Search for vector-like T quarks using events with oppositely-charged lepton pairs and jets in proton-proton collisions at centre-of-mass energy of 13 TeV with CMS detector

by

Dalath Rachitha Asanga Mendis

B.Sc, University of Kelaniya, Sri Lanka, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Abstract

A search is performed for heavy vector-like top quark partner (T) with electric charge +2/3 by using proton-proton collision events from Large Hadron Collider (LHC) at a centre-of-mass energy of 13 TeV. The data corresponds to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ collected by the Compact Muon Solenoid (CMS) experiment during 2016.

The production of this new hypothetical particle is assumed to be in pairs and strong interaction is responsible for such production mechanism. T quarks can decay to various combinations of third generation quarks and standard model bosons: $\mathrm{T} \rightarrow \mathrm{bW}$, tZ, or tH, and hence the final states consist of pair of opposite-sign leptons consistent with coming from a Z boson and jets. No significant excess has been observed and hence 95% CL upper limits are obtained on $\mathrm{T\overline{T}}$ production cross section by assuming different branching ratios. T quark mass values below 1280 GeV are excluded in case of 100% branching fraction for $\mathrm{T} \rightarrow \mathrm{tZ}$.

Search for vector-like T quarks using events with oppositely-charged lepton pairs and jets in proton-proton collisions at centre-of-mass energy of 13 TeV with CMS detector

by

Dalath Rachitha Asanga Mendis

B.Sc, University of Kelaniya, Sri Lanka, 2012

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Andrew Ivanov

Copyright

Abstract

A search is performed for heavy vector-like top quark partner (T) with electric charge +2/3 by using proton-proton collision events from Large Hadron Collider (LHC) at a centre-of-mass energy of 13 TeV. The data corresponds to an integrated luminosity of $35.9\,\text{fb}^{-1}$ collected by the Compact Muon Solenoid (CMS) experiment during 2016.

The production of this new hypothetical particle is assumed to be in pairs and strong interaction is responsible for such production mechanism. T quarks can decay to various combinations of third generation quarks and standard model bosons: $\mathrm{T} \rightarrow \mathrm{bW}$, tZ, or tH, and hence the final states consist of pair of opposite-sign leptons consistent with coming from a Z boson and jets. No significant excess has been observed and hence 95% CL upper limits are obtained on $\mathrm{T\overline{T}}$ production cross section by assuming different branching ratios. T quark mass values below 1280 GeV are excluded in case of 100% branching fraction for $\mathrm{T} \rightarrow \mathrm{tZ}$.

Table of Contents

List of Figures

List of Tables

Acknowledgments

There are many people, with whom I have interacted during the five years of my graduate studies.

First and foremost, I would like to thank my Ph.D. adviser, Dr. Andrew Ivanov for his guidance, motivation, and support throughout my research work. This dissertation wouldn't have been possible without his valuable advice and support.

I would also like to thank Dr. Sadia Khalil, Dr. Devdatta Majumder, Dr. Igor Vorobiev, and Tyler Mitchell, for their support and guidance throughout these years.

I am also grateful to my Ph. D advising committee, Dr. Yuri Maravin, Dr. Tim Bolton, Dr. Tatiana Firsova, and Dr. Zayd Leseman for their valuable support and advice.

I would also like to thank the rest of the HEP group members and faculty of the Physics Department for their support during my academic career as a Ph. D. student. I am also thankful to the supporting staff of the Physics Department, Peggy Matthews, Pamela Anderson, Kim Coy, and Kimberly Elliott for their support during the last five years.

Finally, I would like to give my deepest gratitude to my family members for their encouragement and understanding throughout these years.

Dedication

I would like to dedicate this thesis to my mother, Vijaya Kumari Mendis , my brother, Kasun Chathuranga Mendis, and my wife, Lahiru Namya Piyasundara.

Chapter 1

Introduction

The universe as we know it today has been evolving continuously since its birth. Many scientists have tried to understand the basic concepts of evolution of the universe and thereby to answer any mysteries lies within. Many remarkable theories were developed in the beginning of the 20th century including *Quantum Mechanics*, which ultimately lead to the formalism of *Standard model* (SM) of particle physics, and which is now a complete theory with the discovery of Higgs boson in 2012 by the ATLAS and CMS experiments.

From the beginning of the 20th century, many fundamental questions were unanswered and yet to be resolved. One of them was about basic building blocks of matter. With the discovery of Quantum Mechanics in 1920s and with many theoretical and experimental discoveries by physicists during the later of that era gave insightful ideas, from which the SM was born. It describes the fundamental structure of matter. With the experimental confirmation of Higgs boson in 2012 by the ATLAS and CMS experiments, SM is now a complete theory.

SM successfully explained almost all of the theoretical predictions and experimental observations. But the observed mass of the Higgs boson being at the electroweak scale, suggested that there should be some new physical phenomena involved.

One of the common problems, which came up with the discovery of the Higgs boson, is the so-called *Mass Hierarchy problem*. The quantum corrections to the Higgs mass should make the mass incredibly big, but the observed mass is at the electroweak scale. The only way to explain this is by fine tuning the Higgs mass, which is unnatural. So there should be some new physical phenomena at the TeV scale, which ultimately cancel out the quantum corrections for the Higgs mass. Scientists came up with new theories called *Physics beyond standard model*, to keep the mass of the Higgs boson at its observed value.

These new ideas include introducing new particles such that fine tuning of the bare mass of the Higgs boson could be avoided. Some of these new theoretical extensions of the SM include *supersymmetric partners*, where these new particles are considered to be bosonic partners of the SM quarks. Some other models predict so-called *fermionic partners*. If these new hypothetical particles exist, they should have masses around the TeV scale such that they will cancel out the quantum corrections for the Higgs boson mass to be at the electroweak scale naturally. The *Large Hadron Collider* (LHC) was operating at a centre-of-mass energy of $\sqrt{s} = 13$ TeV since 2015 and search for these new particles is possible by using that data.

The purpose of this thesis is to present a search for one such fermionic resonance called vector-like T quarks by using LHC collision data. At 95% confidence level, the vector-like quarks with masses below 900 GeV are already excluded by the analyses using Run 1 (2010 -2012) LHC data that corresponds to an integrated luminosity approximately 5 fb^{-1}. The main goal of this analysis is to expand the search by using the full 2016 data set that corresponds to 35.9 fb^{-1}.

Chapter 2

Theoretical Background

2.1 The Standard Model

The standard model (SM) contains all the elementary particles discovered up-to-date and explains the interactions between them within the framework of the quantum field theory. The development of the SM started in the 1960s. The main contributions came from the electromagnetic and weak interaction unification by S. Glashow in 1961 and later in 1967 when S. Weinberg and A. Salam incorporated the Higgs mechanism with Glashow's electroweak interaction [2]. The capacity of explaining the experimental observations obtained in the high energy physics experiments by the SM was extraordinary. The remarkable milestone of the SM was achieved in 2012 when both CMS and ATLAS Collaborations [3;4] announced the discovery of the Higgs boson, the missing piece of the SM, which predicted to be responsible for the observed masses of fermions and gauge bosons. All the particles in the SM will be discussed briefly in this chapter with their interactions, accompanied by some mathematical relations. Detailed explanation about the SM can be found in references [3;4]. Natural units, $c = \hbar = 1$ will be used throughout the thesis unless stated otherwise.

Particles that are included in the SM can be divided into two subgroups, namely: *fermions* and *bosons*. The bosons are particles, which has zero or integer spins, and obey the Bose-Einstein statistics [5]. There are four force-carrying gauge bosons (vector bosons), which act as the force mediators that correspond to four fundamental forces of nature. The strong force, which is dominant at short range, is mediated by the *gluon (g)*, and is responsible for binding quarks together to form hadrons. The weak force, which is responsible for nuclear β-decays, is mediated by the Z (neutral weak boson) and $W^{\pm}$ (charged weak bosons). The electromagnetic force, which is responsible for the formation of atoms, is mediated by the *photon* (γ). All of these vector bosons $g, Z, W^{\pm}$ and γ have spin 1. The final hypothetical gauge boson; *graviton* has spin 2 and it is not discovered, thus cannot be incorporated in the standard model yet. The *Higgs boson* is the final particle that belongs to the group of bosons with spin 0. It is responsible for the mechanism, by which the rest of the particles in the SM acquire their masses[6].

Fermions are subdivided into *leptons* and *quarks*. The leptons consist of six spin 1/2 particles; electron (e), muon (μ) and tau (τ) with electric charge +1e and their neutrinos, electron-, muon-, tau-neutrino ($\upsilon_{e,\mu,\tau}$) with no-charge. The quarks also consist of six spin 1/2 particles; up (u), charm (c) and top (t) with electric charge +2e/3 and down (d), strange (s) and bottom (b) with electric charge -1e/3. All of the fermions undergo weak interaction. Only charged particles undergo electromagnetic interaction while only quarks undergo strong interaction. All fermions are divided into three generations. The first generation consists of e, υ_e, u and d. The second generation consists of μ, υ_μ, c and s, while the third generation consists of τ, υ_τ, t and b. There are also particles so-called *anti-particles*, which have opposite quantum numbers that of fermions but have the same mass. This information is summarized in Fig. 2.1.

In 1928 Paul Dirac derived the relativistic wave equation for all spin 1/2 particles, which is given by Eq. 2.1. This was derived by considering the *Lagrangian* ($\mathcal{L}$), which describes the dynamics of the quantum fields. [1]

$$i\gamma^{\mu}\partial_{\mu}\psi - m\psi = 0 \tag{2.1}$$

where, $\mu = 0, ..., 3$ and ∂_μ is the 4-gradient with $\partial_0 = \partial_t$. ψ is the Dirac spinor and

[1] More detailed description about the Lagrangian ($\mathcal{L}$) of the SM can be found in Ref.[7].

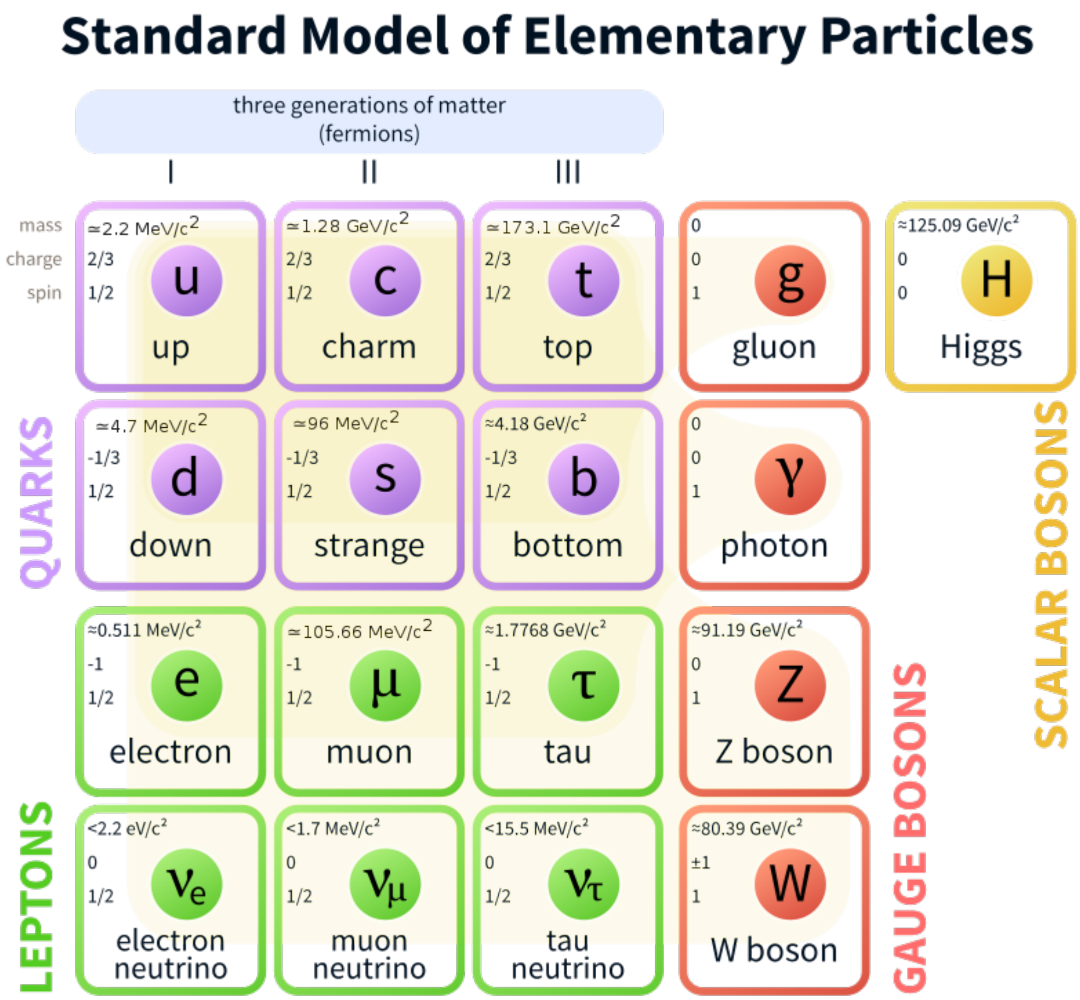


Figure 2.1: *The standard model*[2].

γ^μ are the gamma matrices taken from Pauli matrices and from 2×2 identity matrix (I_2), which are given by Eq. 2.2[8].

$$\gamma^0 = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \gamma^1 = \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix}, \gamma^2 = \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix}, \gamma^3 = \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix} \quad (2.2)$$

The SM is described by the Lie group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where subgroup $SU(3)_C$ represents the strong interaction and $SU(2)_L \times U(1)_Y$ represents the electroweak interaction. These interactions are discussed in detail in Sections 2.1.1 and 2.1.2.

2.1.1 The Strong Interaction

The strong interaction, which is responsible for *strong nuclear force (color force)*, is one of the four fundamental forces. It is approximately 137 times greater than the electromagnetic force at a distance of about 1 fm. Amongst all the other particles in the SM, quarks are the only ones that interact via the strong interaction. The strong force is mediated through *gluons*. The quarks and gluons have a property called *color* and these colored charged particles interact via exchanging gluons. The field theory, which describes this interaction, is known as *Quantum Chromodynamics* (QCD). Quarks can occupy only three color states: red, blue or green and anti-quarks can only occupy anti-red, anti-blue or anti-green[2]. The $SU(3)_C$ gauge group, which is formed by eight generators, explains that quarks can only transform as triplets. The eight matrices, which represent these generators, are known as Gell-Mann matrices T^a [9;10]. The QCD Lagrangian can then be written, as in Eq. 2.3, with the eight colored gluon fields G^μ_a given by Eq. 2.4, where $a = 1, ..., 8$ [6].

$$\mathcal{L}_{QCD} = \bar{q}_i i\gamma_\mu(\partial^\mu - ig_S T^a G^\mu_a)q_i - \frac{1}{4}G_{\mu\nu}G^{\mu\nu} \tag{2.3}$$

$$G^{\mu\nu} \equiv (\partial^\mu G^\nu_a - \partial^\nu G^\mu_a + g_S f_{abc} G^\mu_b G^\nu_c)T^a \tag{2.4}$$

Here, q_i represent the quark fields with contributions coming from all quark flavors, g_S is the coupling constant for strong interaction and f_{abc} are the structure constants representing the $SU(3)$ group [6].

The strong coupling constant g_S is an important parameter when considering the energy scale of the physical process. g_S has an inverse relationship to the *energy scale* ($g_S \propto 1/\mathrm{E}$) and hence at low energies g_S becomes large [6]. This leads to a strong binding force between quarks at low energies, which makes perturbative calculations impossible. This phenomenon is commonly known as *quark confinement*. The quarks are confined into groups of two or three quarks, which are known as *hadrons*, such that these combinations are color neutral. Only

[2]The "color charge" of quarks and gluons has no relation to the visible colors of the electromagnetic spectrum and these main colors are used just to emphasize the charge of the strong force.

groups of three quarks (*baryons*) and two quark-antiquark (*mesons*) objects are possible such that combination of three different colors or two color-anti-color combinations make them color neutral[11]. As the energy of the physical process gets high, g_S becomes very small and quarks behave as free particles, where this phenomenon is known as an *asymptotic freedom*[6;12]. But such free states are never observed in nature. At high energy collider experiments such states are observable due to the phenomenon called *hadronization*. A collimated spray of light bound particles is emitted, as the quark moves forward, which is identified as a *jet*[6].

The quarks can couple to weak force through electroweak symmetry breaking. These couplings are represented by the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix. This matrix determines how the mass eigenstates of quarks x and electroweak eigenstates $x^{'}$ are mixed. It is often represented as a 3×3 unitary matrix, which operates on the charge -1e/3 quark mass eigenstates d, s and b as shown in 2.5[6;13].

$$\begin{pmatrix} d^{'} \\ s^{'} \\ b^{'} \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{ts} & V_{cb} \\ V_{td} & V_{cs} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.5}$$

The matrix elements presented in Eq. 2.5 can be determined by measuring properties of weak decays of the relevant quarks, using the unitarity of the matrix and assuming only presence of three generations of quarks. These matrix elements are shown in 2.6[6;13]. So the transformation of quarks from one flavor to another through weak interaction is possible and is described by the CKM matrix elements. The quarks will decay to the next lightest quark, where the branching fraction is determined by these matrix elements[6].

$$V_{CKM} = \begin{pmatrix} 0.9739 \; to \; 0.9751 & 0.221 \; to \; 0.221 & 0.0029 \; to \; 0.0045 \\ 0.221 \; to \; 0.227 & 0.9730 \; to \; 0.9744 & 0.039 \; to \; 0.044 \\ 0.0048 \; to \; 0.014 & 0.037 \; to \; 0.043 & 0.9990 \; to \; 0.9992 \end{pmatrix} \tag{2.6}$$

2.1.2 The Electroweak Interaction

The gauge groups $SU(2)_L$ and $U(1)_{em}$ represent the *weak* and *electromagnetic* interactions respectively. The electroweak interaction is the unified description of these two fundamental interactions. The photon (γ) represents the generator of the gauge group $U(1)_{em}$, while W^+, W^- and Z bosons represent the generators of the gauge group $SU(2)_L$. The L of the $SU(2)_L$ denotes that the charged weak interaction (mediated by $W^\pm$) only couples to fermions with left-handed chiralities[6]. The neutral weak interaction (mediated by Z^0) was also supposed to behave in the same way but is observed to be coupled to fermions with both left- and right-handed chiralities. The *electroweak unification* theory has been introduced to solve this contradiction. It suggests that the existence of a unified interaction where Z and γ bosons are a mixture of gauge bosons[6]. The symmetry group $SU(2)_L \times U(1)_Y$ describes this *unified interaction*, which is commonly known as *electroweak* interaction.

Four generators of this group correspond to vector fields B^μ and $W^\mu_{1,2,3}$. Transformation of particles under $SU(2)_L$ is described by the *weak isospin* with generators of $W^\mu_{1,2,3}$ fields using the three Pauli-matrices T_i; while transformation under $U(1)_Y$ is described by the *weak hypercharge* (Y). Electromagnetic charge Q is related to the third component of the weak isospin (T_3) and Y is given by Eq. 2.7. The observed gauge bosons of the SM emerged after breaking the symmetry of $SU(2)_L \times U(1)_Y$ group and described by Brout-Englert-Higgs mechanism, which will be discussed in detail in section 2.1.3[6].

$$Q = \frac{Y}{2} + T_3 \tag{2.7}$$

The filed strengths of vector fields B^μ and $W^\mu_{1,2,3}$ are given by Eq. 2.8 and 2.9[14].

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2 \epsilon^{abc} W^b_\mu W^c_\nu \tag{2.8}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.9}$$

Where g_2 is the coupling constant of W^μ_i and ϵ^{abc} is the structure constant of the Lie group $SU(2)$.

The physical gauge bosons W^+, W^-, Z^0 and γ are linear superpositions of W^1, W^2, W^3 and B^0 bosons given by Eq. 2.10, 2.11 and 2.12, where θ_W is the Weinberg angle.

$$W^+ = \frac{1}{\sqrt{2}}(W^1 - iW^2) \tag{2.10}$$

$$W^- = \frac{1}{\sqrt{2}}(W^1 + iW^2) \tag{2.11}$$

$$\begin{pmatrix} Z^0 \\ \gamma \end{pmatrix} = \begin{pmatrix} -sin\theta_W & cos\theta_W \\ cos\theta_W & sin\theta_W \end{pmatrix} \begin{pmatrix} W^3 \\ B^0 \end{pmatrix} \tag{2.12}$$

The dynamics of the electroweak interaction before the symmetry breaking can be represented by the Lagrangian given in Eq. 2.13:

$$\mathcal{L}_{EW} = \bar{\psi}_L i D_L \psi_L + \bar{\psi}_R i D_R \psi_R - \frac{1}{4}(B_{\mu\nu}B^{\mu\nu} + W_{\mu\nu}W^{\mu\nu}) \tag{2.13}$$

with,

$$D_L = \gamma_\mu(\partial^\mu + g_1 Y_W B^\mu - i g_W T^i W_i^\mu$$
$$D_R = \gamma_\mu(\partial^\mu + g_1 Y_W B^\mu)$$

Here, left- and right-handed SM fermion fields are denoted by ψ_L and ψ_R with g_1 being the coupling constant for gauge filed B^μ.

2.1.3 The Higgs Mechanism

The *Brout-Englert-Higgs mechanism* was introduced by theorists Robert Brout, Franois Englert and Peter Higgs to explain how some gauge bosons obtain masses. The Lagrangian of the SM is gauge invariant and it suggests that all the gauge bosons should have a zero mass. The photons and gluons immediately follow this principle, while it is experimentally proven that Z and $W^\pm$ bosons are massive particles with masses 91.1876 ± 0.0021 GeV and 80.385 ± 0.015 GeV respectively[15].

The concept of *spontaneous symmetry breaking* of the electroweak symmetry group $SU(2)_L \times U(1)_Y$ introduced by the Brout-Englert-Higgs mechanism resolved this inconsistency. It defines a new field so-called *Higgs field*, which is a complex doublet field with four real parameters associated with a potential, called *Higgs potential*[6]. The Higgs field and Higgs potential are given by Eq. 2.14 and 2.15, respectively.

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.14}$$

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.15}$$

The parameters μ^2 and λ of the Higgs potential in Eq. 2.15 are specifically chosen to be $\mu^2 < 0$ and $\lambda > 0$. Then the resultant *Higgs potential*, shown in Fig. 2.2, is symmetric under rotation in ϕ space. It has a degenerate minimum, which is not at $<\Phi> = 0$, and is known as the *vacuum expectation value* (ν) given by Eq. 2.16. Electroweak scale is defined by this parameter ν and a small fluctuation around this minimum value *spontaneously breaks* the rotational symmetry of the Higgs field.

$$\nu = \frac{|\mu|}{\sqrt{\lambda}} = 246 \text{ GeV} \tag{2.16}$$

The *Goldstone's theorem* suggests the existence of three massless bosons whenever a continuous symmetry is broken[6;17]. The Higgs field given by Eq. 2.14 has four degrees of freedom. The three Goldstone's bosons correspond to three degrees of freedom. These are *longitudinal degrees of freedom*[3] of the Z and $W^\pm$ bosons. The remaining degree of freedom suggests the existence of a new particle, which has spin 0. This new particle is known to be *Higgs boson* and it acquire, mass through parameter μ^2 of the Higgs potential given by $m = \sqrt{2}|\mu|$[6;18]. The property of the gauge invariance of the Higgs field can be used with a

[3]The spin 1 massive particles have three degrees of freedom. The one, which is parallel to the direction, where particle travels is known as the longitudinal degree of freedom. The spin 1 massless particles have only two degrees of freedom, which are transverse to each other. They do not have a longitudinal degree of freedom.

