9.2 are used to construct a binned likelihood function, Eq. 9.1, as a product of Poisson probabilities from all bins with the strength $R(Y_t) = 1$ corresponding to the signal cross section in the SM. From this we construct a profile likelihood ratio test-statistic $q(Y_t) = -2\, ln \left[ \mathcal{L}(Y_t, \hat{\hat{\theta}}) / \mathcal{L}(Y_t, \hat{\theta}) \right]$.

Here, $\hat{\hat{\theta}}$ in the numerator denotes the value of nuissance parameter $\theta$ that maximizes the likelihood for a specific value of $Y_t$, i.e., it is the conditional maximum-likelihood (ML) estimator of $\theta$ (and thus it is a function of $Y_t$). The denominator is the maximized (unconditional) likelihood function, i.e., $Y_t$ and $\hat{\theta}$ assume the values that simultaneously maximize the likelihood. The statistical procedure to extract the parameter of interest ($Y_t$) is detailed in [69].

We set a limit on the top Yukawa coupling by scanning the likelihood with respect to $Y_t$. At each scanned point, the coupling $Y_t$ is profiled with respect to the SM value 1. Systematics are taken into account via the penalty terms in the likelihood. The intersections of test-statistics with horizontal lines gives upper limits of two sided or one sided intervals which depend on the available scan region. The limits on $Y_t$ are reported in 68% and 95% confidence levels (CL) where intersections of the test-statistics are 1 and 3.84 respectively. The likelihood scan distributions are shown in Fig. 10.1. The expected and the observed limits are presented in Table 10.1.