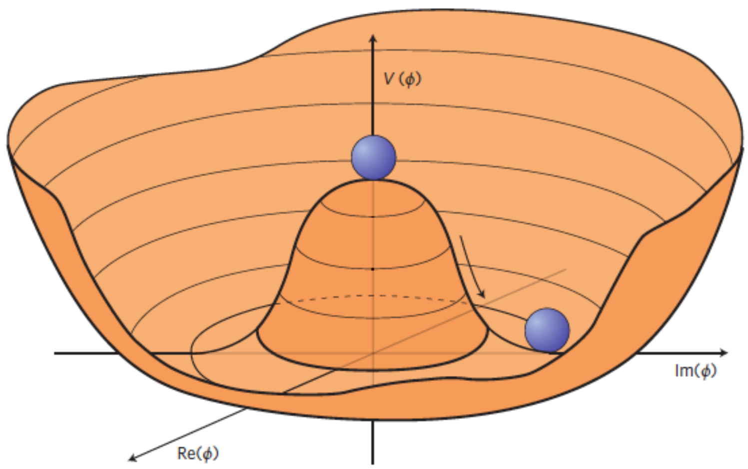


Figure 2.2: *The Higgs potential with $\mu^2 < 0$ and $\lambda > 0$* [16].

unitary gauge to obtain the mass terms of weak charge and neutral interaction as given by Eq. 2.17 where g_2 is the coupling constant of the W_i^μ filed introduced in Eq. 2.8 [6;18].

$$m_W = \frac{g_2 \nu}{2} \;,\; m_Z = \frac{g_2 \nu}{2 \; cos\theta_W} \tag{2.17}$$

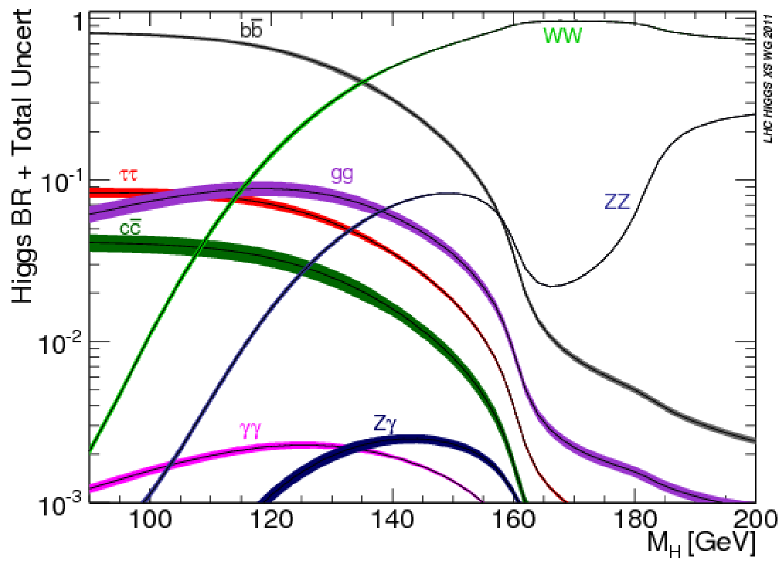
The Higgs boson can couple to all of the particles in the SM including fermions. The couplings to fermions generate their mass as given in Eq. 2.18 and commonly known as *Yukawa coupling*, where g_f denotes the coupling constant [6;18].

$$m_f = \frac{g_f \nu}{\sqrt{2}} \tag{2.18}$$

The Higgs boson has an observed mass of 125 GeV, which was measured by the ATLAS and CMS collaborations in 2012. It has many decay modes with different branching fractions, which are shown in Fig. 2.3. The dominant decays are $H \rightarrow b\bar{b}$, $H \rightarrow W^+W^-$, $H \rightarrow gg$ and $H \rightarrow \tau\bar{\tau}$.

2.1.4 Shortcomings of the Standard Model

The predictions of the SM have been tested and confirmed by many experiments. These results agree to a great precision, but there are some phenomenological observations of nature,



Higgs-boson decay mode	Branching fraction
$b\bar{b}$	0.575
W^+W^-	0.216
gg	0.0856
$\tau^+\tau^-$	0.0630
$c\bar{c}$	0.0290
Z^0Z^0	0.0267
$\gamma\gamma$	2.28×10^{-3}
γZ^0	1.55×10^{-3}
$s\bar{s}$	2.46×10^{-4}
$\mu^+\mu^-$	2.19×10^{-4}
e^+e^-	5.12×10^{-9}

Figure 2.3: *The Decay modes and branching fractions of the Higgs boson as a function of the Higgs boson mass*[19;20].

which could not be explained by using the theories of the SM. Few of these observations are listed below[6].

- **Dark Matter** - The SM only explains 5 % of the matter present in the known universe. Cosmological observations corresponding to the velocity of cosmological objects suggest that observed velocities do not completely explain the mass accumulation only by luminous matter. Therefore, about 26 % of energy should be in the form of non-luminous matter, which is known as *dark matter* (DM). But DM might interact weakly with SM fields and DM candidates cannot be explained by the fundamental particles of the SM[21;22].

- **Neutrino Masses** - SM predicts that neutrinos should be massless particles but recent experiments on neutrino oscillations have evidence that neutrinos indeed have a mass. It is not completely understood as how neutrinos acquire their mass[22].

- **Hierarchy problem** - The elementary particles in the SM acquire their masses through spontaneous symmetry breaking caused by the Higgs field. The observed mass of the Higgs boson (m_h) is given by the Eq. 2.19[23].

$$m_h^2 = m_{bare}^2 + \delta m_h^2 \tag{2.19}$$

  where δm_h represents the quantum corrections due to the presence of virtual particles. Observed masses of gauge bosons and fermions are viable by the symmetries within the SM, but there is no symmetry available for the Higgs boson to keep its observed mass value at 125 GeV[18;23].

  The quantum corrections (δm_h) can be written by using the renormalized mass as $\delta m_h^2 \propto g_f^2 \Lambda^2$[23], where g_f is the Yukawa coupling and Λ defines the maximum scale at which the SM predictions are valid (cutoff scale). The maximum value the Λ can assume to be is at Planck scale with Λ = M_p (Planck constant). If no new physical phenomena exist between the electroweak and Planck scale, the second term in Eq. 2.19 becomes huge. This has to be canceled out by choosing a very unnatural value for the bare Higgs mass (m_{bare}) to keep the observed Higgs mass at 125 GeV. Choosing this value is known as the *fine tuning* and is most unnatural. This phenomenon is known as *mass hierarchy problem.*

There are many theoretical solutions proposed to solve the shortcomings of the SM. One of the most viable solutions is where the SM is assumed to be a successful field theory, which describes the particle interactions at a low energy scale[6]. Prediction of new symmetries at a higher energy scale is the key concept of some of these models, which are commonly known as *physics beyond the Standard model*[6]. If such new physical dynamics exist, they could be observed at the collider experiments such as LHC, which is running at a higher energy scale (13 TeV) during run II.

2.2 Physics Beyond the Standard Model

Many theoretical models have been developed that predict a variety of new particles, which try to address the above-mentioned problems, including the mass hierarchy. The predicted particles follow new symmetries and have new physical interactions and dynamics. Two of these widely used models are described below.

1. **Supersymmetry**

   According to supersymmetry, each particle of the SM has a partner. These new particles differ from the SM particles only by mass but interact similarly via same forces[24]. They tend to cancel out the quantum loop corrections due to the SM particles on the mass of the Higgs boson, which helps to resolve the hierarchy problem. These supersymmetric particles have spin, which differs from their SM partner by 1/2. So SM bosons have fermionic supersymmetric partners and SM fermions have bosonic supersymmetric partners[6;24]. The supersymmetric partners of the SM particles are shown in Fig. 2.4.

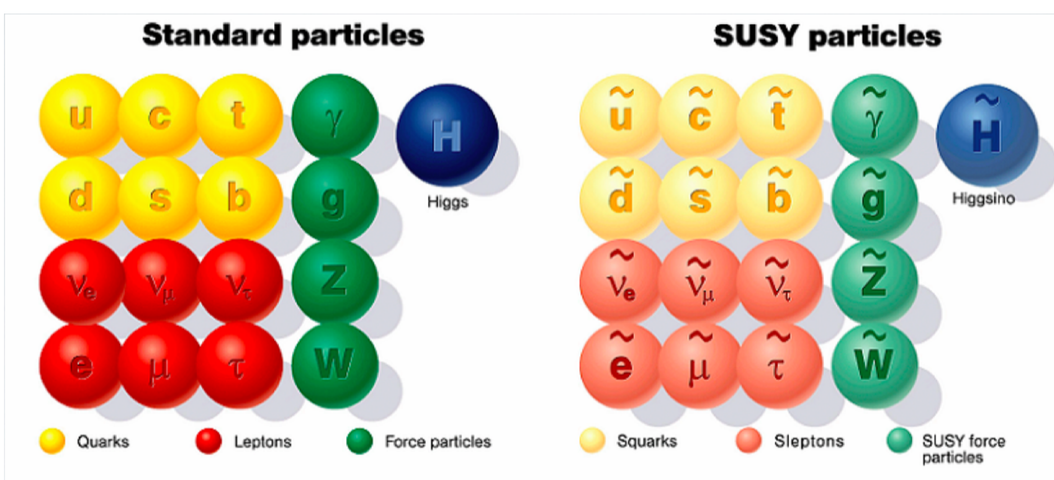


Figure 2.4: *The SM particles and their supersymmetric partners*[25].

2. **Compositeness/extra dimensions**

   According to this model, the Higgs boson (*composite Higgs*) is a bound state of new strong interaction. It predicts the existence of new particles with their masses at the TeV scale, which are excitations of the composite Higgs[6]. Each of the SM particles has a partner, which is heavier but has the same quantum numbers. Therefore, if these particles exist, they can be observed at the LHC at $\sqrt{s} = 13$ TeV. Detailed explanation about the Composite Higgs model can be found in reference[26].

2.3 Vector-like Quarks (VLQ)

The composite Higgs models predict new fermionic resonances, which emerge from a mechanism called *partial compositness*[6]. It suggests the mixing of two fields coming from SM fermions and new fermions which belongs to the strong sector. These new fermions are known to be the partner particles of the SM fermions with unique properties[6]. The quantum numbers are common between the SM fermions and their partners, but they differ by mass. This thesis is focused on searching for one such fermionic resonance called *Vector-like quarks* (VLQ).

The name *vector-like* comes from the unique feature where the transformation of the left- and right-handed fields of the VLQ behaves in the same way under the electroweak symmetry group $SU(2)_L \times U(1)_Y$[6;18]. Unlike the SM quarks, the vector property allows the VLQ to obtain their mass M through gauge invariant mass term in the form of $\mathcal{L}_M = -M\bar{\psi}\psi$, which is not constrained by the Yukawa couplings to the Higgs field[6;27]. The VLQs are also predicted in many other beyond standard model theories like beautiful mirrors[28], Randall-Sundrum[29] and little-Higgs model[30].

The VLQ have the same color charge as their SM partner and hence transform as triplets under gauge group $SU(3)_C$[27]. The composite Higgs models suggest that masses of these partners should be at the TeV scale. This thesis mainly focuses on a search for one such VLQ candidate, which will be discussed in section 2.3.1 using data collected by CMS experiment at center-of-mass energy $\sqrt{s} = 13$ TeV.

2.3.1 Properties of VLQ

The VLQ are colored charged particles with half-integer spin. They tend to undergo strong interaction under the gauge group $SU(3)_C$ with coupling strength being the same as their SM partners. The Yukawa couplings are strongest for the third generation SM quarks than any other quark generations[27]. Following on that, the theory of partial compositeness suggests that it is most likely for VLQ to mix with the third generation SM quarks and hence they are commonly called *top partners*[27]. Therefore, these third generation SM partners have the lightest masses and possibly could appear at the TeV scale. They can modify their coupling to SM gauge bosons and Higgs boson while mixing with third generation SM quarks. These new hypothetical particles are mentioned in Table 2.1, with their electric charge and possible mixing with third generation SM quarks[27]. This thesis is mainly focused on searching for *vector-like* T quarks explained in detail in the following sections.

Table 2.1: *Vector-like quarks and their decay modes.*

$Q^{ElectricCharge}$	Decay modes
$T^{\frac{2e}{3}}$	bW^+, tH, tZ
$B^{-\frac{1e}{3}}$	tW^-, bH, bZ
$X^{\frac{5e}{3}}$	tW^+
$Y^{-\frac{4e}{3}}$	bW^-

Table 2.2: *Isospin multiplets in which vector-like quarks can appear.*

Singlets	Doublets	Triplets
$T_{L,R}, B_{L,R}$	$\begin{pmatrix} X \\ T \end{pmatrix}_{L,R}, \begin{pmatrix} T \\ B \end{pmatrix}_{L,R}, \begin{pmatrix} B \\ Y \end{pmatrix}_{L,R},$	$\begin{pmatrix} X \\ T \\ B \end{pmatrix}_{L,R}, \begin{pmatrix} T \\ B \\ Y \end{pmatrix}_{L,R}$

The VLQ can only appear in seven gauge-covariant multiplets as given by Table 2.2[27]. A reasonable assumption has been made according to a theoretical model such that SM top quarks only mixes with T quarks. This mixing is represented by Eq. 2.20[31].

$$\begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix} = U^u_{L,R} \begin{pmatrix} t^0_{L,R} \\ T^0_{L,R} \end{pmatrix} = \begin{pmatrix} cos\theta^u_{L,R} & -sin\theta^u_{L,R}e^{i\phi_u} \\ sin\theta^u_{L,R}e^{-i\phi_u} & cos\theta^u_{L,R} \end{pmatrix} \begin{pmatrix} t^0_{L,R} \\ T^0_{L,R} \end{pmatrix} \tag{2.20}$$

where $U^u_{L,R}$ is a 2×2 unitary matrix with t^0 and T^0 representing the weak eigenstates. The mixing angle between two types of quarks is given by the $\theta^u_{L,R}$ with ϕ representing the phase between them[27].

The mixing angle is constrained by few important factors, such as the model in which VLQ appear, how VLQ transforms under the electroweak symmetry group and electroweak precision measurements[6]. The appearance of VLQ in possible electroweak isospin multiples given by Table 2.2, is only possible if scalar sector Lagrangian contains $SU(2)_L$ doublets.

2.3.2 Production of VLQ

According to the theoretical models, VLQ can either be *pair* or *singly* produced, which can belong to any of the isospin multiplets mentioned in Table 2.2. The electroweak interaction is responsible for single production and the production cross section depends on both the VLQ mass and electroweak coupling strength. The pair production is a dominant process at the low mass range and produced via the strong interaction. The pair production cross section only depends on the VLQ mass.

The variation of maximum cross-section (σ_{max}) as a function of VLQ mass (m_Q) is shown in Fig. 2.5 for both single and pair production of VLQ, for centre of mass energy of 13 TeV computed using PROTOS[31]. It represents:

- T quark belongs to a singlet in T$\bar{\text{b}}$j production.
- T and B quarks belong to (T B) doublet in T$\bar{\text{b}}$j, B$\bar{\text{b}}$j and T$\bar{\text{t}}$j production.
- B and Y quarks belong to (B Y) doublet in Y$\bar{\text{b}}$j and B$\bar{\text{b}}$j productions.
- X and T quarks belong to (X T) doublet in X$\bar{\text{t}}$j and T$\bar{\text{t}}$j production.

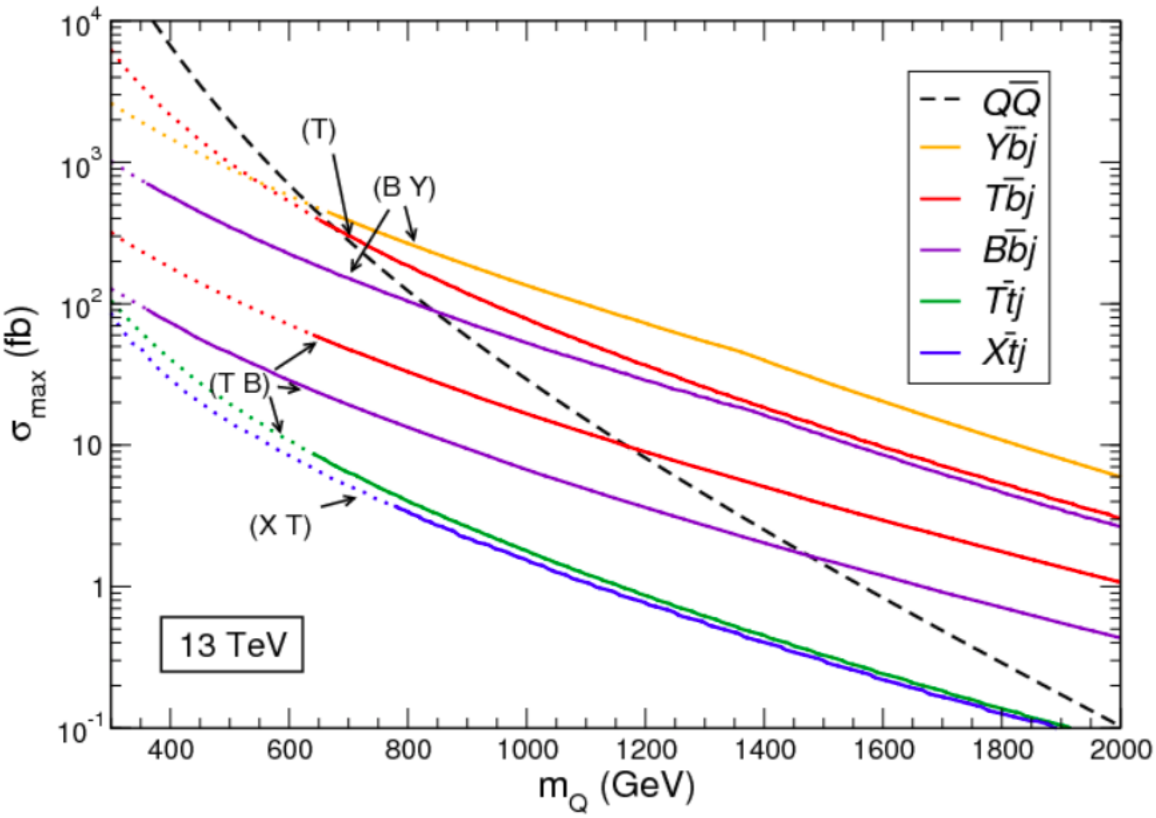


Figure 2.5: *The variation of maximum cross section (σ_{max}) as a function of VLQ mass (m_Q) for selected VLQ isospin multiplets at centre-of-mass energy of 13 TeV computed using PROTOS*[27].

2.3.2.1 Single production of vector-like T quarks

The T quark is singly produced in proton-proton collisions either via neutral (Z) or charge (W) weak interaction. The Feynman diagrams are shown in Fig. 2.6. The characteristics of the interaction include the association of either t or b quark and an extra quark jet (q') propagated towards the forward region of the detector known as a *forward jet*[6]. The coupling constants of the production of T quark are denoted by $C(tZ)$ and $C(bW)$ for the two processes respectively. These electroweak couplings and mass of the T quark determine the cross section of the single production.

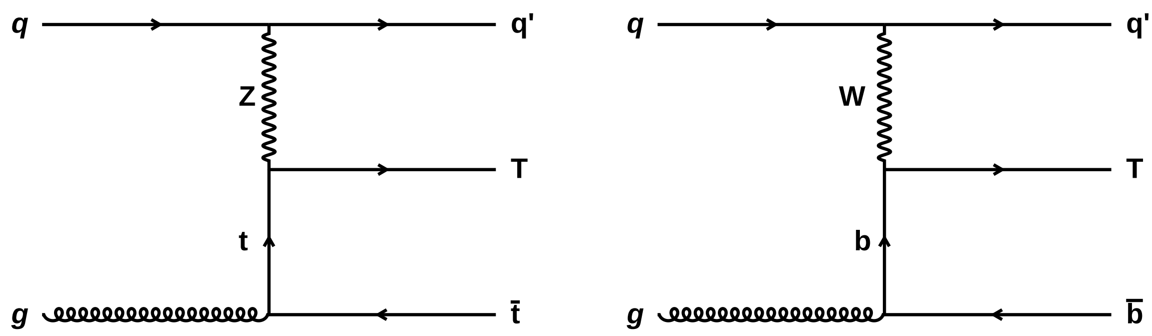


Figure 2.6: *Feynman diagrams for single production of vector-like T quarks via the neutral (left) and charged (right) weak interaction. The diagrams were produced by using ROOT framework*[32].

2.3.2.2 Pair production of vector-like T quarks

The strong interaction is responsible for the production of T quark with its antiquark, $\bar{\mathrm{T}}$. This thesis is mainly focused on pair production of vector-like T quarks. Fig. 2.7 shows the possible $\mathrm{T\bar{T}}$ production at proton-proton collisions.

The predicted cross sections for different vector-like T masses are shown in the Table 2.3 which are calculated at next-to-next-to-leading order (NNLO) using the TOP++ program (v.2.0)[1].

2.3.3 Decay modes and branching fractions of T quarks

The vector-like T quarks can decay into three final states, tZ, tH or bW. The approximate branching fractions for these decay modes are determined by using model-independent calculations. These branching fractions depend upon the isospin multiplet, in which the T quark appears[27;33]. The allowed branching fractions for the three final states are not independent and follow the relationship given by Eq. 2.21[31].

$$\mathcal{B}(T \to bW) + \mathcal{B}(T \to tZ) + \mathcal{B}(T \to tH) = 1 \tag{2.21}$$

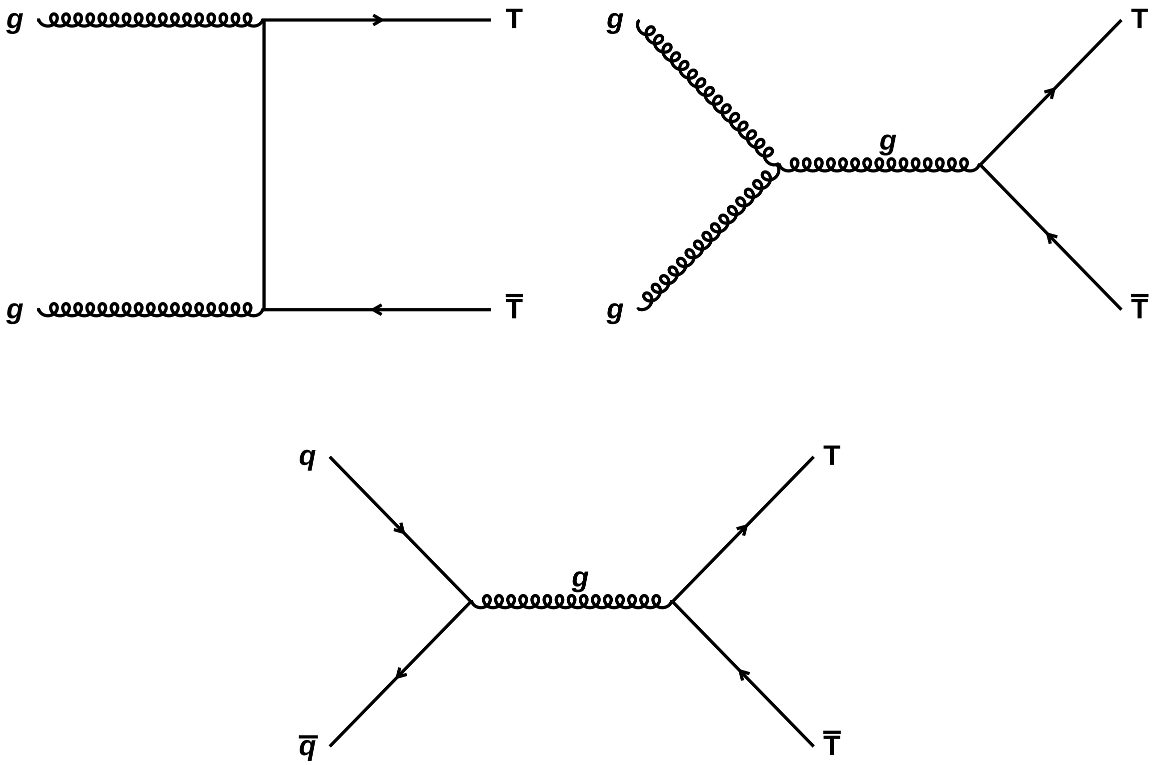


Figure 2.7: *Feynman diagrams for pair production of vector-like T quarks. The diagrams were produced by using ROOT framework*[32].

Possible branching fractions for three final states of T quark are shown in Tab. 2.4, where T quark belongs to either an isospin singlet or doublet. The branching fractions for isospin triplets are similar to that of either isospin singlet or doublet[6].

The variation of branching fractions for different T quark masses are shown in Fig. 2.9. The branching fractions for three decay modes vary slightly at lower masses and are almost constant at higher masses[31]. A full branching ratio scan will be performed for the current search according to Eq. 2.21.

2.3.4 Constraints on Pair Production of Vector-like T quarks

Many direct searches were performed for pair production of vector-like T quarks using LHC *run I* data at centre-of-mass energy $\sqrt{s} = 8$ TeV and also LHC *run II* data at centre-of-

Table 2.3: *The predicted cross sections (σ) for pair production of vector-like T quarks for different masses calculated at NNLO.*

Mass [GeV]	σ [fb]
$T\bar{T}, m_T = 800$	$196\pm^{8.5}_{8.1}$
$T\bar{T}, m_T = 900$	$90.3\pm^{4.0}_{3.8}$
$T\bar{T}, m_T = 1000$	$44.0\pm^{2.1}_{1.9}$
$T\bar{T}, m_T = 1100$	$22.4\pm^{1.1}_{1.0}$
$T\bar{T}, m_T = 1200$	$11.8\pm^{0.64}_{0.56}$
$T\bar{T}, m_T = 1300$	$6.39\pm^{0.37}_{0.32}$
$T\bar{T}, m_T = 1400$	$3.54\pm^{0.22}_{0.19}$
$T\bar{T}, m_T = 1500$	$2.00\pm^{0.14}_{0.12}$
$T\bar{T}, m_T = 1600$	$1.148\pm^{0.087}_{0.072}$
$T\bar{T}, m_T = 1700$	$0.666\pm^{0.056}_{0.045}$
$T\bar{T}, m_T = 1800$	$0.391\pm^{0.037}_{0.029}$

Table 2.4: *Approximated branching fractions for vector-like T quark for different isospin multiplet representations.*

Isospin multiplet	T → bW	T → tZ	T → tH
T Singlet	0.5	0.25	0.25
(T, B) Doublet	0	0.50	0.50
(X, T, B) Triplet	0	0.50	0.50

mass energy $\sqrt{s} = 13$ TeV by both CMS[35–37] and ATLAS[38–41] experiments. None of them have evidence for existence of vector-like T quarks and hence statistical analyses have been performed to obtain 95% confidence level (CL) upper limits on production cross section and lower mass limits corresponds to them.

The lower mass limits on VLQ are given in Fig. 2.9 for Run 1 analyses and also for few of the Run 2 analyses[42]. The strictest lower limit reported for the T quark mass for Run 1 analyses is approximately 900 GeV, while recent searches from Run 2 analyses (using full 2016 data set) ranges from 1200 - 1350 GeV depending on the different decay modes from pair production.

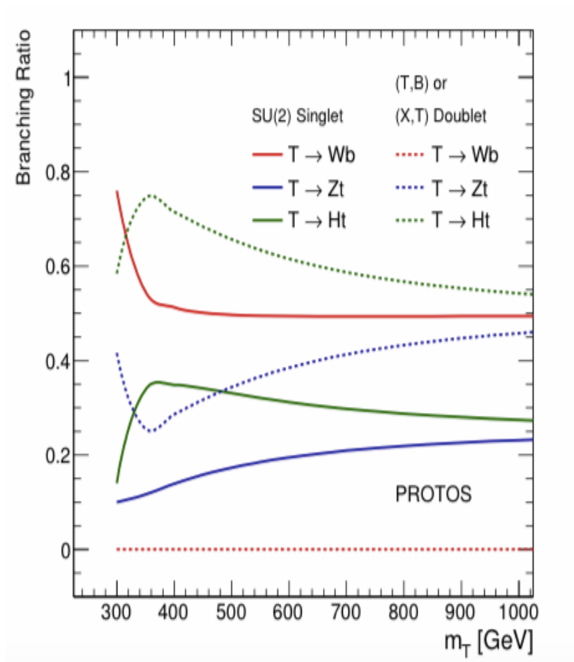


Figure 2.8: *The branching ratios of vector-like T quark as a function of its mass for different decay modes. The T quark is assumed to appear either in SU(2) singlet or doublet scenarios. The branching ratios are computed using PROTOS*[34].

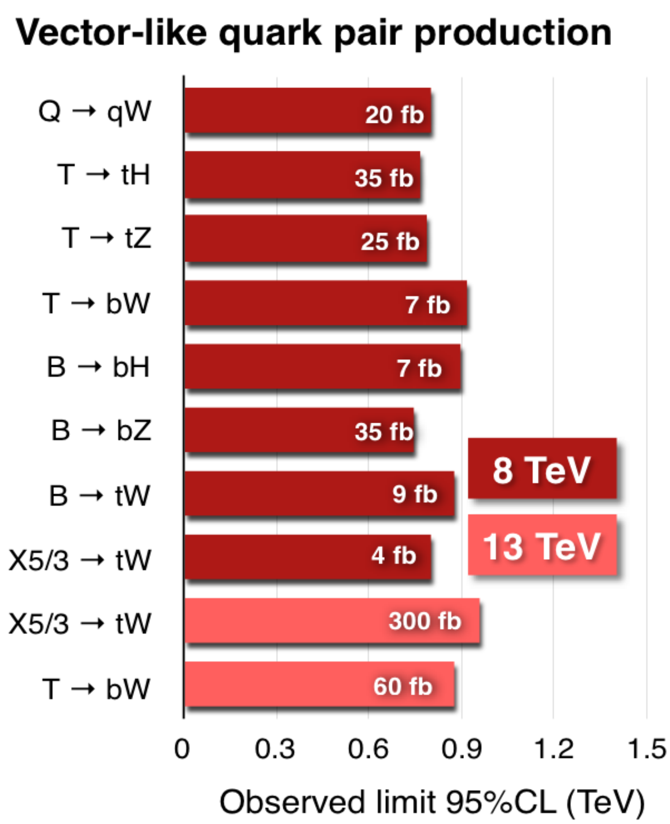


Figure 2.9: *The 95 % CL lower mass limits on VLQ from searches*[42].

Chapter 3

The Experimental Setup

The theoretical aspects of the SM should be validated by means of experimental approaches. This not only aids to improve theoretical approaches but also provides a way to suggest possible extensions to the SM. Particle accelerators play a major role in fulfilling this goal. Different particle accelerators exist, where each of them is designed uniquely to meet the goals of the experiment. All the existing accelerators can be categorized into either *linear* or *circular* accelerators or a combination of both types.

The linear accelerators use a static or dynamic electromagnetic field to accelerate particles in a straight line. These accelerated particles are then collided with another accelerated particle or with a fixed target[6]. The *Stanford Linear Collider* is one of the examples of colliders which belongs to this type[43]. The circular accelerators are widely used today as they can be used to accelerate the particles to higher collision energies. The earliest designs of circular accelerators are *cyclotrons*, where a static magnetic field is used to accelerate particles. The cyclotrons operate such that the radius of particles increases with increasing velocity[44]. A modified version of cyclotrons known as *synchrotrons* are used today, where a dynamic magnetic field has been employed, which adapts according to the momentum of the particle by keeping its radius at a constant value. The so-called *storage rings* use the same mechanism as synchrotrons but are used to keep the accelerated particles at constant energy. Particles are accelerated only when necessary to fill up any potential energy losses.

This improves the operation time of the accelerators by increasing the length of time of data taking[6;45].

Different types of particles are collided at different accelerators depending on the physical process under investigation. The *Stanford Linear Collider* and the *Large Electron-Positron Collider* (LEP) at European Organization for Nuclear Research (CERN), both use electrons and positrons for the collisions[6]. The disadvantage of using electrons is the inability to accelerate them to higher energies as electrons lose a significant amount of energy due to the synchrotron radiation. The *hadron* colliders, on the other hand, do not face this problem as they use either protons or heavy ions for collisions. Therefore, hadron colliders can be accommodated to study physical processes that occur at higher energies. The most powerful hadron collider, which is built to-date is the *Large Hadron Collider* (LHC) and is situated at CERN. It collides either proton-proton (pp) or lead-lead($PbPb$) particles as described in Sec. 3.1 in detail[6].

3.1 The Large Hadron Collider (LHC)

The LHC is the largest and the most powerful particle collider built in the world yet. It is located in CERN and collisions started in 2008. This ring-shaped machine is 27 Km in circumference and consists of superconducting magnets and different structures to accelerate the particles to boost up their energies[46]. Fig. 3.1 shows the complete accelerator complex.

The proton source used here is the hydrogen gas, which is passed through an electric field to yield protons. First, these protons are accelerated by the *Linac 2* up to 50 MeV and then injected to the *Proton Synchrotron Booster* (PSB), *Proton Synchrotron* (PS) and *Super Proton Synchrotron* (SPS) respectively, where at each point the energy of the proton beam is increased up to 1.4, 25 and 450 GeV respectively[47]. This proton beam is then transferred to the two pipes of the LHC, where in one pipe the beam circulates clockwise, while in the other, it circulates counter clockwise. The time duration to fill up each pipe is about 4 minutes and it takes about 20 minutes for the protons in the beam to reach 6.5 TeV. The protons are grouped into *bunches* with 25 ns bunch spacing and the collision of two bunches is referred to *bunch crossing*. The two beams are allowed to collide in four interaction points, where the four detectors CMS, ATLAS, ALICE and LHCb are located with the total energy at the interaction point being 13 TeV [6;47].

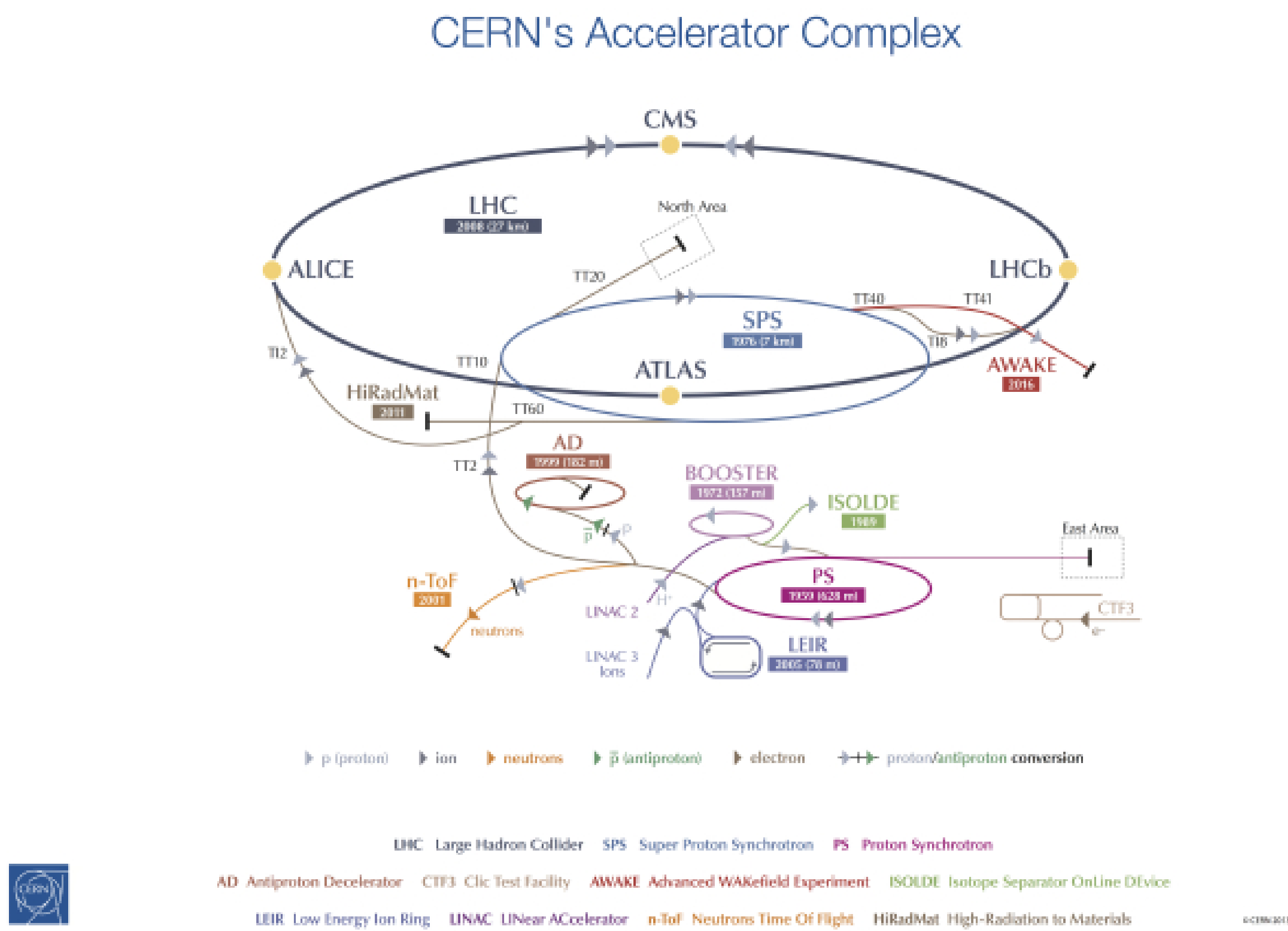


Figure 3.1: *The accelerator complex at CERN along with the interaction points for the CMS, ATLAS, ALICE and LHCb detectors*[48].

The four detectors situated at different interaction points of the LHC ring have different general purposes. The CMS (*Compact Muon Solenoid*) and ATLAS (*A Toroidal LHC ApparatuS*) detectors are generally built for studying the SM physics including the Higgs boson, search for extra dimensions (described in section 2.2) and possible candidates for dark matter (described in section 2.1.4). The technical designs and solutions and the magnet designs are different between these two detectors, while their physics goals are the same[49]. The ALICE (*A Large Ion Collider Experiment*) detector is dedicated to study the heavy ion collisions[50], while the LHCb (*Large Hadron Collider beauty*) detector is used to study B-physics[46;51].

The LHC has been operating since March 2010. The *Run 1* data were collected from 2010 to 2013. The CMS experiment recorded a total integrated luminosity of 45 pb^{-1}, 6.1 fb^{-1} and 23.3 fb^{-1} with a centre-of-mass energy of 7, 7, 8 TeV during the years 2010, 2011 and 2012 respectively. The *Run 2* started in 2015 after two years of maintenance and technical upgrading[52] and ended in 2018. The CMS experiment recorded 4.2, 40.8, 49.8 and 68.2 fb^{-1} in 2015, 2016, 2017 and 2018 respectively with a centre-of-mass energy of 13 TeV. The Fig. 3.2 shows the summary of data collected[53].

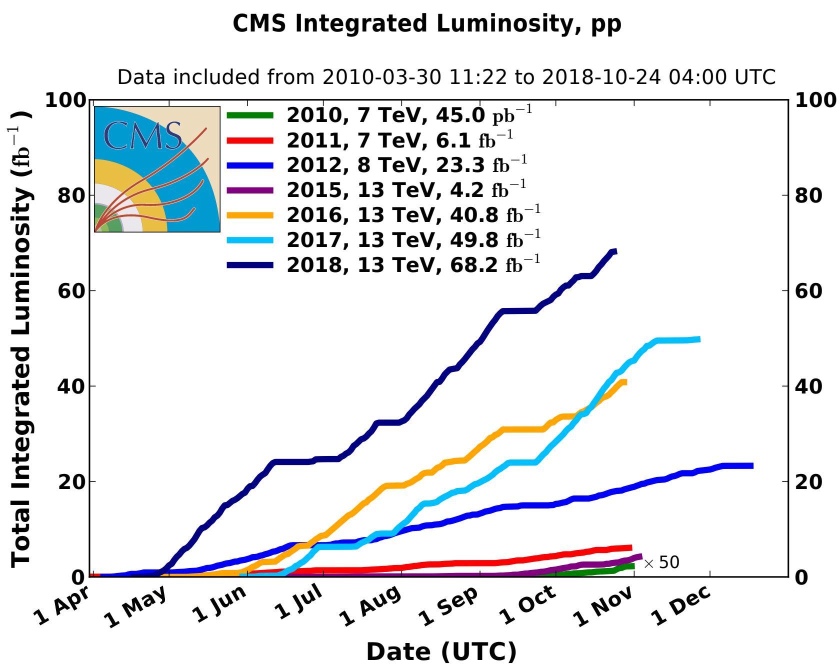


Figure 3.2: *Recorded integrated luminosity by the CMS experiment from 2010 to 2018*[53].

3.2 The CMS detector

3.2.1 General overview

The CMS detector is a general purpose detector, which sits 100 m underground in one of the four collision points. Its main objective is to study the SM physics mainly involving the Higgs boson and also to observe any new physics phenomena that LHC might reveal including the Dark matter candidates.

A huge amount of particles are produced in the CMS detector at each bunch crossing due to different physical processes including the hard scattering. The interactions which occurred by the proton-proton collision during the same bunch crossing are known as *in-time pileup* (IT pileup) interactions, while interactions that occurred during the previous or successive bunch crossing are known as *out-of-time pileup* (OOT pileup) interactions[6].

The detector is composed of many layers of sub-detectors, where each of them is capable of identifying final state particles such as electrons, muons, photons, charged and neutral hadrons from the proton-proton interaction[46]. Fig. 3.3 illustrates the internal structure of the CMS detector, which covers the total solid angle. Starting from the beam interaction point, the innermost sub-detector is the silicon pixel and strip tracker followed by the electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL), superconducting solenoid and muon chambers. These sub-detectors will be explained in detail later in this chapter. The vector-like T quark analysis uses electrons, muons and hadron-jets reconstructed with the CMS detector.

The coordinates of the CMS detector are aligned in a way such that x-axis lies to the center of the LHC ring, y-axis lies directly upwards perpendicular to collision point and z-axis lies parallel the beam to west as shown in Fig. 3.4. The $x-y$ plane is known as the *transverse plane*. The polar angle θ is defined in the $r-Z$ plane, where r is the radial coordinate. The *pseudorapidity* is given by $\eta = -ln\ tan(\theta/2)$[46]. Each sub-detector is subdivided into two regions depending on this η value. The *barrel* region is defined with $|\eta| \lesssim 1$ and the *endcap* region with $|\eta| \gtrsim 1$. The azimuthal angle ϕ is defined in the $x-y$ plane. Transverse

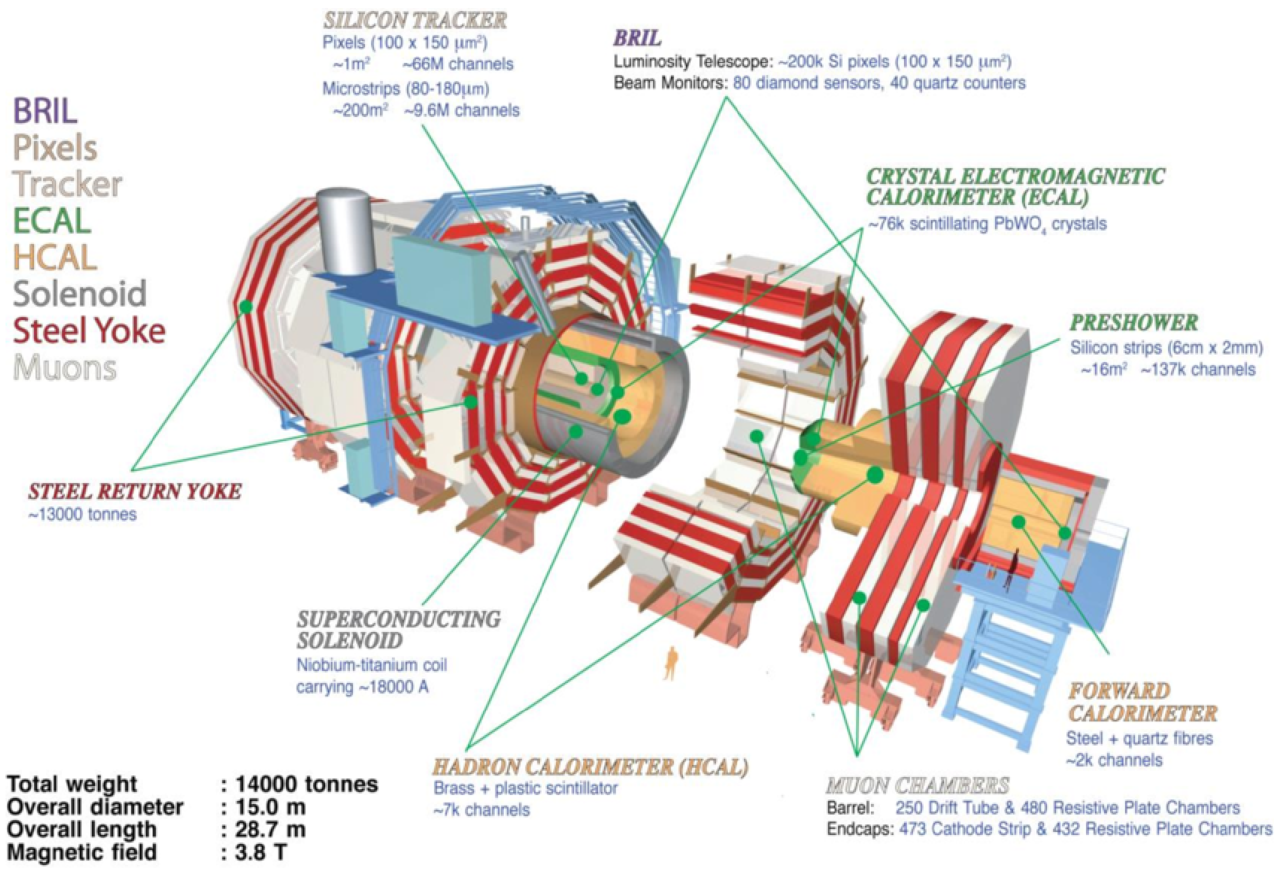


Figure 3.3: *The internal structure of the CMS detector consisting of silicon pixel and strip tracker followed by the electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL), superconducting solenoid and muon chambers*[54].

momentum (p_T) and energy (E_T) are then calculated as $E_T = E\ sin\ \theta$ and $p_T = p\ sin\ \theta$, where E and p denotes total energy and momentum[55]. The angular separation between two particles is usually denoted by ΔR, where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$[46].

3.2.2 The Superconducting Magnet

The central component of the CMS detector is the superconducting solenoid magnet, which is 12.5 m long, 6.3 m in diameter and produces a uniform magnetic field of 3.8 T which is 100,000 times stronger than the Earth's magnetic field[46;57]. It is made of niobium-titanium alloy[46]. A comparative view is shown in Fig. 3.5. The charged particles bend inside this strong magnetic field and their paths can be traced to compute the momentum of particles.

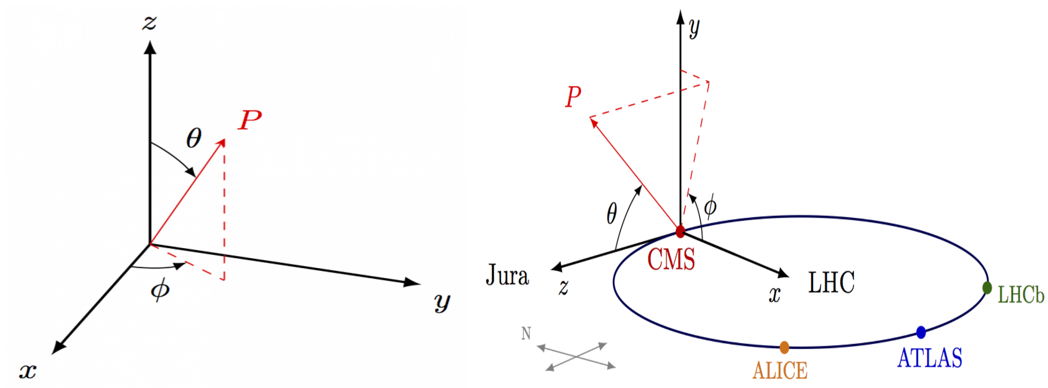


Figure 3.4: *The spherical coordinates for the CMS detector*[56].

Having a strong magnetic field allows the tracker and muon detectors to have highly precise position measurement and also the ability to detect highly energetic charged particles[57;58].