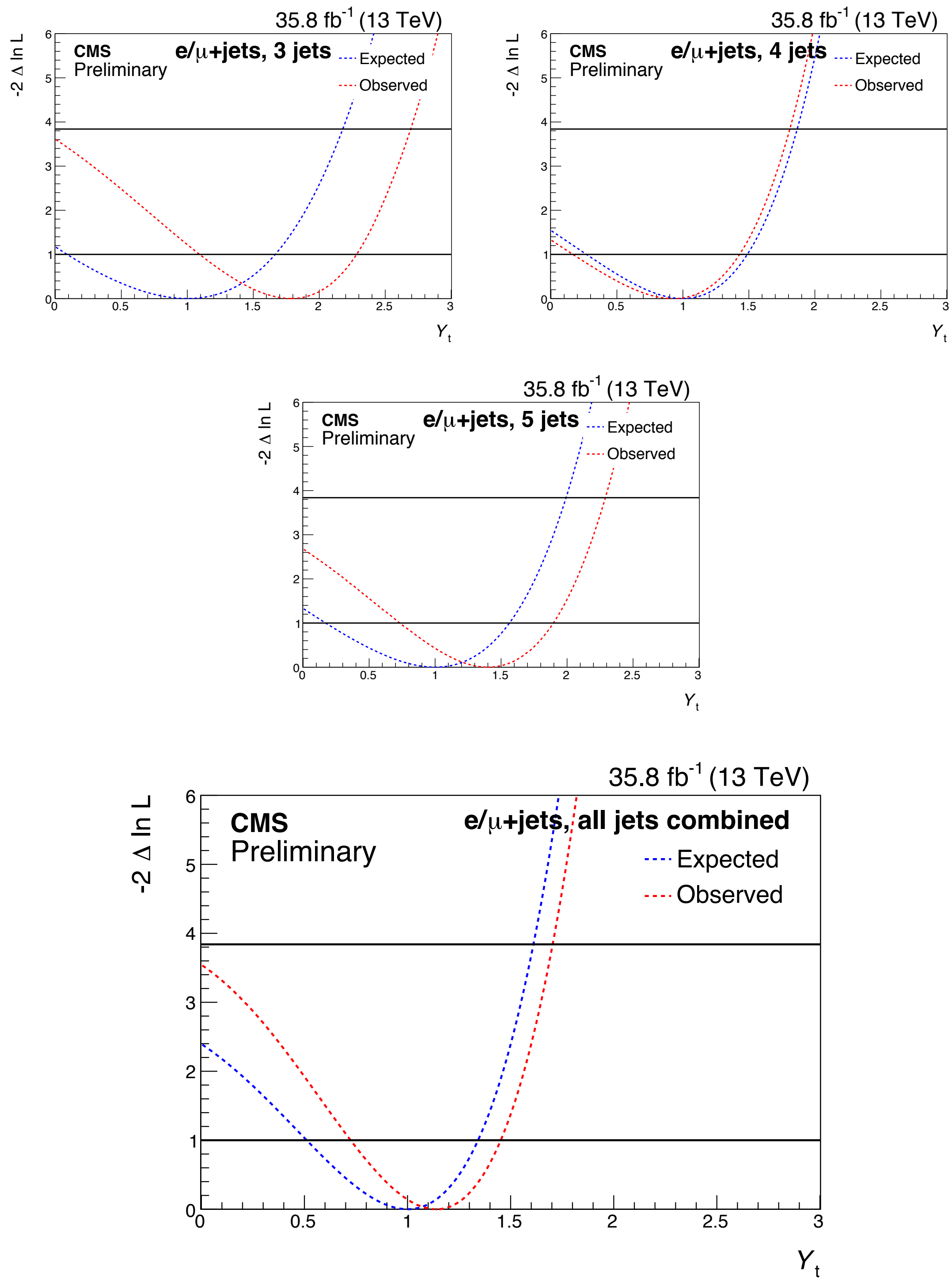


**Figure 10.1:** The test-statistic scan versus $Y_t$ for each channel (three jets, four jets, five and more jets), and all channels combined. The interpolated scan minimum indicates the best fit of $Y_t$. The red lines indicate 68% CL and 95% CL, respectively.

| Channel | Expected 95% CL | Observed 95% CL |
|---|---|---|
| 3 jets | $Y_t < 2.18$ | $Y_t < 2.69$ |
| 4 jets | $Y_t < 1.88$ | $Y_t < 1.82$ |
| $\geq$5 jets | $Y_t < 2.01$ | $Y_t < 2.30$ |
| Combined | $Y_t < 1.62$ | $Y_t < 1.70$ |

**Table 10.1:** The expected and observed 95% CL intervals on $Y_t$.

# Chapter 11

# Conclusion

A measurement of the the top-Higgs Yukawa coupling is performed using top pair production in the semi-leptonic channel. The dataset corresponds to proton-proton collisions at $\sqrt{s} = 13$ TeV with an integrated luminosity of 35.9 fb$^{-1}$. The analysis uses two kinematic variables of the top pair system, $M_{t\bar{t}}$ and $|\Delta y_{t\bar{t}}|$, to infer any possible deviation in $Y_t$ from the SM prediction. A new reconstruction algorithm is developed to reconstruct the top pair system in the case of one jet lost due to the detector acceptance. Residual multijet backgrounds are estimated by data-driven techniques.

The top-Higgs Yukawa coupling can also be constrained indirectly by the measurement of $tttt$ production [70]. Following this approach, a search for $tttt$ production in the same-signed and three-lepton final state [71] reports a upper limit of $|Y_t/Y_t^{SM}| < 2.1$, corresponds to an integrated luminosity of 35.9 fb$^{-1}$. For direct measurements, although the $ttH$ production has been observed by the CMS collaboration [19], no direct measurement has been performed on the top-Higgs Yukawa coupling.

This analysis is systematics dominanted, which means the upper limit will not be largely improved by only using more CMS collected data. It can only be sizable improved if we have further understanding of the systematics uncertainties. There are several future improvements in some aspects could be considered for the next study using similar conditions. The most significant and challenging is the estimation of the multijet background. It is the dominant background source in three jets category. A well-defined background estimation method is applied and corresponding uncertainties are assigned. Because the assigned uncertainty is not small, it dilutes in part the improvement on the upper limit we originally expected from including the three jets category. Another improvement could be on the binning optimization when the MC samples are generated with more statistics. A finner binning can be applied near the $M_{\mathrm{t\bar{t}}}$ threshold region to gain more shape information in the likelihood model. With more statistics in the five or more jets reconstruction category, it could be broken down into a few more channels. It could help on the upper limit of the combine channel result. It would also be interesting to extend the $M_{\mathrm{t\bar{t}}}$ to higher region, which is expected to be sensitive to the Yukawa coupling as well.