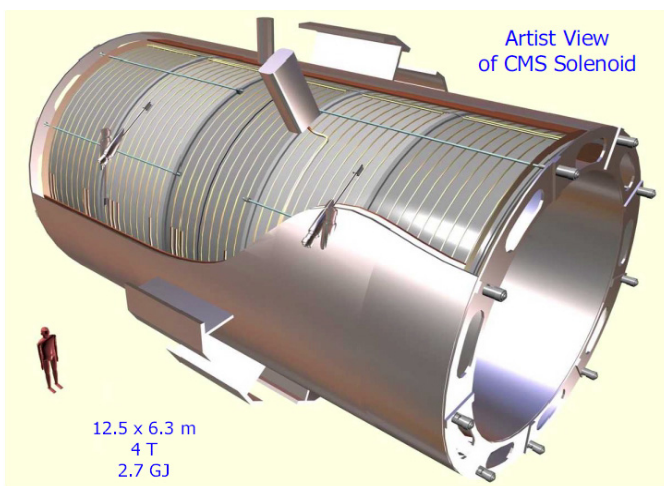


Figure 3.5: *The Superconducting solenoid magnet of the CMS detector*[57].

3.2.3 The Inner Tracking System

The CMS tracking system is designed to identify the tracks of the charged particles; electrons, muons and charged hadrons. This is the closest sub-detector to the beam pipe and nominal interaction point, which is 2.6 m in diameter, 5.8 m long and consists of 25,000 silicon strip sensors[59]. With the know magnetic field of 3.8 T, the transverse momentum of the particles can be calculated using the curvature of their tracks. It also determines the interaction vertices namely; *primary* and *secondary* vertices. Primary vertex is the position of the hard scattering process and the secondary vertex is the position where intermediate particles decay[6].

The detection of the tracks of the charged particles is done using semiconductors. The entire tracker consists of a large number of modules. Each module comprises small units of silicon semiconductors[6]. A charged particle knocks out electrons from the material, which creates a net positive charge on the pixel, which can be amplified and measured as tracks. The silicon returns to its neutral state by passing an electric current[60].

The tracker can be subdivided into two systems. The one which is closest to the proton-proton interaction point is known as the *pixel detector* and the one which is further away from the beam axis and that encloses the pixel detector is known as the *strip tracker*. The main difference between these two subsystems is the size of the small silicon semiconductors. The size increases as moving away from the beam axis. The entire tracking system covers up to $\eta < 2.5$[60].

The pixel detector - This is the innermost layer, which is closest to the nominal interaction point and hence has an immense track density. Therefore, this subsystem has a low occupancy (0.002 - 0.02 %) and a high granularity to accommodate the high number of tracks. It is composed of 66 million silicon pixels in total; each with a size of 100×150 μm^2 distributed over 1,440 modules[6]. These modules are arranged in three layers at radial distance $r = 4.4, 7.3$ and 10.2 cm in barrel region and two disk situated at each side of the detector with 6 cm < r < 16 cm and $|z| = 34.5, 46.5$ cm in endcap region[6;61]. A fourth pixel layer has been added during the technical upgrade in 2017 to increase the accuracy in reconstruction of tracks[62].

The strip detector - This is the outermost layer covering the pixel detector with low track density. The size of the strips vary from 10 $cm \times 80$ μm to 25 $cm \times 180$ μm closer to the beam axis and outermost point respectively. 9.6 million silicon strips are placed inside the detector[6]. The detector is comprised with *tracker inner barrel* (TIB) with four layers, *tracker outer barrel* (TOB) with 6 layers ranging from $20\ cm < r < 116\ cm$ in barrel region and *tracker inner disk* (TID) with three disks, *tracker endcap* (TEC) with nine disk ranging from 60 $cm < |z| < 280$ cm in the endcap region. A schematic view is shown in Fig. 3.6.

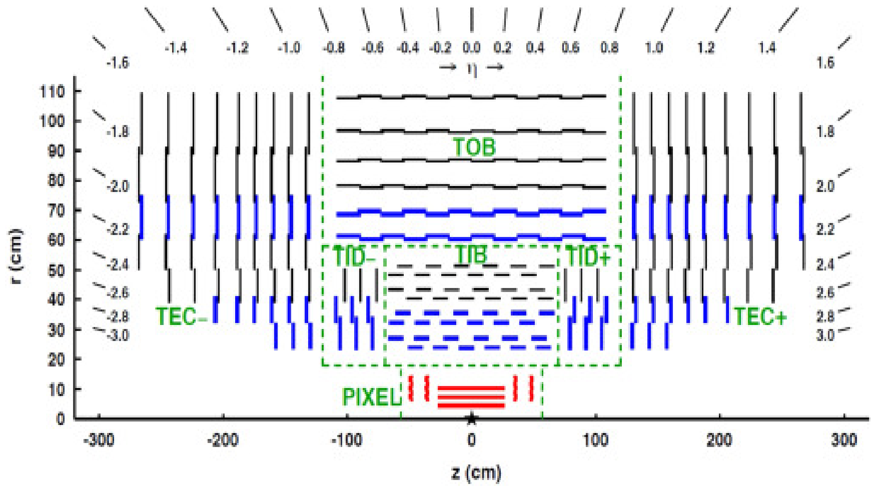


Figure 3.6: *A schematic view of the CMS inner tracking system including the pixel and strip detector*[63].

3.2.4 Electromagnetic Calorimeter (ECAL)

ECAL is the next layer of sub-detectors, which encloses the inner tracker system and spans between $r = 129$ to 177 cm. The main purpose of the ECAL is to detect and measure the energy of the electrons and photons emitted from the particle interactions. The ECAL consists of $\sim 68,000$ crystals that are made of lead tungstate ($PbWO_4$), where each crystal has a length of 23 cm[6]. The stopping power and scintillation properties of $PbWO_4$ make them the best material to use in ECAL. They have a radiation length of 0.89 cm and are capable of capturing the bulk of the energy of particles[6].

ECAL is divided into two regions namely; ECAL barrel (EB) with 61,200 crystals and spans the area $|\eta| < 1.479$ and ECAL endcap (EE) with 7324 crystals and spans the area $1.653 < |\eta| < 3.0$. The EB and EE use silicon avalanche photodiodes (APD) and vacuum photo-triodes respectively to convert the collected scintillation light to energy. In order to identify the decays of $\pi_0 \rightarrow \gamma\gamma$, there is an addition detector ranging from $1.566 < |\eta| < 2.6$ called the *preshower detector* (PS) with a thickness of 20 cm[6]. It enhances the ability to detect photons. A schematic diagram of the ECAL is shown in Fig. 3.7. The ECAL energy resolution is given by Eq. 3.1, where three contributions represent stochastic, noise and constant terms respectively[6;64].

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + \left(\frac{12\%}{E/\mathrm{GeV}}\right)^2 + (0.3\%)^2 \tag{3.1}$$

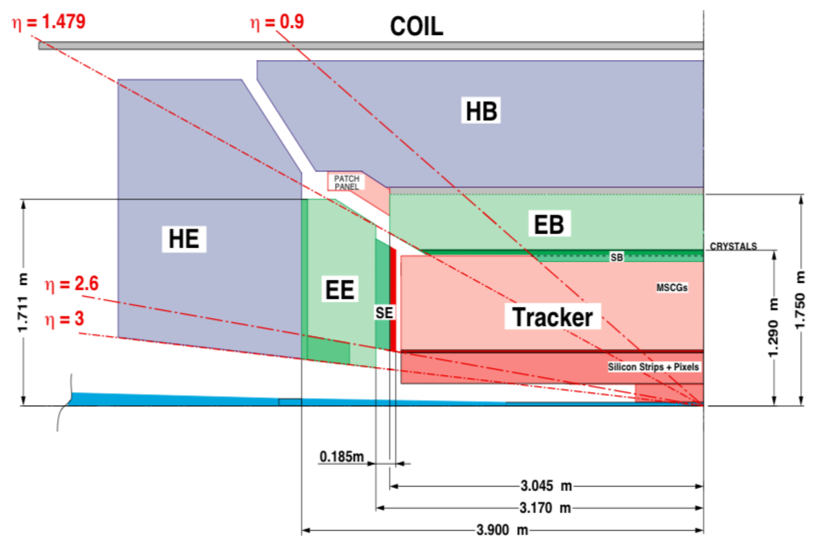


Figure 3.7: *A schematic view of the electromagnetic calorimeter*[65].

3.2.5 Hadronic Calorimeter (HCAL)

The main purpose of the HCAL is to measure the energy of the hadronic particles emerged from the proton-proton interactions. This sub-detector is divided into four sub-components; the hadron barrel (HB) which spans $|\eta| < 1.4$, the hadron endcap (HE) which spans $1.3 < |\eta| < 3.0$, the hadron forward (HF) which spans $3.0 < |\eta| < 5.2$ and the hadron outside (HO) which spans $|\eta| < 1.4$. Both HB and HE are within the solenoid magnet, while HO is situated outside next to the magnet[6;46].

The HB, HE and HO are composed of large segments of brass material and plastic scintillator tiles[66]. Brass material acts as the absorber, while the produced light from the scintillator is read by *hybrid photodiodes* (HPD). Optical fibers are used to connect the plastic tiles and HPDs. Not all the hadronic particles are absorbed by EB and HB as they do not have large enough absorption lengths and hence the HO is placed outside the solenoid magnet to measure any leakage[6;46]. The HF is placed 11.2 m away from the nominal interaction point closer to the beam axis and hence exposed to a large flux of particles. For this reason, HF uses steel as the absorber and quartz fiber as the scintillating material, which are more radiation hard materials[66]. A schematic diagram of the HCAL is shown in Fig. 3.8. The HCAL energy resolution for particles with momenta between 2 - 350 GeV is given by Eq. 3.2[67].

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{87.7\%}{\sqrt{E/\text{GeV}}}\right)^2 + (7.4\%)^2 \tag{3.2}$$

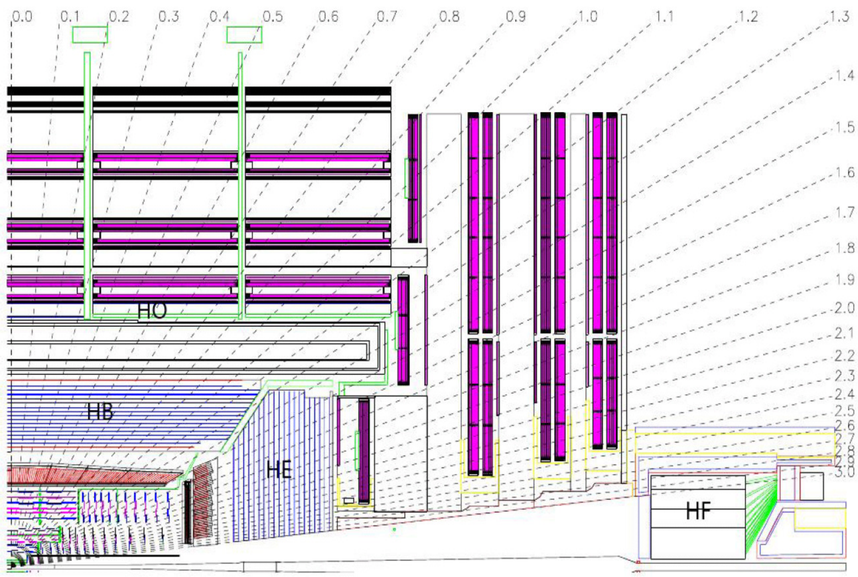


Figure 3.8: *A schematic view of the hadronic calorimeter[68].*

3.2.6 Muon System

The muon system is the outermost sub-detector, which is used to identify muon tracks. Unlike any other particles, muons are not stopped and absorbed by any of the calorimeters. Muons have higher decay lengths than any other particle and hence undetectable within the dimensions of the calorimeters. Therefore, muons are reconstructed using a track based reconstruction method by using information from both tracker and muon chambers. The muon system consists of three sub-systems; drift tube chambers (DTs), resistive plate chambers (RPCs) and cathode strip chambers (CSCs)[69]. These chambers are filled with gases, which is ionized by muons. This ionization can be used to reconstruct the muon path. These chambers are also connected in groups to form so-called *muon stations*. Fig. 3.9 shows a schematic diagram of the muon system.

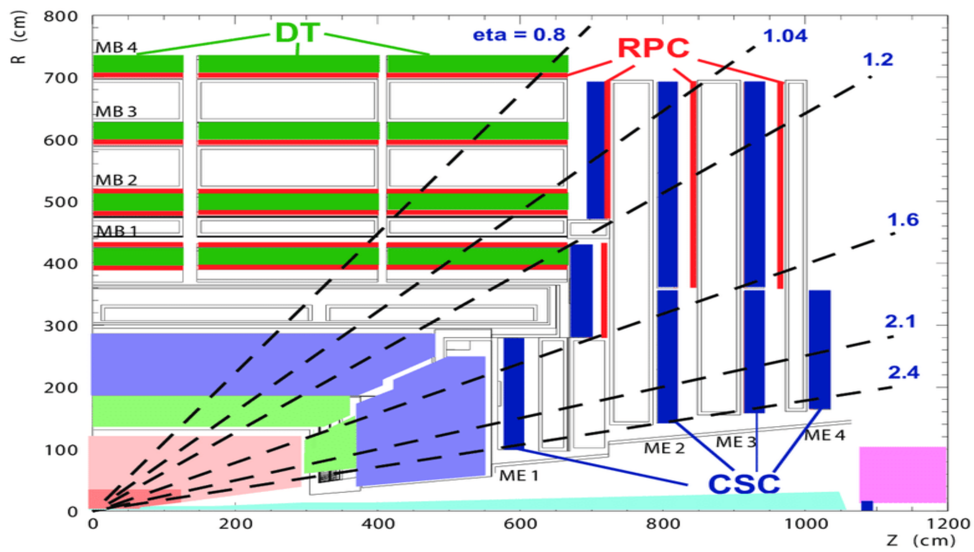


Figure 3.9: *A schematic view of the muon system*[70].

Drift Tube Chambers (DTs) - The drift tubes are filled with a gas mixture of Ar and CO_2 in the ratio of 85:15 respectively and span $|\eta| < 1.2$ region[46]. They are arranged in three groups of four layers[71]. The middle four chambers determine the position of the muon in Z direction, while the two outer layers with eights chambers determine the position of the muon in $r - \phi$ plane[46].

Resistive Plate Chambers (RPCs) - The RPCs are filled with a gas mixture of $C_2H_2F_4$ and i-C_2H_{10} in the ratio of 95:5 respectively between two parallel plates with readout strips in the middle[46]. They have a much faster timing response to the muon trigger system with resolution around 1ns[72]. The RPCs span up to $|\eta| < 1.6$. Six layers of RPCs are installed with two layers at either side of the first two drift tube stations and one layer in each of the last two drift tube stations in barrel region. In endcap region, a total of three layers of RPCs are installed with one layer at each of the first three muon stations [46].

Cathode Strip Chambers (CSCs) - The CSCs take the shape of a trapezoidal with 6 layers of 80 copper strips cathode panels and several groups of 16 anode wires oriented perpendicular to the cathode strips. The gaps are filled with a gas mixture of Ar, CO_2 and CF_4 in the ratio of 40:50:10 respectively [71]. The 4 CSP stations span $0.9 < |\eta| < 2.4$ in the endcap region. The anodes and cathodes are used to determine the r and ϕ positions of the muons respectively [6].

3.2.7 Trigger System

At every 25 ns time interval, data is produced from proton-proton collision by the LHC at a very large rate (> 1 GHz) [73]. Most of the data produced does not contain any interesting physical interactions. A decision has to be made as what events to be kept and stored at the expense of storage space. Very sophisticated *trigger system* has been implemented to succeed this scope which is described in detail in Ref. [73].

Two level decision making is performed at the triggering system. *level one* (L1) trigger, where the event rate is reduced to $\sim$ 100 kHz by selecting events upon the information gathered from the sub-detectors only. The next level of the triggering process is known as the *high-level trigger* (HLT). The events that are selected by L1 are served as the input to HLT. The HLT takes a long time compared to L1 to process and perform a basic event reconstruction online and reduces the event rate to $\sim$ 400 Hz. These selected events can be used in offline reconstruction [6].

The L1 trigger system is a complicated hardware system. Separate hardware systems are mounted at several places within the CMS detector. The decision on accepting or rejecting the event is completely based on the information from calorimeters and muon detectors and a decision is made within $4\mu s$ [73]. The overview of the L1 trigger decision is shown in Fig. 3.10. The *global calorimeter trigger* (GCT) involves gathering information from the calorimeters (ECAL and HCAL). This information is then used to identify any possible electron, photon and jet candidates. The *global muon trigger* (GMT) involves in gathering information from muon subsystems; DTs, CSCs and RPC. The muon candidates are identified by using this information by using a fast track finding algorithm. The *global trigger* combines information from both GCT and GMT to produce a trigger menu. The trigger menu is based on the kinematics of the possible identified objects. A final decision is made based upon the trigger menu to either accept or reject the event. For those events, which are accepted, further information is being stored by the sub-detectors to process and perform the online event reconstruction at HLT level[6].

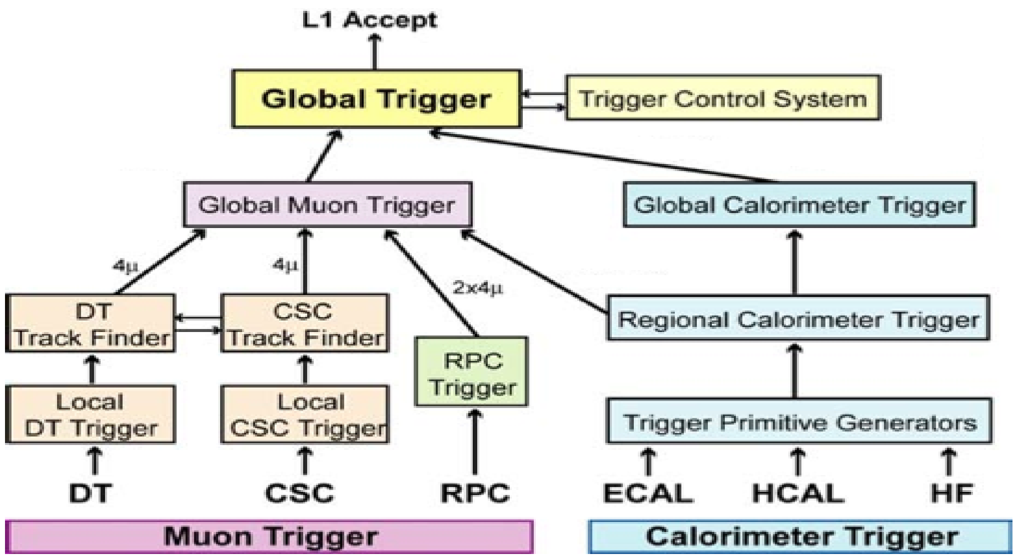


Figure 3.10: *Decision flow of the CMS L1 trigger[74].*

The HLT trigger system is an advanced software based system. It consists of predefined event selection criteria so-called *HLT paths*. The stored information from the sub-detector systems for the accepted events is used for the online event reconstruction. Less time consuming reconstruction steps are performed at first to verify that the event passes the predefined requirements from the HLT menu to preserve time. The physics objects reconstructed can be categorized as :

- **Charged tracks** - The seeds containing at least two hits in the tracking system are reconstructed as charged tracks using the Kalman filter[75].

- **Electrons** - Information from ECAL and HCAL combined with a matching charged track from the inner tracker is used to reconstruct electrons.
- **Muons** - Information from Muons system combined with a matching charged track from the inner tracker is used to reconstruct muons
- **Photons** - Information from ECAL and HCAL is used to reconstruct photons.
- **Jets** - The particle flow (PF) algorithm[76] is used to reconstruct jets using information form calorimeters and tracks.
- **Primary interaction vertex** - Either jets or clustering algorithm[61] is used to reconstruct the primary interaction vertex.
- **Hadronic τ leptons** - The particle flow algorithm is used to reconstruct τ leptons using information form calorimeter and tracks.

The events that are selected by using the HLT menu and passes the HLT trigger are stored in CMS Tier-0 for offline use. This stored data set is known as the *primary data stream*.

Chapter 4

Event simulation, reconstruction and particle identification

4.1 Event Simulation

The data from collider experiments should be compared to theoretical predictions in order to make a quantitative assessment of the standard model parameters or to discover any new physical phenomena[6]. Theoretical predictions are modeled using so-called *Monte Carlo technique* (MC). It involves generation of random numbers based on the rates and probabilities of various physical processes[6]. So-called *event generators* represent software based frameworks and are used to simulate events. The proton-proton collisions corresponding to various physical processes are simulated in two steps:

1. **Event Generation** - Physics processes that produce measurable stable particles as a result of a proton-proton collision are simulated first. There are several physics processes, which should be considered: hard scattering, initial state showers, resonance decays, final state showers, hadronization and additional scattering within the same collision[6]. This is accomplished by simulating these processes using event generators, which are discussed further in this chapter[77].

2. **Detector Simulation** - After generating all the stable particles, the actual response to propagation of those particles through the CMS detector is simulated. The software tool used for this purpose is known as GEANT 4[78], which is discussed briefly in Sec. 4.1.2.

The event generation process begins by simulating the hard scattering process. For the process, $pp \rightarrow ab \rightarrow cd$, where ab and cd represent scattered and final state stable particles respectively, the cross sections are related by the *factorization theorem* given by Eq. 4.1[79] for all the possible partons, which participate in the process.

$$\sigma_{pp \rightarrow cd} = \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \sigma'_{ab \rightarrow cd}(x_a, x_b, Q^2, \mu_R^2, \mu_F^2) \tag{4.1}$$

Where, $\sigma_{pp \rightarrow cd}$ is the cross section of the physics process , $\sigma_{ab \rightarrow cd}$ is the cross section of the parton-parton scattering process , x_a and x_b are fractions of total parton momenta, f_a and f_b are parton distribution functions (PDF). The PDF provides the probability density of a particle i to have the momentum fraction x_i[80]. The total momentum transfer of the parton-parton scattering process is given by Q^2[6]. μ_R^2 and μ_F^2 are the renormalization and factorization scales, respectively. μ_R^2 is a measure of energy scale at which the partonic scattering process occurred, while μ_F^2 is a measure of the threshold between perturbative and non-perturbative regimes[6]. The values of μ_R^2 and μ_F^2 are chosen such that they match to the energy scale of Q^2.

The cross section of the parton-parton scattering process can be calculated using the matrix element, where $\sigma'_{ab \rightarrow cd} \propto |M|^2_{ab,cd}$. The Feynman rules can be incorporated to calculate these matrix elements[6]. They are calculated numerically by using so-called *ME generators* such as POWHEG and MADGRAPH (_aMC@NLO). The hard and soft energy scale transitions are modeled differently in two generators. The POWHEG generator can provide calculations at the next-to-leading order (NLO) accuracy for a set of physical processes[81–83]. The MADGRAPH generator provides calculations for both at leading order (LO) and NLO. It is used to process simulations for new models[6].

The PDFs are used to model effects from both low energy regime and perturbations up to the energy scale of μ_F^2. The dependency of PDFs on non-perturbative energy regime is determined from experiments [6]. About 109 replicas are used, which are generated from different random seeds (labeled as NNPDF3.0), to extract the PDFs used for this analysis. They were calculated by the NNPDF collaboration [84] using experimental data from ATLAS, CMS and LHCb experiments during Run 1 of the LHC, as well as from HERA experiment [6;85].

4.1.1 Parton showering and hadronization

A parton from the hard scattering can emit gluons, which can again either emit more radiation or quark-antiquark pairs. This process continues until these partons begin to form hadrons at the *confinement scale* (Λ_{QCD}). The particle clusters formed from this process are known as *jets* [6;18]. These effects are calculated using the matrix elements of the radiative processes. Some approximations are done using *parton showering algorithms* (PS) for those processes. They operate starting from the energy scale of the hard scattering and moving to lower energy scales by progressive iterations. The process stops after forming mesons and baryons, which is known as *hadronization* [6]. These are approximated by using two famous models, *Lund string model* [86;87] and the *cluster model* [88;89].

Two different matching schemes are applied to avoid possible double counting of final state topologies. MLM scheme [90] is used to simulate LO, and *FxFx* [91] scheme is used to simulate NLO processes. Finally, above-mentioned processes are simulated using a so-called *matrix element generator* that combines several software frameworks. Most commonly used generators are:

- Pythia [92;93] - This is a multi-purpose event generator, which is capable of simulating entire event chain including underlying event[1] using stored generic matrix elements [6]. Transverse momentum is used as the ordering variable to model parton showering, while hadronization is simulated based on the Lund String model [6].

- Herwig++ [94] - In this generator, parton fragmentation is modeled using cluster

[1] Any process which is capable of producing jets other than the hard scattering process is known as an underlying event

hadronization model, while parton showers are simulated using angular ordering[6].