Moreover, a two dimensional likelihood model constraining simultaneously the Yukawa coupling and the top mass is well motivated. A preliminary study has been done in this analysis using the same dataset and the same likelihood model. It indicates a promising result, Figure. 11.1, showing that is worth pursuing in the next study. We can expect a better result with the improvements mentioned above. The results can also be improved by the other SM top measurements, because this helps the understandings of many systematic uncertainties in this analysis.

The LHC is expected to deliver data up to 3000 fb$^{-1}$. Possibilities of further constraining $Y_t$ would require exploitation of this large dataset and reductions in various systematic uncertainties. Any deviation in $Y_t$ could point to potential Beyond-Standard-Model signatures. It is pivotal to perform precise measurements on the properties of Higgs boson and other SM processes, to complete our understanding of physics at TeV scale, and provide possible new directions for particle physics.

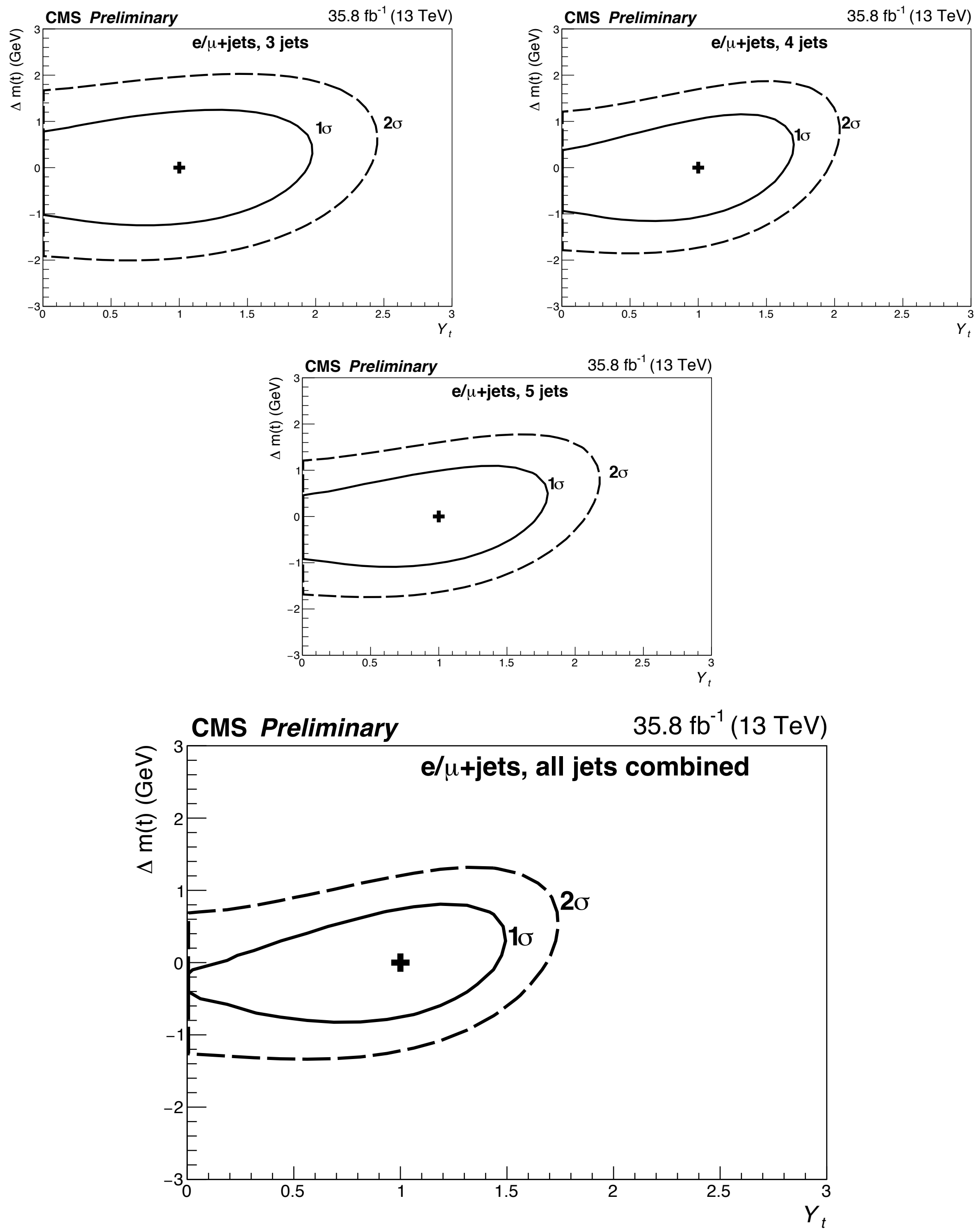


**Figure 11.1:** The test-statistic scan of $Y_t$ versus the derivation of the top mass MC setting (172.5 GeV) for each channel (three jets, four jets, five and more jets), and all channels combined. The interpolated scan minimum indicates the best fit of $Y_t$. The contours indicate 68% CL and 95% CL, respectively.

# Bibliography

[1] F.Abe et al. (CDF Collaboration), "Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (1995), no. 2626,.

[2] S. Abachi et al. (D0 Collaboration), "Observation of the Top Quark Production", *Phys. Rev. Lett.* **74** (1995), no. 2632,.

[3] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29.

[4] 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.

[5] J. H. Oort, "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems", *Bull. Astron. Inst. Netherlands* **6** (1932) 249–287.

[6] M. Markevitch, A. H. Gonzalez, D. Clowe et al., "Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56", *Astrophys. J.* **606** (2004) 819–824, doi:10.1086/383178, arXiv:astro-ph/0309303.

[7] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", *Astron. Astrophys.* **594** (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589.

[8] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C40** (2016), no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.

[9] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[10] T. Muta, "Foundations of Quantum Chromodynamics: An Introduction to Perturbative Methods in Gauge Theories". World Scientific lecture notes in physics. World Scientific, 2010.

[11] V. Barger and R. Phillips, "Collider Physics". Frontiers in Physics. Addison-Wesley Publishing Company, 1987.

[12] UA1 Collaboration, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV", *Phys. Lett.* **122B** (1983) 103–116, doi:10.1016/0370-2693(83)91177-2. [,611(1983)].

[13] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider", *Phys. Lett.* **126B** (1983) 398–410, doi:10.1016/0370-2693(83)90188-0.

[14] UA2 Collaboration, "Evidence for Z0 —> e+ e- at the CERN anti-p p Collider", *Phys. Lett.* **129B** (1983) 130–140, doi:10.1016/0370-2693(83)90744-X.

[15] UA2 Collaboration, "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider", *Phys. Lett.* **122B** (1983) 476–485, doi:10.1016/0370-2693(83)91605-2.

[16] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[17] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[19] CMS Collaboration Collaboration, "Observation of $t\bar{t}H$ Production", *Phys. Rev. Lett.* **120** (Jun, 2018) 231801, doi:10.1103/PhysRevLett.120.231801.

[20] T. A. Collaboration, "Measurement of the top quark-pair production cross section with ATLAS in pp collisions at 7 TeV", *Eur. Phys. J. C* **71** (2011) 1577.

[21] T. C. Collaboration, "First measurement of the cross section for top-quark pair production in proton-proton collisions at 7 TeV", *Physics Letters B* **695** (2011), no. 5, 424 – 443, doi:https://doi.org/10.1016/j.physletb.2010.11.058.

[22] CMS Collaboration Collaboration, "Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton + jets events from $pp$ collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **97** (Jun, 2018) 112003, doi:10.1103/PhysRevD.97.112003.

[23] 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.

[24] J. H. KÃijhn, A. Scharf, and P. Uwer, "Weak Interactions in Top-Quark Pair Production at Hadron Colliders: An Update", *Phys. Rev. D* **91** (2015), no. 1, 014020, doi:10.1103/PhysRevD.91.014020, arXiv:1305.5773.

[25] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *Submitted to: Phys. Rev. D* (2016) arXiv:1610.04191.

[26] M. Aliev, H. Lacker, U. Langenfeld et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034–1046, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[27] W. Beenakker, A. Denner, W. Hollik et al., "Electroweak one loop contributions to top pair production in hadron colliders", *Nucl. Phys.* **B411** (1994) 343–380, doi:10.1016/0550-3213(94)90454-5.