4.1.2 Detector Simulation

The final state stable particles, which are produced from the scattering processes described in previous section are then used to perform the detector simulation. The CMS experiment uses GEANT4 software tool to perform this step. This software includes all actual geometries of the sub-detectors of the CMS apparatus, including the detector response by considering possible noises and interactions with material[6]. The output from this step is passed on to the next level of analyzing the data, which is called an event reconstruction[6].

4.1.3 Signal simulation samples

The $\mathrm{T\overline{T}}$ signal samples are generated in the range $800 - 1800$ GeV in steps of 100 GeV. Equal mixtures of the three final states for each decay of T quark $\mathrm{T \rightarrow tZ}$, $\mathrm{T \rightarrow tH}$ and $\mathrm{T \rightarrow bW}$ are generated and events are further filtered at the generator level to separate six decay channels $tZ\bar{t}Z$, $tZ\bar{t}H$, $tZ\bar{b}W$, $tH\bar{t}H$, $tH\bar{b}W$ and $bW\bar{b}W$ for further re-weighting to simulate different decay probabilities.

The signal processes are produced at the leading order (LO) using the MadGraph5_aMC@NLO 2.3.3 event generator[95]. The parton distribution functions (PDFs) are obtained from NNPDF3.0 by using up to two additional partons for matrix element calculations[84]. PYTHIA 8.212[93] and underlying event tune CUETP8M1[96] are used to simulate the showering and hadronization. The underlying event tune adjusts parameters for simulating events with jets that are not originated from the hard scattering process. Top++ program (v.2.0)[1] is used to obtain the theoretical cross sections for the simulated samples at the NNLO accuracy. These cross sections are used to normalize the signal contributions to observed integrated luminosity. The cross sections, as a function of T mass, are shown in Table 4.1.

The six final states, where $\mathrm{T\overline{T}} \rightarrow tZ\bar{t}Z$, $tZ\bar{t}H$, $tZ\bar{b}W$, $tH\bar{t}H$, $tH\bar{b}W$ and $bW\bar{b}W$ are used to compute the correct event composition corresponding to arbitrary values of branching

Table 4.1: *The cross sections*[1]*, as a function of T quark mass.*

Process	σ (pb)
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 800$ GeV	0.196
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 900$ GeV	0.0903
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1000$ GeV	0.0440
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1100$ GeV	0.0224
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1200$ GeV	0.0118
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1300$ GeV	0.00639
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1400$ GeV	0.00354
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1500$ GeV	0.00200
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1600$ GeV	0.001148
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1700$ GeV	0.000666
$\mathrm{T\overline{T}}$, $\mathrm{m_T} = 1800$ GeV	0.000391

ratios. This is done by re-weighting these six samples according to the relationships given in Table 4.2.

Table 4.2: *The weights for six decay channels of* $\mathrm{T\overline{T}}$*, as a function of T decay branching ratios.*

Channel	Weight
$\mathrm{T\overline{T}} \to tZ\bar{t}Z$	$\mathcal{B}(T \to tZ) \times \mathcal{B}(T \to tZ)$
$\mathrm{T\overline{T}} \to tZ\bar{t}H$	$2 \times \mathcal{B}(T \to tZ) \times \mathcal{B}(T \to tH)$
$\mathrm{T\overline{T}} \to tZ\bar{b}W$	$2 \times \mathcal{B}(T \to tZ) \times \mathcal{B}(T \to bW)$
$\mathrm{T\overline{T}} \to tH\bar{t}H$	$\mathcal{B}(T \to tH) \times \mathcal{B}(T \to tH)$
$\mathrm{T\overline{T}} \to tH\bar{b}W$	$2 \times \mathcal{B}(T \to tH) \times \mathcal{B}(T \to bW)$
$\mathrm{T\overline{T}} \to bW\bar{b}W$	$\mathcal{B}(T \to bW) \times \mathcal{B}(T \to bW)$

For example, consider the benchmark branching fraction $\mathcal{B}(\mathrm{T \to tZ}) = 100\%$. By substituting $\mathcal{B}(\mathrm{T \to tZ}) = 1.0$, $\mathcal{B}(\mathrm{T \to tH}) = 0.0$ and $\mathcal{B}(\mathrm{T \to bW}) = 0.0$ from Table 4.2, we obtain weights 1, 0, 0, 0, 0 and 0 for six samples, $tZ\bar{t}Z$, $tZ\bar{t}H$, $tZ\bar{b}W$, $tH\bar{t}H$, $tH\bar{b}W$ and $bW\bar{b}W$ respectively. Hence, as expected, only $\mathrm{T\overline{T}} \to tZ\bar{t}Z$ sample is used for this benchmark $\mathcal{B}$ scenario with weight 1. In case of another benchmark, where branching fractions are $\mathcal{B}(\mathrm{T \to tZ}) = \mathcal{B}(\mathrm{T \to tH}) = 50\%$, using $\mathcal{B}(\mathrm{T \to tZ}) = 0.5$, $\mathcal{B}(\mathrm{T \to tH}) = 0.5$ and $\mathcal{B}(\mathrm{T \to bW}) = 0.0$ from Table 4.2 we obtain weights 0.25, 0.50, 0.0, 0.25, 0.0 and 0.0 for six samples, $tZ\bar{t}Z$, $tZ\bar{t}H$, $tZ\bar{b}W$, $tH\bar{t}H$, $tH\bar{b}W$ and $bW\bar{b}W$ respectively. Hence, $\mathrm{T\overline{T}} \to tZ\bar{t}Z$,

$\mathrm{T\overline{T}} \to tZ\bar{t}H$ and $\mathrm{T\overline{T}} \to tH\bar{t}H$ samples are combined using weights 0.25, 0.50 and 0.50 respectively, to obtain the final signal event composition for this benchmark $\mathcal{B}$ scenario. Table 4.3 summarizes the weights for three examples of benchmark $\mathcal{B}$s scenarios.

Table 4.3: *Weights calculated for six* $\mathrm{T\overline{T}}$ *decay channels for the three benchmark* $\mathcal{B}$*s.*

$\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = 100\%$			$\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{tH}) = 50\%$			$\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{bW}) = 50\%$		
$tZ\bar{t}Z$	$= 1 \times 1$	$= 1.0$	$tZ\bar{t}Z$	$= 0.5 \times 0.5$	$= 0.25$	$tZ\bar{t}Z$	$= 0.5 \times 0.5$	$= 0.25$
$tZ\bar{t}H$	$= 2 \times 1 \times 0$	$= 0$	$tZ\bar{t}H$	$= 2 \times 0.5 \times 0.5$	$= 0.50$	$tZ\bar{t}H$	$= 2 \times 0.5 \times 0$	$= 0$
$tZ\bar{b}W$	$= 2 \times 1 \times 0$	$= 0$	$tZ\bar{b}W$	$= 2 \times 0.5 \times 0$	$= 0$	$tZ\bar{b}W$	$= 2 \times 0.5 \times 0.5$	$= 0.50$
$tH\bar{t}H$	$= 0 \times 0$	$= 0$	$tH\bar{t}H$	$= 0.5 \times 0.5$	$= 0.25$	$tH\bar{t}H$	$= 0 \times 0$	$= 0$
$tH\bar{b}W$	$= 2 \times 0 \times 0$	$= 0$	$tH\bar{b}W$	$= 2 \times 0.5 \times 0$	$= 0$	$tH\bar{b}W$	$= 2 \times 0 \times 0.5$	$= 0$
$bW\bar{b}W$	$= 0 \times 0$	$= 0$	$bW\bar{b}W$	$= 0 \times 0$	$= 0$	$bW\bar{b}W$	$= 0.5 \times 0.5$	$= 0.25$

4.1.4 Background simulation samples

As mentioned in Sec. 5.1, the $\mathrm{T\overline{T}}$ event signatures can be similar to SM processes. These include, production of a Z boson decaying into two opposite sign leptons ($\mathrm{Z} \to e^+e^-$ or $\mu^+\mu^-$), production of Z, W, H bosons and top and bottom quarks. The SM processes, Drell–Yan (Z/γ^*)+jets (Z decays into e^+e^- or $\mu^+\mu^-$), $\mathrm{t\bar{t}}$ + jets, $\mathrm{t\bar{t}Z}$, dibosons (WW, ZZ or WZ), tZq, tWZ, and $\mathrm{t\bar{t}W}$ represent SM backgrounds for the current analysis. They are simulated using MC@NLO and POWHEG event generators described in Sec. 4.1.

The main background contribution comes from the Drell–Yan (Z/γ^*)+jets (DY+jets) production, where Z boson decays into two opposite sign leptons ($\mathrm{e^+e^-}$ or $\mu^+\mu^-$). The MC@NLO event generator[97] at NLO accuracy is used with FxFx jet matching scheme[91](see Sec. 4.1) to generate the DY+ jets sample. In order to increase statistics of this samples in the phase space considered for the current search, several samples are generated corresponding to different *bins* of transverse momentum of the Z boson ($p_{\mathrm{T}}(\mathrm{Z})$), 100-250 GeV, 250-400 GeV, 400-650 GeV and 650 GeV-infinity. PYTHIA 8.1[98] and underlying event tune CUETP8M1[96] are used for showering and hadronization of generated events. The cross sections for these samples are computed by the event generator, and are determined at the NLO accuracy.

Smaller contributions come from the $t\bar{t}$ + jets and $t\bar{t}Z$ events. POWHEG 2.0[82;83;99] is used to generate $t\bar{t}$+jets events with PYTHIA 8.212 and underlying event tune CUETP8M2T4[100] for showering and hadronization. The production cross section is determined using top quark mass of 172.5 GeV, at the NNLO accuracy. MadGraph5_aMC@NLO 2.3.3 generator[95] at LO accuracy is used to generate $t\bar{t}Z$ events with up to one additional parton in the calculation of matrix elements. The MLM matching scheme (see Sec. 4.1) is used for parton showering[101].

Relatively smaller contributions come from diboson (ZZ, WZ and WW), tZq, tWZ, and $t\bar{t}W$ production. PYTHIA 8.212 is used to generate SM diboson events. The rare single top processes, tZq and tWZ, with $t\bar{t}W$ are generated using MadGraph5_aMC@NLO 2.3.3 generator with MLM matching scheme for parton showering.

The background MC samples used in the analysis are summarized in Table. 4.4 along with their respective cross sections.

Table 4.4: *MC background samples used in the analysis and their cross sections.*

Process	σ (pb)
DY+jets, $p_{\mathrm{T}}(Z) \in [100, 250]$ GeV	83.12
DY+jets, $p_{\mathrm{T}}(Z) \in [250, 400]$ GeV	3.047
DY+jets, $p_{\mathrm{T}}(Z) \in [400, 650]$ GeV	0.3921
DY+jets, $p_{\mathrm{T}}(Z) \in [650, \infty]$ GeV	0.03636
$t\bar{t}$ + jets	831.76
WW	118.7
WZ	46.74
ZZ	16.91
$t\bar{t}Z$	0.783
$t\bar{t}W$	0.611
tZq	0.0758
tWZ	0.01123

4.2 Event reconstruction

A huge amount of data from proton-proton collisions are collected from sub-detectors of the CMS detector. A dedicated system of algorithms exists in order to process this information and to reconstruct physical objects. This chapter provides a brief introduction to those algorithms and the process of particle identification.

4.2.1 Track and primary vertex reconstruction

4.2.1.1 CTF algorithm for general track reconstruction

The very first step, which is performed, is the track identification and reconstruction. The traces that charged particles leave in both pixel and strip detectors can be used to reconstruct *tracks*. For objects like leptons and charged hadrons, the highest energy resolution is provided by the charged tracks and hence keeping the high efficiency of track reconstruction is crucial [6].

The electronic output from the silicon pixel and strip detectors is used to identify the hit clusters for charged track reconstruction. Different algorithms are used to minimize the electronic noise by assigning different charge thresholds to each readout pixel [61].

The *combinatorial track finder* (CTF) algorithm [102–104] is used to reconstruct charged tracks. It is capable of mapping possible particle trajectories and finding a so-called *seed hit* to begin the track reconstruction. The next hit is determined by considering several track based variables and the related probability densities. All possible hits are identified in this way and a final fit is performed using the least-square method to provide the final trajectory of the charged particle [6]. The CTF performs in a way such that the reconstruction starts from easiest tracks using a limited number of hits. Tracks that are already used are masked during the next iteration to reduce computational load. The first iteration starts with tracks having $p_{\mathrm{T}} > 0.8$ GeV and at least three hits in the pixel detector [6]. These requirements are reduced in each subsequent iterations to reconstruct low p_{T} tracks.

For a given iteration, the tracks should satisfy a certain set of requirements to be accepted. These include, the number of hit layers, the number of missed hit layers, χ^2/n_{dof}, transverse and longitudinal impact parameters with respect to the beam spot[2] and the closest pixel vertex divided by their associated uncertainties. The exact requirements are listed in Ref. [61].

[2]The beam spot is the centre of the estimated luminous region which is reconstructed over many events [61].

4.2.1.2 GSF algorithm for electron track reconstruction

Electrons are reconstructed using both tracks and energy deposits from the ECAL, compatible with each other. The bremsstrahlung radiation causes a significant decrease in energy of electrons. This results in a lower number of track hits needed for reconstructing the trajectory of electrons by CTF algorithm, as well as kinks in the electron trajectory. This drawback is overcame by using a *Gaussian sum filter* (GSF) algorithm, which is capable of refitting those poorly reconstructed tracks [6]. It combines information from both the tracker and ECAL to reconstruct electron tracks [6;105].

4.2.1.3 Reconstruction of the primary vertex

The *primary vertex* (PV) is the position, where the hard scattering process is occurred. The positions of tracks identified by the CTF algorithm are reconstructed first. The considered tracks should have a transverse impact parameter significance w.r.t the beam spot ($|d_0|/\sigma_{d_0}$) < 5, at least 2(5) hits in the pixel(strip) layer and the χ^2 of the fit for particle trajectory less than 20 [6].

The *deterministic annealing* (DA) [106] algorithm followed by *adaptive vertex fitter* (AVF) [107] are then used to cluster those tracks according to their z-position of the closest approach from the beam spot and to fit the vertex respectively. The AVF uses specific weights w_i, for each track to make sure that the track is originated from the vertex. These weights are related to the number of degrees of freedom (n_{dof}) of the fit by $n_{dof} = -3 + 2\sum_i w_i$ [6].

Then a primary vertex is identified as the vertex with the maximum transverse momentum of associated tracks ($\sum p_{\mathrm{T}}^2$). They should have $n_{dof} \geq 4$, impact parameter w.r.t the beamspot in the x-y plane ($|\rho|$) < 2 cm and impact parameter w.r.t the beamspot in z direction ($|z|$) < 24 cm.

4.2.2 Particle-flow (PF) algorithm

The PF algorithm incorporates the entire information from sub-detectors to reconstruct final state stable particles, such as electrons, photons and muons [108]. It uses tracks reconstructed by CTF and GSF algorithms described in Sections 4.2.1.1 and 4.2.1.2. It also uses tracks reconstructed from hits in the muon chambers. The extrapolated trajectory of the considered muon tracks should be compatible with the nominal interaction point. These tracks are called *standalone-muon tracks* [6;109].

Calorimeter clusters in ECAL and HCAL are used to measure energy. Clusters are identified in two steps, where at first the cell having the maximum energy is tagged as a *seed cell*. Any nearby cells having energy greater than the noise of the detector are combined with the seed cell. Additional set of clusters in ECAL are identified as *superclusters*, if the hit pattern is identified to be coming from the bremsstrahlung radiation emitted by photons [6;105].

So-called *blocks* are formed by the *link algorithm* [110] by combining all of this information. Next, categorization is performed according to the availability of tracks, either $= 1$ or > 1 accounted from either inner or GSF tracks. The number of clusters/superclusters from both ECAL and HCAL and the number of muon tracks are $= 0$ or > 0. Links are established by extrapolating inner tracks and their tangents to find compatible calorimeter clusters. The tangents of tracks that are compatible with calorimeter clusters can be used to identify bremsstrahlung photons. Clusters/superclusters, which are close in the $\eta - \phi$ plane in PS, ECAL and HCAL are also linked together. Finally, links are also made between inner and muon tracks, if a reasonable χ^2 can be obtained by performing a global fit [57].

The process is iterative, such that reconstruction of the least time-consuming candidates is done first using appropriate blocks. The blocks, which are already used, are masked in the following iteration. These reconstructed candidates are known as *particle-flow candidates* and are listed below.

- **Electrons** - Appropriate blocks are chosen such that, they do not have more than two tracks in total, while at least one of them is a GSF track. The energy measured in ECAL and HCAL are linked and measured energy in the HCAL clusters should be less than 10% of measured energy in the ECAL clusters [6]. A multivariate technique is used to identify the PF electrons by considering various block elements like energy loss of GSF tracks, while linking together the ECAL clusters with tracks[111]. The identified tracks and calorimeter clusters are then used to determine the momentum and the energy of the PF electrons.

- **Muons** - The blocks containing muon tracks linked with inner tracks are used to identify PF muons. Three types of muons can be reconstructed in this way, which are called *standalone*, *tracker* and *global* muons. Standalone muons are reconstructed using only tracks in the muon chambers. Tracker muons are identified using inner tracks that when extrapolated are compatible with at least one hit in the muon chambers[6]. The inner tracks that are compatible with tracks from standalone muons are used to reconstruct global muons. Identified PF muons should be either tracker or global muons, depending on their p_{T}. Muons with $p_{\mathrm{T}} < 10$ GeV are reconstructed as tracker muons, while muons with $p_{\mathrm{T}} > 10$ GeV are reconstructed as global muons. For very high p_{T} muons ($p_{\mathrm{T}} > 200$ GeV), a combinatorial fit is performed between inner and muon tracks to obtain the muon p_{T}[6].

- **Charged hadrons** - Blocks with inner tracks compatible, with energy deposits in both ECAL and HCAL are used to reconstruct the PF charged hadrons. Energy deposits in both ECAL and HCAL are re-calibrated by charged hadron simulation to increase the accuracy in the energy measurement[108]. This step of energy re-calibration is important at high energies where the energy resolution of inner tracks is low and since the higher η region ($\eta > 2.5$) is not covered by pixel or strip detectors. In those cases, the four-momenta of the charged hadrons are determined using the information from calorimeters only[6].

- **Neutral hadrons and photons** - The *isolated PF photon* candidates are identified in a similar way as PF electron candidates, i.e. the links between both ECAL and HCAL are considered, where energy measured in the ECAL superclusters should exceed 10 GeV and the energy measured in the HCAL clusters must be less than 10 % of the energy deposited in linked ECAL cluster. The identified ECAL clusters should be isolated from inner tracks and HCAL clusters. The *non-isolated PF photons* and *PF neutral hadrons* are identified by considering any remaining calorimeter clusters from ECAL and HCAL that are not linked to inner tracks[6;57].

- **Missing transverse momentum (MET)** - The identification of neutrinos emitted during the hard scattering process are determined by the *transverse momentum* of all the other particles. According to the momentum conservation law, the sum of momenta of all the particles in the transverse plane should be equal to zero. This momentum imbalance can be used to determine the energy and momentum of neutrinos[18]. The Eq. 4.2 and 4.3 give the *missing transverse momentum* and *missing transverse energy* respectively.

$$\vec{p}_{\mathrm{T}}^{\,miss} = -\sum_{i} \vec{p}_{\mathrm{T}(i)} \tag{4.2}$$

where, i runs over all the particles within an event.

$$E_{\mathrm{T}} = |\vec{p}_{\mathrm{T}}^{\,miss}| \tag{4.3}$$

.

4.2.3 Relative lepton isolation (I_{rel})

This variable is defined to identify leptons originated from the hard scattering process, which are isolated from the jet activity[18]. It is given by Eq. 4.4, where p_{T} in the denominator denotes the transverse momentum of the lepton.

$$I_{rel} = \frac{1}{p_{\mathrm{T}}(lepton)}\Big(\sum p_{\mathrm{T}}(charged\ hadrons) + \sum p_{\mathrm{T}}(neutral\ hadrons) + \sum p_{\mathrm{T}}(photons)\Big) \tag{4.4}$$

This variable is usually calculated within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the lepton, where $\Delta\eta$ and $\Delta\phi$ are differences in pseudorapidity and azimuthal angle w.r.t the direction of lepton. The contributions from pileup are removed before calculation[112], and ΔR is chosen to be 0.3(0.4) for electrons(muons). In addition, for electrons, I_{rel} is also corrected for effective area of the jets[112;113].

4.2.4 Identification of Electrons

The electrons used for this analysis should satisfy an additional set of criteria apart from the PF electron selection. These requirements are implemented by the CMS electron physics object group to identify so-called *tight* electrons and are known as *cut-based electron identification and isolation criteria*[114]. The following variables are used for the selection of electrons that fall in EB($|\eta_{SC}| < 1.479$) and EE ($1.479 < |\eta_{SC}| < 2.5$) regions of the ECAL detector with selections summarized in Table 4.5.

- **Full5x5**_$\sigma_{i\eta i\eta}$ - This is a variable, which depends on a shape of the energy deposited in an area of 5×5 cluster crystals around the seed crystal in ECAL[46].
- $|\Delta\eta_{\,\mathbf{InSeed}}|$ - The absolute distance between the supercluster seed and the extrapolated GSF track in the η direction.
- $|\Delta\phi_{\,\mathbf{In}}|$ - The absolute distance between the supercluster and the extrapolated GSF track in the ϕ direction[46].
- **H/E** - The ratio of energies measured in HCAL and ECAL[46].
- **RelIso (withEA)** - The relative electron isolation with corrections for effective jet areas mentioned in Sec. 4.2.3, calculated within a cone of $\Delta R = 0.3$.
- **|1/E - 1/p|** - The absolute difference in the inverse super cluster energy and the track momentum[46].
- **Expected Missing Hits** - The expected number of missed hits in the electron trajectory[46].
- **pass conversion veto** - The rejection of tracks originating from photon conversion.

- $|d_{\mathbf{0}}|$ - The transverse impact parameter w.r.t the primary vertex.

- $|d_{\mathbf{z}}|$ - The longitudinal impact parameter w.r.t the primary vertex.

Table 4.5: *Tight Electron selection requirements.*

Observable	Value (EB)	Value (EE)
full5x5$\sigma_{i\eta i\eta}$	< 0.00998	< 0.0292
$\|\Delta\eta_{\text{ InSeed}}\|$	< 0.00308	< 0.00605
$\|\Delta\phi_{\text{ In}}\|$	< 0.0816	< 0.0394
H/E	< 0.0414	< 0.0641
RelIso (withEA)	< 0.0588	0.0571
|1/E - 1/p|	< 0.0129	< 0.0129
Expected Missing Hits	<= 1	<= 1
pass conversion veto	yes	yes
$\|d_{0}\|$	< 0.05	< 0.10
$\|d_{z}\|$	< 0.10	< 0.20

This criteria result in the electron identification efficiency for the *tight* id working point of about 70%[114]. Scale factors are derived by the CMS electron object group for any inefficiency of the ID and reconstruction[115] by using so-called *tag and probe* method[105], which depends on both electron p_{T} and supercluster $|\eta|$. These scale factors are applied to correct MC simulation, as a function of the p_{T} and η of the electron.

"Tight" electrons, which are identified according to the criteria mentioned above, are used for the current analysis. In addition, selected electrons should reside in the $|\eta| < 2.4$ region excluding the gap, $1.442 < |\eta| < 1.556$ in ECAL.

4.2.5 Identification of Muons

The muons used for this analysis should satisfy an additional set of criteria apart from the PF muon selection. These requirements are implemented by the CMS muon physics object group to identify so-called *tight* muons[116].

- **is Global Muon** - The muon is identified as a global muon as described in Sec. 4.2.2.

- **is PF Muon** - The muon is identified as a PF muon by the PF algorithm as described in Sec. 4.2.2.

- **χ^2 /ndof of the global-muon track fit** - This quantity represents the quality of the muon track, which reflects how well the hits are modeled in the fit[117].

- **no. of valid muon hits** - The number of muon hits identified by the PF algorithm.

- **no. of matched muon station** - The number of muon stations identified with muon hits.

- $|d_{\mathbf{0}}|$ - The muon impact parameter in the transverse plane w.r.t the primary vertex.

- $|d_{\mathbf{z}}|$ - The muon impact parameter in the longitudinal plane w.r.t the primary vertex.

- **no. of pixel hits** - Number of hits in the pixel detector.

- **no. of track layers** - Number of layers in both of the tracking detectors.

- $Iso_{\mathbf{rel}}$ - The relative muon isolation described in Sec. 4.2.3, calculated within a cone of $\Delta R = 0.4$.

Table 4.6: *Tight muon identification requirements.*

Observable	Value or range
is Global Muon	yes
is PF Muon	yes
χ^2 /ndof of the global-muon track fit	< 10
no. of valid muon hits	> 0
no. of matched muon station	> 1
$\lvert d_0 \rvert$	< 0.2
$\lvert d_z \rvert$	< 0.5
no. of pixel hits	> 0
no. of track layers	> 5
Iso_{rel}	< 0.15

Scale factors are derived by the CMS muon object group for any inefficiency of the ID and reconstruction[118] using the *tag and probe* method[119] as a function of both muon p_{T}

and supercluster $|\eta|$. These scale factors are applied to correct MC simulation, as a function of the p_{T} and η of the muon.

"Tight" muons, which are identified according to the criteria mentioned above are used for the current analysis. In addition, selected muons should reside in the $|\eta| < 2.4$ region.

4.2.6 Identification and reconstruction of jets

Quarks and gluons produced during collisions do not exist as single particles. They undergo a process called *hadronization*, as described in Sec. 2.1.1, and produce a collimated spray of lightly bound particles in the form of charged and neutral hadrons[6]. Dedicated clustering algorithms exist, which take the four-momenta of these particles as their input and reconstruct *jets*. The particles, which are reconstructed by the PF algorithm mentioned in Sec. 4.2.2 are used for this purpose.

The *sequential recombination algorithm*[120] is widely used to reconstruct jets. This algorithm runs iteratively based on two variables, d_{ij} and d_{iB} given by Eq 4.5 and 4.6 respectively[121].

$$d_{ij} = min(p_{Ti}^{2q}, p_{Tj}^{2q}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.5}$$

$$d_{iB} = p_{Ti}^{2q} \tag{4.6}$$

Here, the angular separation between the objects i and j is given by ΔR_{ij}. The shape and cone size of the jet is influenced by the pre-defined parameters, q and R respectively[6]. At each iteration, the object i is clustered with object j, if it satisfies the condition $d_{ij} < d_{iB}$. This iteration continues until all the input objects are considered and clustered into jets[6;18]. This iterative procedure is shown in Fig. 4.1. The final output of the sequential recombination algorithm does not contain very soft and low-angle radiation.

The clustering algorithm adds the four-momenta of the particles at each iterating step and most important jet quantities like mass, energy, momentum and direction can be extracted from the four-momentum of the clustered jet. The axis of the clustered jet is determined by

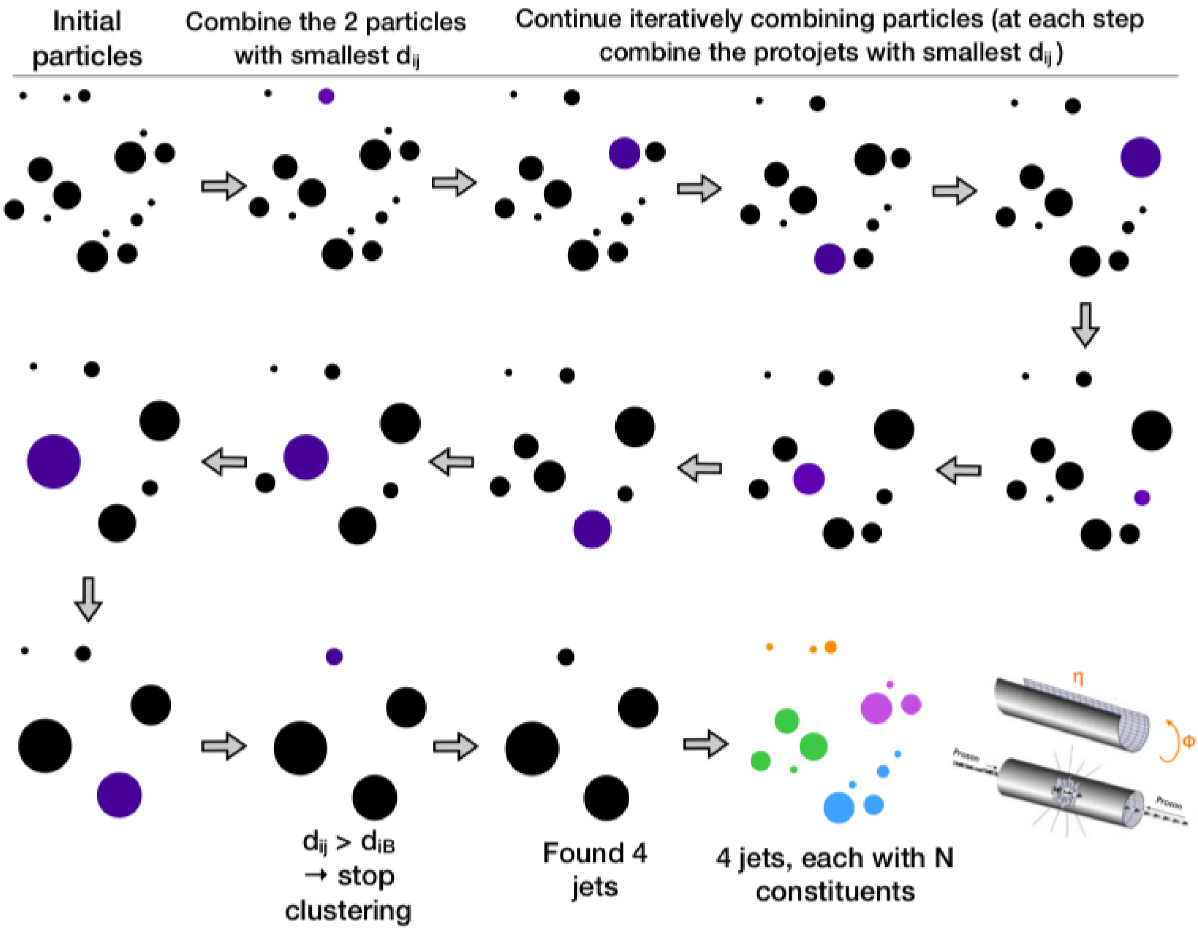


Figure 4.1: *The mechanism of jet clustering algorithm*[121].

the sum of three momenta (p_x, p_y, p_z) of the constituent particles[6]. Depending on the choice of q value in Eq. 4.5 and 4.6, three common algorithms exist, which are *k_T algorithm* (KT), *Cambridge Aachen algorithm* (CA) and *anti-k_T algorithm* (AK) as shown in Table. 4.7.

Table 4.7: *Jet clustering algorithms.*

q value	Name	description
1	$\mathrm{k_T}$ algorithm	Soft particles are clustered first
0	Cambridge Aachen algorithm	Clustering based on angular seperation
-1	anti-$\mathrm{k_T}$ algorithm	hard particles are clustered first

CMS uses anti-$\mathrm{k_T}$ algorithm to reconstruct jets[122]. This algorithm clusters starting from the hardest particles within the collection. The hardest particles are the ones that have the highest momentum and hence jets, which are reconstructed using this algorithm, have the largest fraction of momentum of the initial parton. The reconstructed jets have a circular cone size for a given R value[6]. Two types of jets are used in this analysis based on the R value in Eq. 4.5.

- **AK4 jets**- These are reconstructed jets using $R = 0.4$. They are capable of capturing the hard gluons and light flavor quarks such as u, d, c, s and b.
- **AK8 jets** - These are reconstructed jets using $R = 0.8$. They are useful in cases where there are jets originated from two or more nearby partons.

Both types of jets must satisfy $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ to be considered as candidates for this analysis. They also should satisfy the *loose* PF jet identification given in Table. 4.8 as recommended by the CMS JetMET Physics object group[123]. Any charged hadron track originated from non-primary vertices are removed before clustering in order to minimize the effect from pileup interactions. This process is commonly known as *charged hadron subtraction* (CHS)[124]. This is applied to all jets and the reconstructed jets are known as PFCHS jets.

Table 4.8: *Loose and Tight jet identification requirements.*

Observable	Value or range (loose)	Value or range (Tight)
neutral hadron energy fraction of the jet	< 0.99	< 0.90
neutral EM fraction	< 0.99	< 0.90
number of constituent particles	> 1	> 1
$\|\eta\|$	≤ 2.4	≤ 2.4
charged hadron energy fraction of the jet	> 0.00	> 0.00
charged hadron multiplicity	> 0	> 0
charged EM energy fraction of the jet	< 0.99	< 0.99

AK4 jets that satisfy the *loose* jet identification criteria with jet $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ are selected for the current analysis. Leptons, which are misidentified as jets, are removed from the AK4 jet collection, if they are within a cone of $\Delta R(\text{lepton}, \text{jet}) = 0.4$.

Since many hadronic decays are involved in this analysis, the events that have at least three AK4 jets are selected.

In addition, the jets should have a fair amount of boost. Jets that are originated from a high energetic parton with high transverse momentum are known as *boosted jets.* The variable H_{T} defined as, $H_{\mathrm{T}} = \sum_{p_{\mathrm{T}}}(AK4)$ is used to select events having boosted jets with $H_{\mathrm{T}} > 200$ GeV for the current analysis. Since the decay products from a massive T quark decay are expected to be boosted, the leading and second-leading jet thresholds are raised to $p_{\mathrm{T}} > 100, 50$ GeV respectively for the signal search, where any additional jets should have $p_{\mathrm{T}} > 30$ GeV.

4.2.6.1 Jet energy correction (JEC)

Many factors affect the energy of the reconstructed jets. Most important contribution is from pileup interactions (mostly from neutral hadrons, which are not removed by charged hadron subtraction process) and how sub-detectors, such as calorimeters, respond to the particle interactions with the material in the detector. The energy of the initial parton divided by the energy of the reconstructed jet is defined as the JES [6]. The simulated jet energy scale response is shown in Fig. 4.2 as a function of jet η for different values of jet p_{T}. There is almost stable response in EB, while there is a strong p_{T} dependence in both EC and HF [125].

The differences between observed and simulated jet p_{T} are accounted for by applying jet energy corrections to the simulated jet. These corrections are

1. Pileup corrections

2. Relative and absolute corrections

3. Residual corrections

4.2.6.2 Pileup corrections

Any remaining contributions from pileup interactions to JES that are not removed by CHS procedure are corrected by applying the pileup corrections [6]. It corrects the individual jet

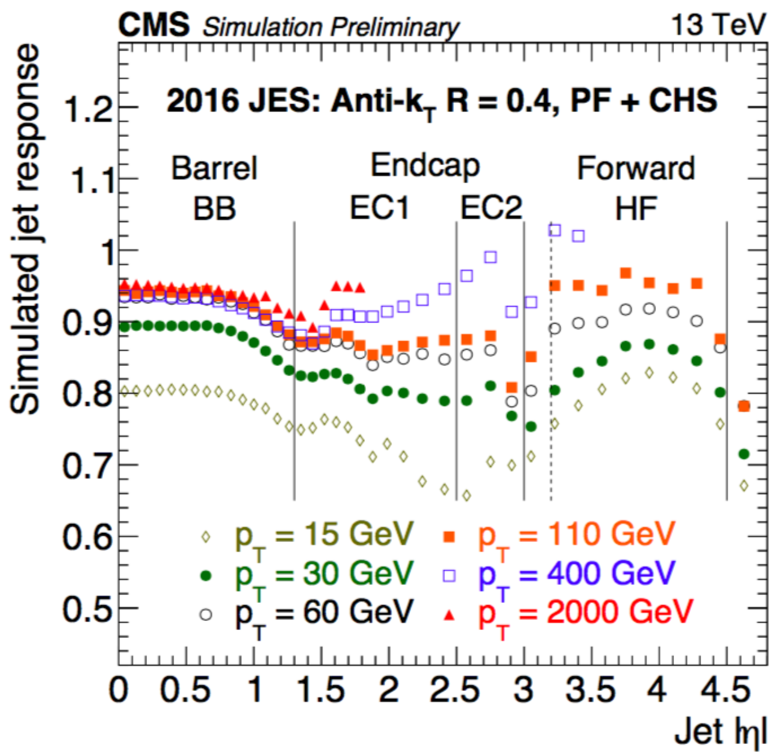


Figure 4.2: *The simulated jet energy scale response as a function of jet η for different jet p_T values*[125].

response by using a correction factor based on jet p_{T}, η, area (A) and diffuse offset energy (ρ). The variables A and ρ are defined as area and mean grid cell energies occupied by the jet in the $\eta - \phi$ plane[6].

The correction factors are obtained by simulating events with and without pileup interactions using di-jet samples. They are applied to jets that are reconstructed in both data and simulation[6]. Any difference remaining between the data and simulation, after applying those scale factors is also accounted using so-called *Zero bias* data stream and applied to simulation as an additional scale factor. The detailed procedure can be found in Ref.[125;126]. The Fig. 4.3 illustrates the average offset in transverse momentum (p_{T}) per average additional pileup interaction (μ) in data and simulation (left) and additional scale factors to correct simulation, as a function of η (right) derived using data collected by CMS experiment during different run periods in 2016.

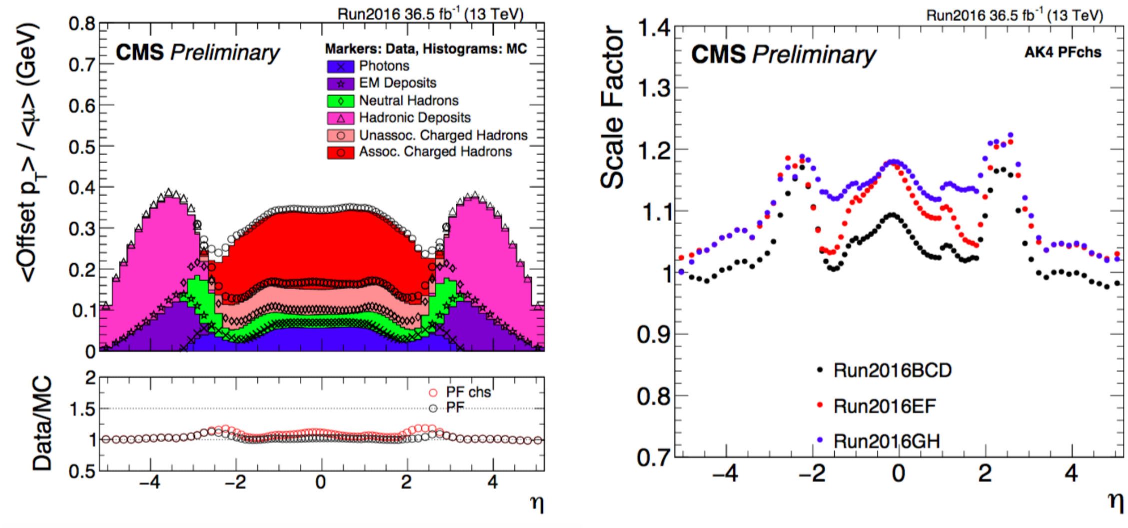


Figure 4.3: *Average offset in transverse momentum (p_T) per average additional pileup interaction (μ) in data and simulation (left) and scale factors as a function of η (right)*[125].

4.2.6.3 Relative and absolute corrections

These are used to correct the JES for the full η range considered by trying to match it to the original parton energy. The initial parton and their corresponding reconstructed jet are matched using ΔR(initial parton, reconstructed jet) < 0.2 requirement, using simulated multi-jet sample[6]. The correction factor R_{ptcl}, which is a function of average jet p_{T} and η, is given by Eq. 4.7 is applied to both data and simulation.

$$R_{ptcl}(\langle p_{\mathrm{T}} \rangle, \eta) = \frac{\langle p_{\mathrm{T}} \rangle}{\langle p_{\mathrm{T},ptcl} \rangle}[p_{\mathrm{T},ptcl}, \eta] \tag{4.7}$$

Where *ptcl* stands for quantities from particle-level jets.The term $\frac{\langle p_{\mathrm{T}} \rangle}{\langle p_{\mathrm{T},ptcl} \rangle}[p_{\mathrm{T},ptcl}, \eta]$ represents the ratio of arithmetic means of transverse momenta of reconstructed and particle level jets in bins of p_{T} of the particle level jet and η of the reconstructed jet[126].

4.2.6.4 Residual corrections

These corrections are involved in matching the simulated jet response to that of data. Di-jet events are used to derive the correction factors by using the *tag and probe method*[126]. The *tag jet* should satisfy $|\eta| < 1.3$, while the *probe jet* does not have any restriction on η. Two different correction factors, $R_{rel}^{p_\mathrm{T}}$ and R_{rel}^{MPF} given by Eq. 4.8 and 4.9[126] are used, where the former uses transverse momenta and the latter considers all hadronic activities including MET(also known as MPF method)[6].

$$R_{rel}^{p_\mathrm{T}} = \frac{1 + \langle A \rangle}{1 - \langle A \rangle} \tag{4.8}$$

$$R_{rel}^{\mathrm{MPF}} = \frac{1 + \langle B \rangle}{1 - \langle B \rangle} \tag{4.9}$$

where,

$$A = \frac{p_{\mathrm{T,probe}} - p_{\mathrm{T,tag}}}{2p_{\mathrm{T,ave}}};\ B = \frac{\vec{p}_\mathrm{T}^{\,miss}.(\vec{p}_{\mathrm{T,tag}}/p_{\mathrm{T,tag}})}{2p_{\mathrm{T,ave}}} \tag{4.10}$$

By using the ratio given by Eq. 4.11, the scale factors are derived to correct p_T and MET in the simulation. The Fig. 4.4 shows the correction value at average p_T in each bin of η for data collected by CMS experiment at different run periods during 2016[126].

$$R_{rel} = \frac{R_{rel,data}}{R_{rel,simulation}} \tag{4.11}$$

The $Z \rightarrow \mu^+\mu^-(e^+e^-) + jets$ and $\gamma + jets$ events are used to calculate the residual corrections as a function of p_T. This procedure makes use of the fact that the mass of the Z boson, photon and leptons energies are more precisely measured than jet energies[6]. Therefore, the measured energy of the boson candidate can be used to calibrate the energy of the corresponding reconstructed jet using the energy balance[6]. The corrections are also extrapolated to the higher p_T regime by smearing multi-jet events and a global fit is performed in all of these measurements to calculate the corrections[6]. These corrections for different data taking periods are shown in Fig. 4.5[125].

4.2.6.5 Jet energy resolution (JER)

The JER is determined by using the distribution of $p_\mathrm{T}/p_{\mathrm{T,ptcl}}$, where p_T and $p_{\mathrm{T,ptcl}}$ are transverse momenta of the reconstructed and particle level jet respectively. This distribution

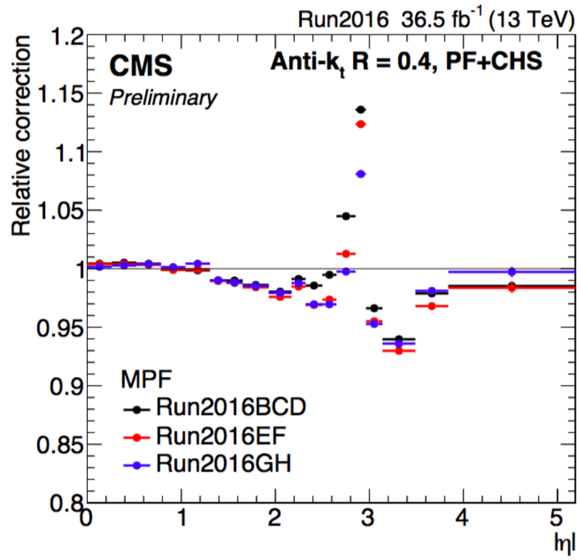


Figure 4.4: *correction value at average p_T in each η bin for data collected by CMS experiment during different run periods*[125].

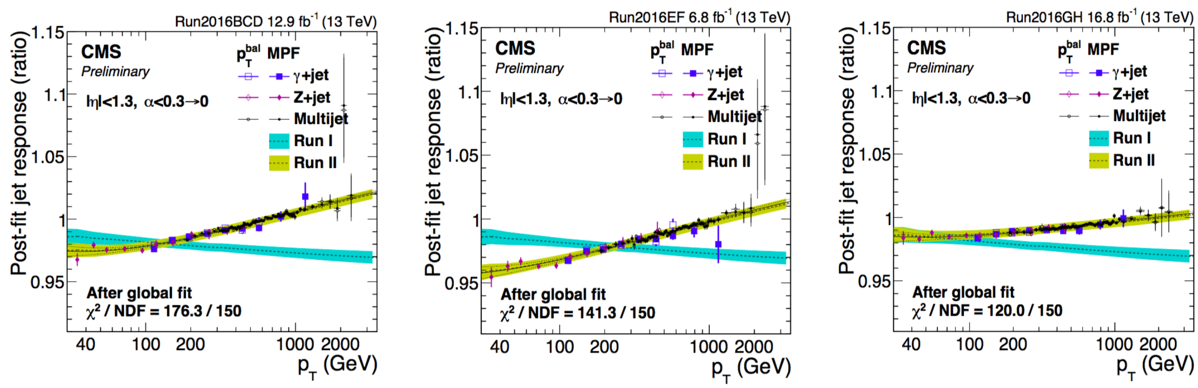


Figure 4.5: *Residual corrections as a function of jet p_T for data collected by CMS experiment during different run periods*[125].

has a shape of a *Gaussian* distribution as shown on the left plot of Fig. 4.6 and the resolution is given by the width of this distribution[127].

The JER scale factors are derived using $Z \rightarrow \mu^+\mu^-(e^+e^-) + jets$ and $\gamma + jets$ events. The measured energy of the boson candidate is used to calibrate the reconstructed jets using the energy balance and to measure the η dependent scale factors (SF(η_{reco})). They are used to smear the jet p_{T} according to Eq. 4.12.

$$p_{\mathrm{T}}(smeared) = p_{\mathrm{T}}(ptcl) + SF(\eta_{reco})(p_{\mathrm{T}}(reco) - p_{\mathrm{T}}(ptcl)) \tag{4.12}$$

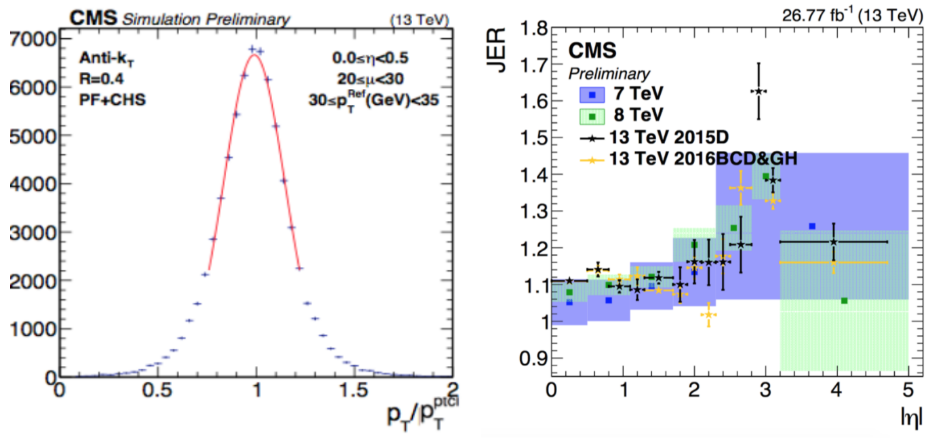


Figure 4.6: *The distribution of $p_T/p_{T,ptcl}$(left) and JER scale factors as a function of $|\eta|$ (right)*[127].

4.2.6.6 Identification of b jets

The jets originating from light flavor quarks are almost similar and difficult to differentiate from each other. However, there are some differences between jets originating from c and b quark decays and from decays of the light quarks[6]. These differences can be used to identify them. Identification of jets originating from b quarks (known as b *jets*) is essential for this analysis, as it is sensitive to the decays, $T \to bW$, $T \to tZ$ and $T \to tH$, where *top* quark decays to a b quark and a W boson.