[28] P. Brogi, "Study of inelastic processes in proton-proton collisions at the LHC with the TOTEM Experiment", Master's thesis, INFN, Pisa, 2011.

[29] CMS Collaboration, "CMS physics: Technical design report",.

[30] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (May, 2014) P10009. 80 p, doi:10.1088/1748-0221/9/10/P10009.

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

[32] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[33] W. Adam, R. Frühwirth, A. Strandlie et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *J. Phys. G: Nucl. Part. Phys.* **31** (2005) N9, doi:10.1088/0954-3899/31/9/N01.

[34] CMS Collaboration, “Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev. Lett.* **112** (2014) 191802, doi:10.1103/PhysRevLett.112.191802, arXiv:1402.0923.

[35] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k t jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063.

[36] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *The European Physical Journal C* **72** (Mar, 2012) 1896, doi:10.1140/epjc/s10052-012-1896-2.

[37] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”,.

[38] CMS Collaboration Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, (2016).

[39] CMS Collaboration Collaboration, “Measurement of jet activity in $t\bar{t}$ events using lepton + jets final states in proton-proton collisions at 13 TeV”, Technical Report CMS-AN-17-003, CERN, 2017. Geneva, (Jun, 2017).

[40] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040,

arXiv:hep-ph/0409146.

[41] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[42] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[43] J. M. Campbell, R. K. Ellis, P. Nason et al., "Top-pair production and decay at NLO matched with parton showers", *JHEP* **04** (2015) 114, doi:10.1007/JHEP04(2015)114, arXiv:1412.1828.

[44] T. Sjöstrand, 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.

[45] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[46] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

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

[48] 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.

[49] M. H. Seymour and A. Siodmok, "Constraining MPI models using $\sigma_{eff}$ and recent Tevatron and LHC Underlying Event data", *JHEP* **10** (2013) 113, doi:10.1007/JHEP10(2013)113, arXiv:1307.5015.

[50] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[51] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to

dilepton production in FEWZ", *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[52] P. Kant, O. M. Kind, T. Kintscher et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

[53] N. Kidonakis, "NNLL threshold resummation for top-pair and single-top production", *Phys. Part. Nucl.* **45** (2014) 714, doi:10.1134/S1063779614040091, arXiv:1210.7813.

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

[55] B. A. Betchart, R. Demina, and A. Harel, "Analytic solutions for neutrino momenta in decay of top quarks", *Nucl. Instrum. Meth. A* **736** (2014) 169, doi:10.1016/j.nima.2013.10.039, arXiv:1305.1878.

[56] D. O. Regina Demina, Amnon Harel, "Reconstructing top quark-antiquark events with one lost jet", arXiv:1310.3263.

[57] "LHCTopWG Summary Plots". https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit. Accessed: 2010-09-30.

[58] "LHCTopWG Summary Plots". https://twiki.cern.ch/twiki/bin/view/CMSPublic/%20PhysicsResultsEWK. Accessed: 2010-09-30.

[59] CMS Collaboration, "CMS Luminosity Measurement for the 2016 Data Taking Period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, (2017).

[60] "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions atÂ£ sÂ£=Â£8Â£TeV", *Journal of Instrumentation* **10** (2015), no. 06, P06005.

[61] T. C. collaboration, "Performance of CMS muon reconstruction in pp collision events at Â£s = 7 TeV", *Journal of Instrumentation* **7** (2012), no. 10, P10002.

[62] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in

pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[63] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, (2016).

[64] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of tt at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, (2016).

[65] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[66] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

[67] M. Czakon, D. Heymes, and A. Mitov, "fastNLO tables for NNLO top-quark pair differential distributions", arXiv:1704.08551.

[68] "Peter Uwer, private communication during TOP Readiness Meeting on April 28, 2017.".

[69] J. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra",. Comments: Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report.

[70] Q.-H. Cao, S.-L. Chen, and Y. Liu, "Probing Higgs width and top quark Yukawa coupling from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ productions", *Phys. Rev. D* **95** (Mar, 2017) 053004, doi:10.1103/PhysRevD.95.053004.

[71] USCMS, ECAL/HCAL Collaboration, "Search for standard model production of four top quarks with same-sign and multilepton final states in protonâĂŞproton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C78** (2018) 78–140.