The property that allows a b jet to be identified within the CMS detector comes from their relatively long lifetime compared to the rest of the light flavor quarks. The b quarks travel a small distance away from the primary interaction vertex before hadronization (to form a B meson), which leads to *displaced tracks* as shown in Fig. 4.7. The position, where B meson decays, is known as the *secondary vertex* (SV) and can be identified using the pixel detector[6].

b *tagging algorithms* use this information to distinguish b jets from other light flavor

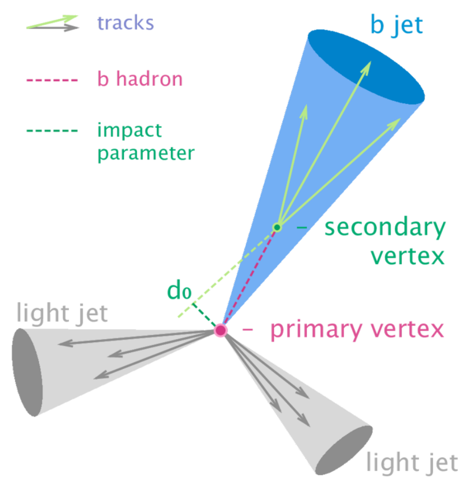


Figure 4.7: *Identification of a b jet from displaced tracks*[128].

jets. CMS uses the *combined secondary vertex* (CSV) algorithm to identify b jets[129]. It makes use of both tracks in a jet and secondary vertices, which are reconstructed using the *inclusive secondary vertex finder* (IVF) and the *adaptive vertex filter* (AVF) mentioned in Sec. 4.2.1.3[130]. The complete selections on tracks and jets can be found in Ref[130]. The output of the CSV algorithm provides a numerical value ranging from $0-1$, where b jets have a higher value closer to 1, while the light jets have a lower value and hence can be used as a discriminator. c-jets could be misidentified as b-jets since they have a discriminator value in the medium range between b-jets and light flavor jets.

Fig. 4.8 shows the discriminator (CSVv2) for AK4 jets obtained using a muon enriched QCD jet sample[131]. The different colored stacked histograms show the discriminator for different jet flavors. It can be seen that majority of the b-tagged AK4 jets (red colored histogram) have discriminator values closer to 1, while majority of the light flavor u, d, s and g tagged AK4 jets (blue colored histogram) have lower discriminator values. The c-jets have discriminator values in the medium range.

Different working points for the CSV algorithm are introduced by the CMS *b-tagging* physics object group (BTV POG)[132] based on the discriminator (CSVv2) value. This analysis uses *medium* working point (CSVv2M = 0.8484), which corresponds to a b-tagging efficiency $\sim 70\%$ and misidentification rate for light jets $\sim 1\%$. The scale factors for b and

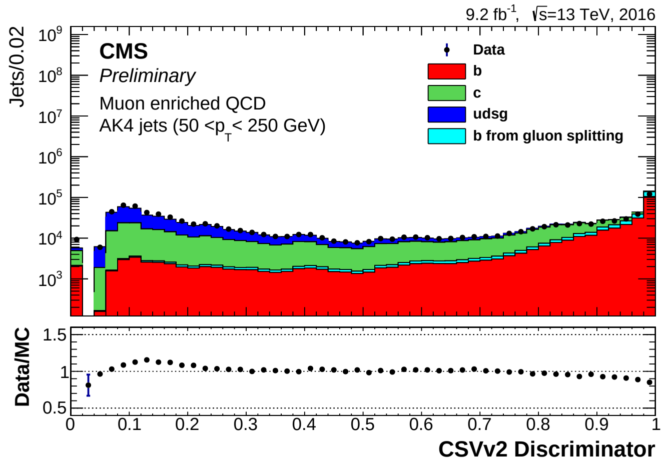


Figure 4.8: *CSVv2 discriminator for AK4 jets obtained using a muon enriched QCD jet sample*[131].

light flavor tagging are provided by the CMS BTV physics object group, where the scale factor is given by the ratio of ϵ_{data} and $\epsilon_{simulation}$. Here, ϵ_{data} and $\epsilon_{simulation}$ are tagging efficiencies in data and simulation respectively[133]. The scale factors and efficiencies are functions of jet flavour, p_{T} and η. The simulation efficiencies for each background and signal processes considered in this analysis are obtained as a function of jet p_{T} and η after some pre-event selection (which is described in Sec. 5.3.4). The simulation efficiencies are used to obtain the scale factors and applied as an event weight to the simulated MC events. Signal selection for the current analysis requires at least one b-tagged AK4 jet per event with $p_{\mathrm{T}} > 50$ GeV.

4.2.6.7 Jet substructure techniques

Jet substructure techniques are extensively used to identify *boosted heavy particles* such as top quarks and SM bosons. At low p_{T}, decay products of these particles can be reconstructed as separate AK4 jets. When p_{T} of the heavy particle is considerably high (boosted), emitted jets become collinear (small separation in $\eta - \phi$ plane) and it is hard to reconstruct them as separate AK4 jets[6].

An example is shown in Fig. 4.9, where top quark decays into a $W \rightarrow q\bar{q}'$ and a b quark.

At low p_{T}, all three quarks, q, q' and b can be reconstructed as three separate AK4 jets. But at high p_{T}, the decay products are more boosted so that they become collinear with each other.

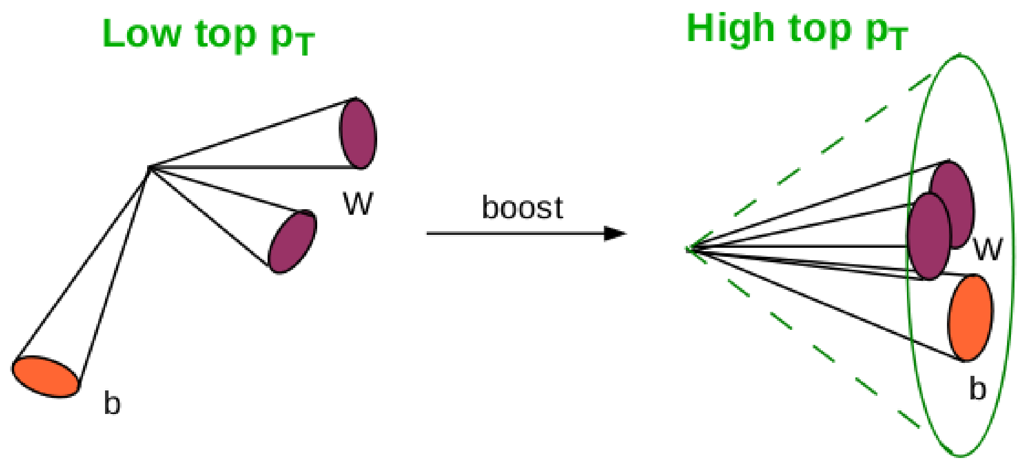


Figure 4.9: *Boosted top jet[134].*

A new technique has been developed known as *jet substructure technique* by the CMS collaboration, where decay products of this type of boosted particles are reconstructed as a single jet with a higher distance parameter than AK4 jets[6]. Then the inner structure of this jet is used to identify the original heavy particle. This technique is employed to identify boosted objects used in this analysis[135;136]. CMS uses AK8 jets to identify boosted objects.

There are two major jet grooming algorithms, *soft-drop* and *pruning*, which are used to aid the substructure technique. They basically remove the soft radiation and large-angle emissions in jets and are capable of providing the measurement if mass of the original heavy particle very accurately, which is known as the *groomed jet mass*. Employing this technique also reduces the influence by pileup interactions on jets[18].

4.2.6.8 The soft-drop algorithm

The soft-drop algorithm is based on Cambridge- Aachen (CA) algorithm mentioned in Sec. 4.2.6, which re-clusters jet constituents according to their angular separation. Pair of jets $(j1, j2)$ is considered by declustering the original jet at first. Then soft-drop condition is given by Eq. 4.13:

$$\frac{min(p_{\mathrm{T},j1}, p_{\mathrm{T},j2})}{p_{\mathrm{T},j1} + p_{\mathrm{T},j2}} > z_{cut} \left(\frac{\Delta R(j1, j2)}{R_C} \right)^{\beta} \tag{4.13}$$

where the cone size parameter of the original jet is given by R_c, $\Delta R(j1, j2)$ is the separation between $j1$ and $j2$ in $\eta - \phi$ plane, z_{cut} refers to the soft threshold, and β is the angular exponent [6;137].

If subjets, $j1$ and $j2$, satisfy the above condition, the new jet (j) is constructed as the final soft-drop jet and then the iteration stops. Otherwise, the jet j is defined as the sub-jet with the highest p_{T} of the pair and the iteration continues until further de-clustering cannot be done. Then the jet j is defined to be the final soft-drop jet. This is illustrated in Fig. 4.10. The jet mass obtained after re-clustering is known as *soft-drop mass* [6;18].

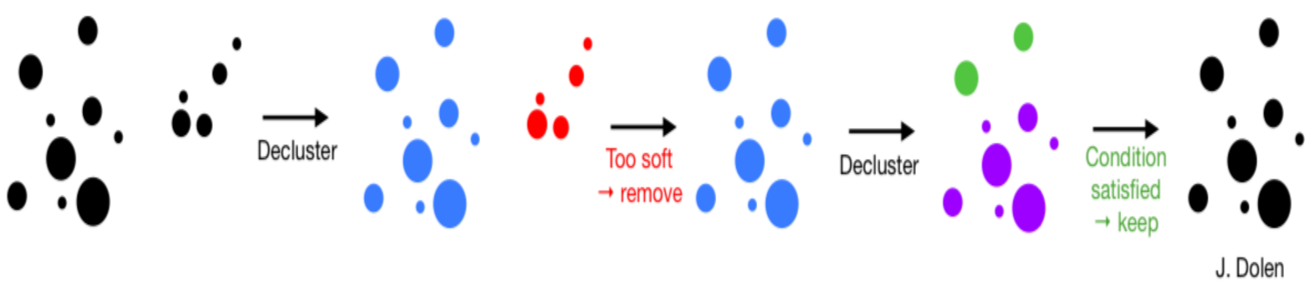


Figure 4.10: *Illustration of soft drop algorithm*[121].

If z_{cut} is chosen to be 0.1, then different β values give different characteristics of the algorithm. $\beta = 0$ will remove the softer sub-jet given that $p_{\mathrm{T}} < 0.1$ and is equivalent to the *modified mass-drop tagger* (MMDT)[138]. $\beta < 0$ tags hard and large angle jets instead of grooming them. $\beta > 0$ removes the softer sub-jet given that $p_{\mathrm{T}} < x$, where x is a ΔR dependent parameter with a maximum value of 0.1. If $\beta \to \infty$, no grooming is performed[18;121]. These characteristics are shown in Fig. 4.11.

Fig. 4.12 illustrates the significance of using the soft-drop mass when identifying boosted *top* quark jets. The left plot shows the difficulty of separating QCD jets with merged *top* quark jets by using the un-groomed jet mass. The right plot shows how easily a merged *top* jet can be identified by selecting a soft-drop mass window, which is symmetric around the

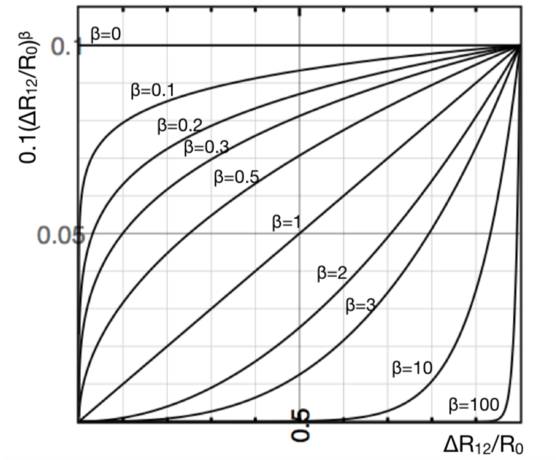


Figure 4.11: *Characteristics of the soft-drop algorithm with $z_{cut} = 0.1$ and for different β values*[121].

top quark mass.

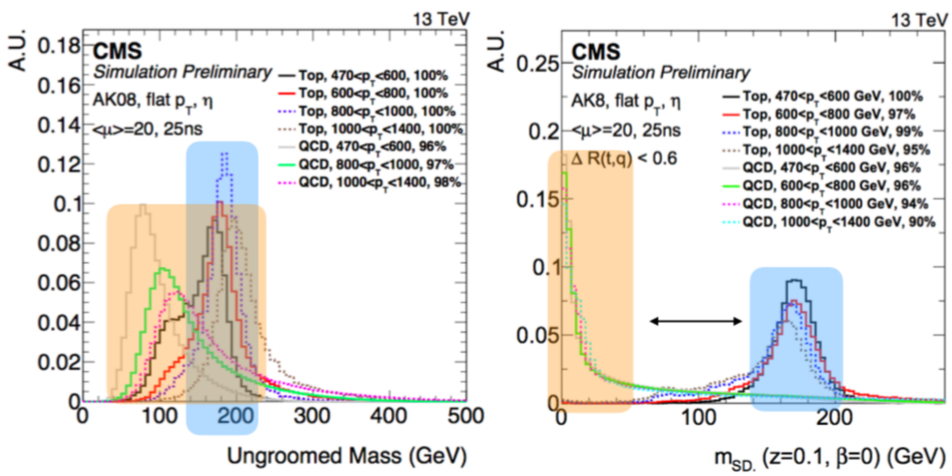


Figure 4.12: *Significance of using soft-drop algorithm to identify boosted top quark jets. mass distributions without(left) and with (right) soft-drop algorithm are shown*[121].

4.2.6.9 The pruning algorithm

The pruning algorithm iterates similarly to the soft-drop algorithm except for different conditions[139]. A pruning radius (R_{prune}) is defined for each jet by using the mass (m) and the transverse momentum (p_{T}) and new radius parameter R_{fact} given by Eq. 4.14[18].

$$R_{prune} = R_{fact}\frac{2m}{p_{\mathrm{T}}} \tag{4.14}$$

There are two conditions that should be satisfied by the two jets, *jet1* and *jet2* under consideration, which are given by Eq. 4.15. The jet with smaller p_{T} is dropped if the condition is not satisfied, otherwise both jets are kept. Decay products of the heavy particle are determined by the final subjets identified by the algorithm. z_{cut} and R_{fact} values are set to 0.1 and 0.5 respectively[18].

$$\begin{aligned} \Delta R(jet1, jet2) > R_{prune} \\ min(p_{\mathrm{T},jet1} p_{\mathrm{T},jet2}) < z_{cut}(p_{\mathrm{T},jet1} + p_{\mathrm{T},jet2}) \end{aligned} \tag{4.15}$$

4.2.6.10 N-subjettiness(τ_N)

The N-subjettiness (τ_N) is a variable, which measures consistency of a jet with having N hypothesized sub-jets. It is given by Eq. 4.16[121;140]. The value can vary between $0 < \tau_N < 1$. The lower the τ_N is, the higher is the probability that the jet consists of N sub-jets. The ratio τ_N/τ_M where $M = N - 1$ is also used to obtain better discriminating power[121]. For example, τ_2/τ_1 ratio can be used to identify hadronic decay of a Z boson, since the decay consists of only two quarks that are reconstructed as two sub-jets within an AK8 jet.

$$\tau_N = \frac{1}{\sum_i p_{\mathrm{T},i} R} \sum_i p_{\mathrm{T},i}.min(\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}) \tag{4.16}$$

4.2.6.11 Subjet b tagging

In the process of subjet b tagging, the b tagging procedure described in Sec. 4.2.6.6 is applied to the subjets identified by the soft-drop algorithm. This process is highly efficient in identifying boosted H boson decays, where $H \to b\bar{b}$[6].

4.2.6.12 W/Z boson tagging

This analysis uses CHS AK8 jets ($p_{\mathrm{T}} > 200$ GeV, $|\eta| < 2.4$) with the pruning algorithm to identify boosted W/Z bosons. AK8 pruned jet mass should be within $65 < \mathrm{M}_{pruned} < 105$ GeV and should only have two subjets. This is commonly known as the *W/Z tagging*. The performance of this method is evaluated by using a data sample enriched in $t\bar{t} + jets$ events, which contains hadronic W boson decays. Detailed information of the procedure can be found in Ref.[135]. The scale factors and related uncertainties for *jet mass scale* and *jet mass resolution* for this tagging method for given pruned mass window are computed by the CMS *Jet MET Algorithms and Reconstruction* (JMAR) group for different τ_2/τ_1 requirements[135;141].

4.2.6.13 *Higgs* tagging

This analysis uses CHS AK8 jets ($p_{\mathrm{T}} > 300$ GeV, $|\eta| < 2.4$) with the pruning algorithm to identify boosted H boson. AK8 pruned jet mass should be within $105 < \mathrm{M}_{pruned} < 135$ GeV and should have only two subjets. This is commonly known as *H tagging*. The subjet b-tagging method mentioned in Sec 4.2.6.11 has been used to identify b-jets for this tagger, both identified subjets are required to be identified as b-jets. This method gives 70-90% efficiency of identifying a b-tagged subjet depending on the p_{T} of the subjet and 10% misidentification rate for subjets from gluons and light-flavored quarks. Corrections to the jet mass scale and resolution are done according to Ref.[142].

4.2.6.14 *top* quark tagging

This analysis uses CHS AK8 jets ($p_{\mathrm{T}} > 400$ GeV, $|\eta| < 2.4$) with the softdrop algorithm to identify boosted *top* quarks. The AK8 softdrop jet mass should be within $105 < \mathrm{M}_{softdrop} < 220$ GeV with $\tau_3/\tau_2 < 0.81$. The scale factors and related uncertainties for *jet mass scale* and *jet mass resolution* for this tagging method for given softdrop mass window are computed by the CMS *Jet MET Algorithms and Reconstruction* (JMAR) group for different τ_3/τ_2 requirements[143]. In addition, the W/Z, H and top jets can be mistagged. Mistag scale factors and related uncertainties for such mistagged jets are derived and applied to both background and signal samples as described below.

4.2.7 Mistag rates for W/Z, H and top taggers

Regions that are orthogonal to signal region are known as control regions(see Sec. 5.4), where SM background processes provide the major contribution, and signal contributions are negligible. In such regions, the standard model backgrounds are not expected to have boosted objects such as W/Z, H and top jet candidates identified by the jet substructure techniques as mentioned in Sec. 4.2.6.7. But, the jet substructure taggers may still identify a small fraction of boosted jets in those regions. This is known as "mistag" of a boosted jet and the associated rates should be studied. The mistag rate per event is calculated by using the definition given by Eq. 4.17.

$$\text{Mistag rate} = \frac{\text{Number of jets passed the respective jet substructure tagger}}{\text{Total number of AK8 jets}} \tag{4.17}$$

A dedicated control region labelled as CR3, is defined in Sec. 5.4 to study this effect, where we do not expect the boosted W/Z, H and top jet candidates from SM backgrounds. This region is similar to signal region with $S_{\mathrm{T}} > 1000$ GeV but with zero b jets, where $S_{\mathrm{T}} = H_{\mathrm{T}} + p_{\mathrm{T}}(\mathrm{Z}) + \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$. The Fig. 5.14 shows the S_{T} distribution in this region where signal contamination is found to be 0.7 and 0.5 % respectively for $Z \to \mu^{+}\mu^{-}$ and $Z \to e^{+}e^{-}$ channels. The sub-jet b tagging requirement is removed from the H tagger to be compatible with zero b jet requirement.

In order to study the difference between data and background simulation, the p_{T} distributions of the W/Z-, H-, top-tagged and AK8 jets are investigated. DY sample is used for background simulation, as it provides the dominant background contribution in this analysis. The mistag rates for data and DY MC are calculated separately as shown in left side plots of Fig. 4.13. The mistag rate scale factors and their associated uncertainties are calculated by taking the ratio between data and DY MC mistag rates for the W/Z, H and top taggers separately and by performing a linear fit as there is no strong p_{T} dependence observed, which can be seen in right side plots of Fig. 4.13. The obtained scale factors and their uncertainties are shown in Table 4.9.

Table 4.9: *Derived mistag rate SFs and their related uncertainties.*

Tagger	SF
W tagger	1.08 ± 0.05
H tagger	0.98 ± 0.14
top tagger	1.04 ± 0.07

The mistag scale factors in Table 4.9 are applied to all background simulation as an additional event weight per tagged jet identified by the jet substructure taggers.

Mistag rate for signal samples - In order to match reconstructed boosted jets originated from the hard scattering process, we require ΔR between boosted jet and generator level particle to be < 0.8. If the condition is satisfied, appropriate efficiency scale factors are applied as mentioned in Sections 4.2.6.12, 4.2.6.13 and 4.2.6.14, otherwise, appropriate mistag scale factors from Table 4.9 are applied.

boosted jet overlapping in signal and efficiency scale factor - Due to the small overlapping mass window between H and top tagger, small fraction of jets can be tagged by both taggers. In such cases, corresponding scale factors are applied by first performing ΔR(reco jet, gen particle) < 0.8 matching and confirming the jet to be either a top or H-tagged AK8 jet.

4.2.8 Pileup reweighting

Dependency of observables on pileup interactions, remaining after CHS procedure and pileup corrections, are accounted by applying the *pileup reweighting* procedure[6]. The scale factors are derived by using the distribution of number of pileup vertices in both data and simulation. The distribution of number of pileup vertices (μ) for data is obtained by using instantaneous luminosity and total inelastic cross section (69.2 mb)[6]. The similar distribu-

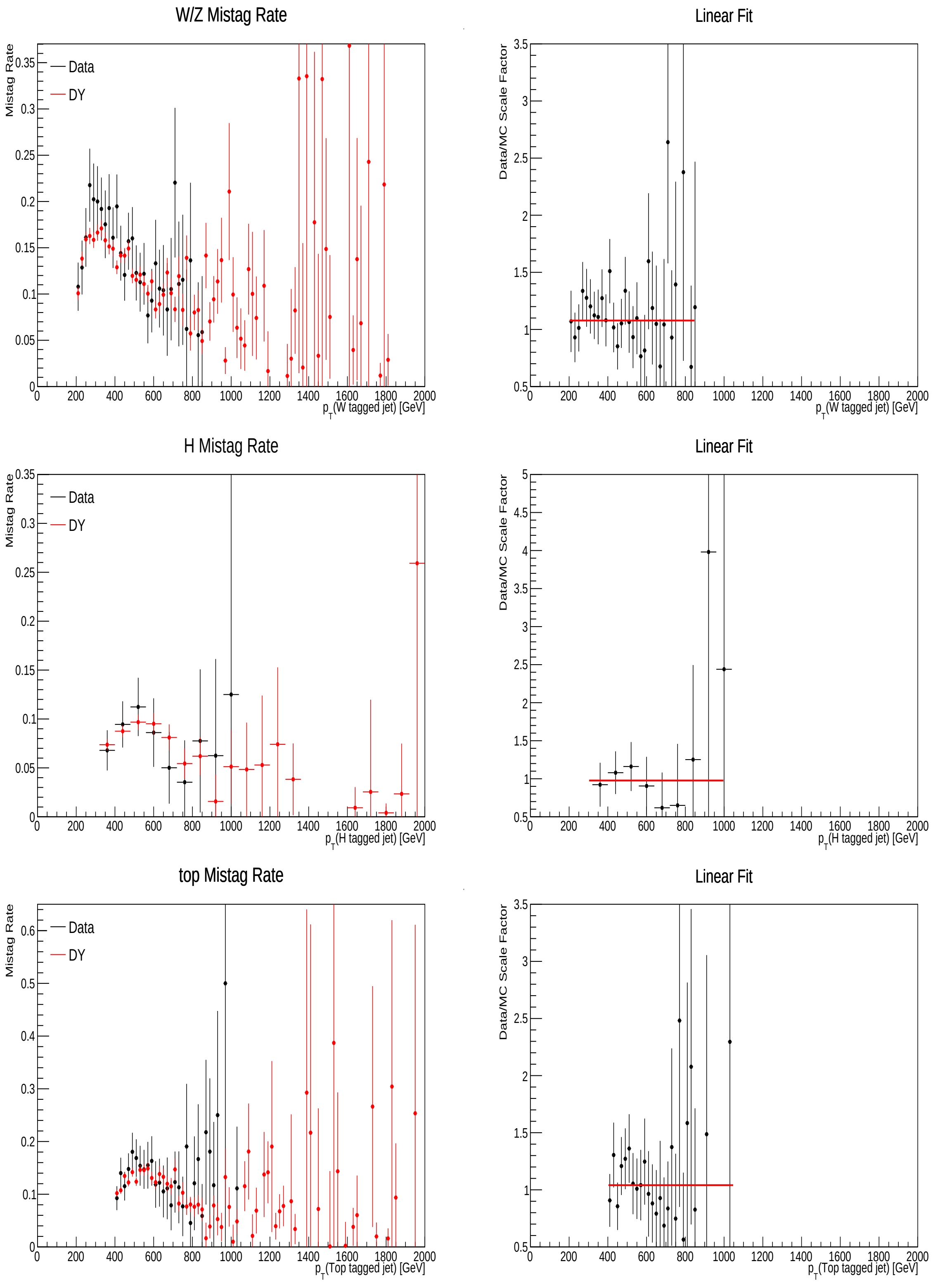


Figure 4.13: *The mistag rates for W/Z, H and top taggers measured as a function of* p_{T} *for both data and DY MC (left) and the data/MC scale factors (right).*

tion for simulation is obtained by producing a Poisson distribution that have a mean value of μ, which is also known as *true* pileup distribution. The ratio between the distributions of data and simulation provides the scale factors, which depend on number of pileup vertices present in an event [6].

These scale factors are finally applied to simulated events based on the number of pileup vertices found in simulation. The uncertainty of the procedure is calculated by varying the inelastic cross section 69.2 mb by 5 % up and down and repeating the same procedure.

Chapter 5

Event selection and control regions

5.1 Introduction

In this chapter, we describe the search for pair produced vector-like top quarks ($\mathrm{T\bar{T}}$ pairs), using the proton-proton collision data collected by the CMS experiment in 2016 at a centre-of-mass energy $\sqrt{s} = 13$ TeV, which corresponds to a total integrated luminosity of 35.9 fb^{-1}.

The search is performed assuming that one of the T quark decays into a third generation top quark and a Z boson, where the Z boson decays into two opposite-sign leptons, e^+e^- or $\mu^+\mu^-$. The other T quark might decay into any of the three possible states tZ, tH or bW, where *top* and *bottom* quarks and W, Z and H bosons decay hadronically and are reconstructed as jets. Fig. 5.1 shows the Feynman diagram for $\mathrm{T\bar{T}}$ production. Events with two opposite-sign leptons consistent with coming from a Z boson decay, $Z \rightarrow e^+e^-$ or $\mu^+\mu^-$ and jets are considered in this analysis.

The VLQ gauge group determines the final decay channels of $\mathrm{T\overline{T}}$ production and corresponding benchmark branching ratios($\mathcal{B}$)[27]. Since a massive T quark is assumed to couple to third generation SM quarks, only three decay modes are possible. The $\mathcal{B}$s are chosen using a benchmark model, where mass differences between Z, W and H bosons are assumed to be small and should satisfy the relation, $\mathcal{B}(T \rightarrow bW) + \mathcal{B}(T \rightarrow tZ) + \mathcal{B}(T \rightarrow tH) = 1$[27] as described in Sec. 2.3.3. Combinations of all possible $\mathcal{B}$s are scanned to cover the full phase

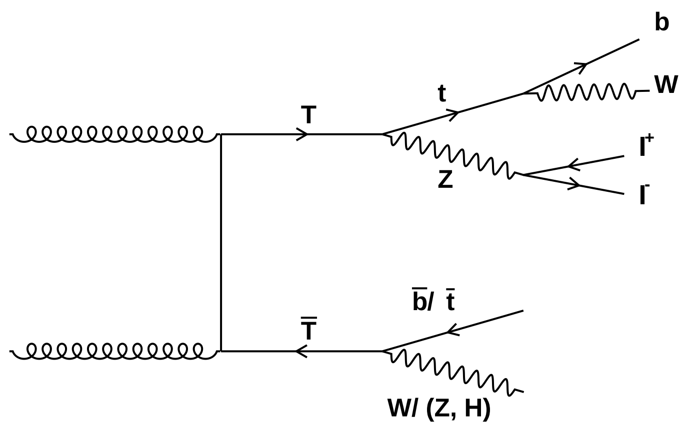


Figure 5.1: *Feynman diagram of* $\mathrm{T\bar{T}}$ *production, where one of the T quark decays into a top quark and Z boson and the Z boson decays into opposite-sign lepton pair ($Z \rightarrow l^+l^-$) (shown at the upper branch of the diagram). The other T quark can decay into any of the three states $T \rightarrow tZ, tH$ or bW. The diagram was produced by using ROOT framework*[32].

space.

Since T quarks with masses below 700 GeV are excluded by the Run 1 analyses as described in Sec. 2.3.4, T quarks with masses above 700 GeV are considered in the current analysis. In this mass regime, decay products that are originated from the T quarks have a fair amount of Lorentz boost. Therefore, they have high transverse momentum and are collimated. Collimated particles are close to each other in the $\eta - \phi$ plane with a low ΔR.

Fig. 5.2 shows the $\mathrm{T\overline{T}}$ signal distributions for ΔR between two opposite sign leptons in $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ channels after event preselection (see sec. 5.3.4 for more detailed description on event pre-selection). The signal processes shown here correspond to three different T quark masses, $m_T = 800, 1300$ and 1800 GeV in the $\mathrm{T\bar{T}} \rightarrow$ tZtZ decay channel. According to Fig. 5.2, most of the pre-selected signal events lie in the $\Delta R(l^+, l^-) < 0.5$ region, for both $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ channels. This suggests that the Z bosons originated from the massive T quarks are boosted. The $\Delta R(l^+, l^-) < 0.5$ region is known as the *boosted Z regime*. Also, the peak of the $\Delta R(l^+, l^-)$ distribution moves toward low ΔR regime with increasing T quark masses. This suggests that the Z bosons originated from

T quarks with larger masses become more boosted, where more pre-selected events tend to have a $\Delta R(l^+, l^-) < 0.5$. The scope of this analysis is to search for the $\mathrm{T\overline{T}}$ pairs in the *boosted Z regime.*

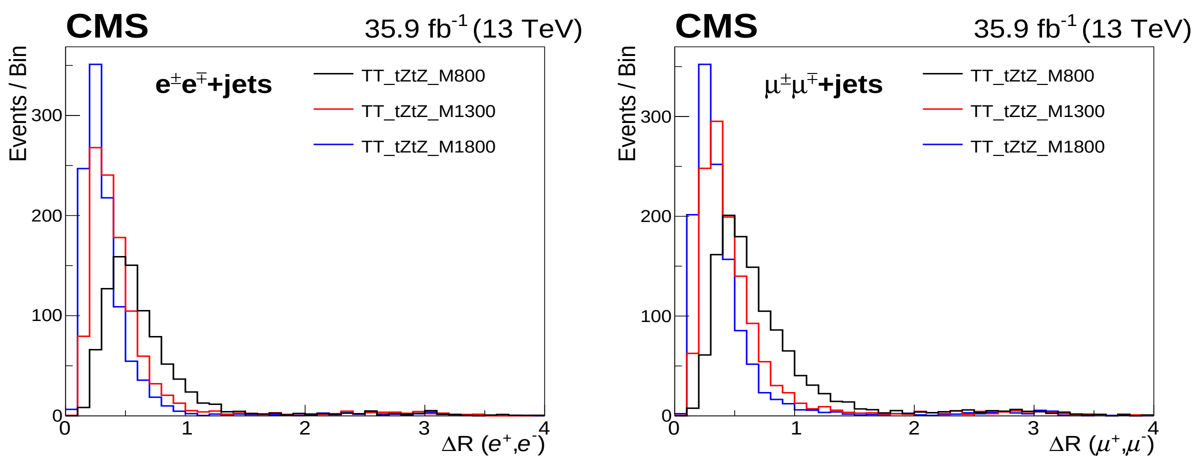


Figure 5.2: $\mathrm{T\overline{T}}$ *signal distributions for* ΔR *between two opposite sign leptons in* $Z \to e^+e^-$ *(left) and* $Z \to \mu^+\mu^-$ *(right) channels after event preselection. The signal processes correspond to three different T quark masses,* $m_T = 800, 1300$ *and* 1800 GeV *in the* $\mathrm{T\bar{T}} \to$ tZtZ *decay channel. The signals are normalized to 1 pb cross section for better visualization.*

Many standard model (SM) processes may have similar signatures as $\mathrm{T\overline{T}}$ events. The main contribution comes from the Drell–Yan (Z/γ^*)+jets (DY+jets) production, where Z boson decays into two opposite-sign leptons, e^+e^- or $\mu^+\mu^-$ pair. This signature is similar to the Z boson originated from one of the massive T quark and acts as the main background for the current search. Smaller background contributions come from $\mathrm{t\bar{t}}$+jets and $\mathrm{t\bar{t}Z}$ production, where signatures are similar to t and $\bar{\mathrm{t}}$ quarks and Z bosons originated from the massive T quarks. Much smaller backgrounds are the SM diboson production: WW, ZZ and WZ, and SM rare processes such as tZq, tWZ, and $\mathrm{t\bar{t}W}$ productions. These background processes are simulated using MC generators described in Sec. 4.1 and are discussed in Sec. 4.1.4 in detail. Their contributions and kinematics are studied to separate them from the signal process.

The strategy of this search is to use pre-selected events, with two opposite sign leptons and jets, and then to reconstruct all hadronically-decaying objects that are originated from the $\mathrm{T\overline{T}}$ decay channel. This involves reconstruction of the W, Z and H bosons and top and bottom quarks. As mentioned earlier, the decay products from T quarks have a fair amount of Lorentz boost and hence have a high transverse momentum compared to decay products from SM processes. Therefore, a variable, sensitive to the transverse momentum (p_{T}) of the reconstructed final state objects is a good discriminator between the signal and SM background events. The variable S_{T} is used for this purpose, which is defined as the sum of p_{T} of all the final state objects in an event. This includes leptons, jets and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$).

Fig. 5.3 shows the S_{T} distributions for the SM background and signal processes in the $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels after event pre-selection (see Sec. 5.3.4 for more detailed description on event pre-selection). The signal processes shown here correspond to three different T quark masses, $m_T = 800, 1300$ and 1800 GeV in the $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$ decay channel. Area of the signal and background distributions are normalized to 1 pb and 4 pb respectively for better visualization.

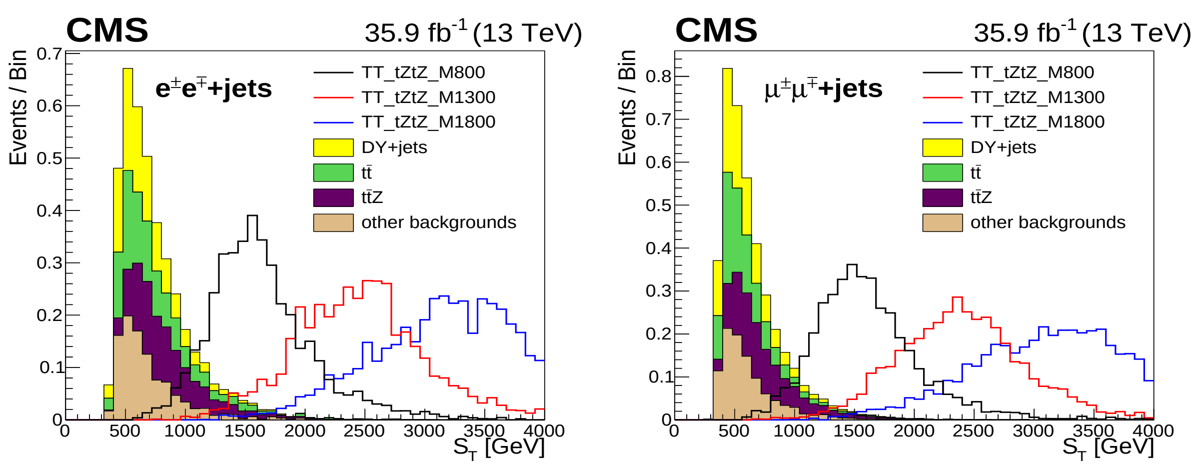


Figure 5.3: *S_{T} distributions for SM background and signal processes in $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) channels after event preselection. The signal processes shown here correspond to three different T quark masses, $m_T = 800, 1300$ and 1800 GeV in the $\mathrm{T\overline{T}} \to$ tZtZ decay channel. The area of the signal and background distributions are normalized to 1 pb and 4 pb respectively for better visualization. "other backgrounds" represents the combination of smaller background contributions such as dibosons, $\mathrm{t\bar{t}W}$, tZq and tWZ*

From Fig. 5.3, it can be seen that most of the SM background events have $S_{\mathrm{T}} < 1000$ GeV. More than 95% of the signal events have $S_{\mathrm{T}} > 1000$ GeV and the peak shifts towards the high S_{T} regime with increasing T quark mass. Hence, S_{T} can be used as the discriminating variable between the signal and SM backgrounds. Since $\mathrm{T\overline{T}}$ decay consists of many hadronically decaying candidates, the events can be categorized according to the number of such reconstructed candidates, top (N_t), W/Z (N_V), Higgs (N_H), and b jets (N_b) present in an event. The search is optimized using the S_{T} distribution in those event categories. This procedure is described in Chapter 6 in detail.

5.2 Datasets

5.2.1 Collision data

The proton-proton collision data used in the analysis were collected by the CMS detector at centre-of-mass energy $\sqrt{s}$ = 13 TeV in 2016 with 25 ns bunch spacing. The primary data sets to identify two oppositely charged leptons are chosen with events having one electron (**SingleElectron**) and one muon (**SingleMuon**) data streams at the HLT level.

Due to the inefficiency of reconstructing high p_{T} electrons ($p_{\mathrm{T}} > 300$ GeV) by the HLT trigger used for **SingleElectron** data stream, the data stream with events having one photon (**SinglePhoton**) is also added to the **SingleElectron** data stream. The HLT triggers used for the analysis are discussed in Sec. 5.3.1.

Data events are selected, when all the sub-detectors of the CMS detector are fully functional, including the superconducting solenoid. This is to verify that the data events selected for the analysis are of high quality and without any detector malfunction. A set of additional filters are used to remove events with detector noises and signal interferences that are not originated from the hard scattering processes. These include:

- Events, which do not have at least one good primary vertex reconstructed, as described in Sec. 4.2.1.3.
- Events with pre-identified noise patterns in HB and HE.
- Events with pre-identified noise patterns in EB and EE.

- Events with masked ECAL crystals, where energy cannot be detected.
- Events with dead tracker cells, where detected energy from the ECAL superclusters cannot be correlated.
- Events where proton beam interacts with residual gas, which is commonly known as a *beam halo*.

All the generated signal and background MC events described in sections 4.1.3 and 4.1.4 are simulated using the GEANT4 software to model the response of the CMS detector [78;144]. The same software configurations are used for both data and simulation to reconstruct events. The effect of pileup interactions are simulated using PYTHIA 8.1 generator as described in Sec. 4.2.8 with a total inelastic proton-proton cross section of 69.2 mb [145].

In text, legends or tables, the term "other backgrounds" represents the combination of smaller background contributions such as dibosons, $\mathrm{t\bar{t}W}$, tZq and tWZ. From this point onward, in plots/tables, only exclusive $\mathrm{T\bar{T}} \rightarrow$ tZtZ, $\mathrm{T\bar{T}} \rightarrow$ tZtH and $\mathrm{T\bar{T}} \rightarrow$ tZbW decays are used to represent the signal without considering the real mixtures of events due to $\mathcal{B}$s. All the background and signal MC samples are normalized to their cross sections, unless stated otherwise. The hashed band shown in plots represents the statistical and systematic uncertainties added in quadrature. The systematic uncertainties considered for the analysis are discussed in Chapter 7.

5.3 Event Selection

This section describes the event selection for the $\mathrm{T\overline{T}}$ search. The main purpose of the event selection is to choose events that are only relevant to $\mathrm{T\overline{T}}$ search and to study and reduce the contributions from SM background as much as possible. So-called *event pre-selection*, described in Sec. 5.3.4, is used to reduce the number of events that are used for the analysis. Subsequent selections, so-called *signal selections* are used on the pre-selected events to define a boosted regime, where the search for $\mathrm{T\overline{T}}$ pairs is performed. Finally, these selected events according to signal selection criteria are used to optimize the $\mathrm{T\overline{T}}$ search using the S_{T} variable.

5.3.1 Trigger and lepton selection

The trigger selection is the first step of event selection. Several triggers are studied, which provide the maximum signal efficiency after trigger selection. As the search uses events with two opposite sign leptons, the HLT triggers (described in Sec. 3.2.7) that select events according to presence of reconstructed leptons are considered for the analysis. The triggers studied are,

- Double Muon (**Mu17 Mu8**) - Events with two isolated (see lepton isolation criteria described in Sec. 4.2.3) tracker muons (see Sec. 4.2.2), with minimum transverse momenta of 17 and 8 GeV.

- Double Electron (**Ele23 Ele12**) - Events with two isolated electrons identified using the information from tracker and ECAL, with minimum transverse momenta of 23 and 12 GeV.

- Single Muon (**Mu24**) - Events with an isolated muon, either identified as a tracker muon or a global muon with minimum transverse momentum of 24 GeV.

- Single Electron/ Single Photon (**Ele115 Pho175**) - Events with an electron identified using information from tracker and ECAL with minimum transverse momentum of 115 GeV, or a photon identified with minimum transverse momentum of 175 GeV.

The primary data sets with events having two electrons and muons are used to study **Ele23 Ele12** and **Mu17 Mu8** triggers respectively, while primary data sets with events having one electron, photon and muon are used to study **Ele115 Pho175** and **Mu24** triggers respectively. As mentioned in Sec. 5.1, the signal region in this analysis represents the boosted Z regime and hence, pair of leptons originating from a Z boson decay should have a fair amount of Lorentz boost. Hence, transverse momentum of the leading and second-leading leptons (the most energetic leptons reconstructed within an event) are chosen such that the signal efficiency is preserved and it should be compatible with the trigger choices. The leptons selected according to minimum p_{T} requirements are given in Table 5.1.

Table 5.1: *Selection on lepton p_T for different triggers.*

Lepton p_T	**Ele23 Ele12**	**Mu17 Mu8**	**Ele115 Pho175**	**Mu24**
Leading lepton p_T [GeV]	≥ 45	≥ 45	≥ 120	≥ 45
Second leading lepton p_T [GeV]	≥ 25	≥ 25	≥ 25	≥ 25

The isolation requirement at the trigger level affects the signal acceptance. Therefore, the $\Delta R(l^+, l^-)$ distribution after the event pre-selection (see Sec. 5.3.4) is used to determine the most optimal trigger that preserves the signal efficiency. Fig. 5.4 shows the $\Delta R(l^+, l^-)$ distributions for triggers, **Ele23 Ele12**, **Mu17 Mu8**, **Ele115 Pho175** and **Mu24** after event pre-selection. The $T\overline{T}$ processes shown here correspond to three different T quark masses, $m_T = 800, 1000$ and 1200 GeV in the $T\bar{T} \rightarrow tZtZ$ decay channel. The signal processes are normalized to 1 pb cross section for better visualization. The number of signal events in the boosted Z regime, with $\Delta R(l^+, l^-) < 0.3$ are shown in Table. 5.2.

According to the Fig. 5.4 and Table 5.2, it is seen that there are more signal events preserved for the triggers, **Ele115 Pho175** and **Mu24** than, **Ele23 Ele12** and **Mu17 Mu8**. Therefore, **Ele115 Pho175** and **Mu24** triggers are selected as the default triggers for the current analysis. The photon trigger (**Pho175**) is used to capture electrons with $p_T > 300$ GeV. It recovers the electrons that are not captured by the electron trigger (**Ele115**) alone. Combining the **Pho175** with **Ele115** recovers 2.4% of the data events and 0.4% of signal events for $T\bar{T} \rightarrow tZtZ$ decay channel.

Table 5.2: *Number of signal events in the boosted Z regime. The signal processes correspond to three different T quark masses, $m_T = 800, 1000$ and 1200 GeV in the $T\overline{T} \rightarrow tZtZ$ decay channel. The signal processes are normalized to 1 pb cross section for better visualization.*

Signal sample	**Ele23 Ele12**	**Mu17 Mu8**	**Ele115 Pho175**	**Mu24**
$T\bar{T} \rightarrow tZtZ$, $m_T = 800$ GeV	191.96	203.64	201.34	230.23
$T\bar{T} \rightarrow tZtZ$, $m_T = 1000$ GeV	360.15	388.42	383.93	405.16
$T\bar{T} \rightarrow tZtZ$, $m_T = 1200$ GeV	486.56	533.01	510.51	548.37

The signal acceptance after applying trigger requirements is shown in Fig. 5.5 for T quark

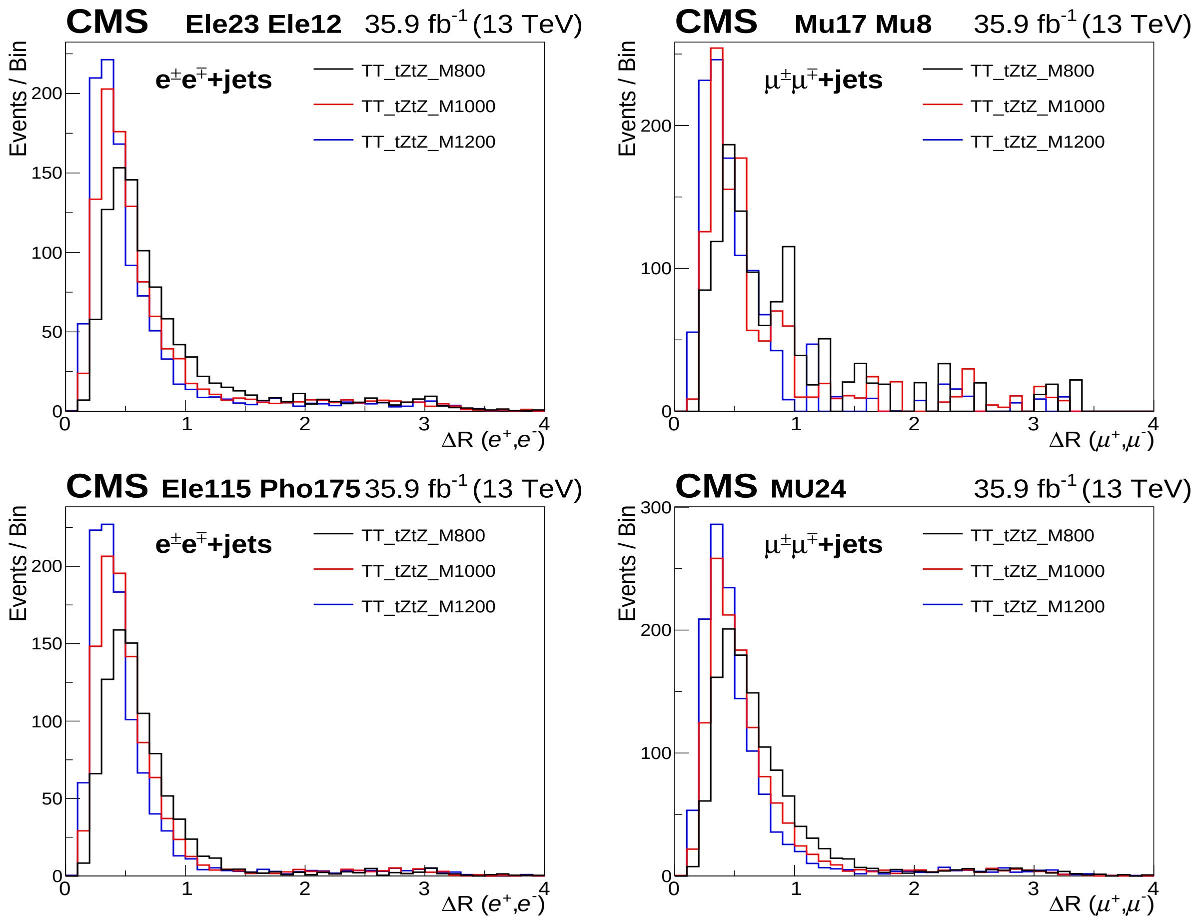


Figure 5.4: $\mathrm{T\overline{T}}$ *signal distributions for* ΔR *between two opposite sign leptons in* $Z \to e^+e^-$ *(left) and* $Z \to \mu^+\mu^-$ *(right) channels after event preselection. The signal processes shown here correspond to T quark masses,* $m_T = 800, 1000$ *and* 1200 GeV *in the* $\mathrm{T\overline{T}} \to tZtZ$ *decay channel. The signal processes are normalized to 1 pb cross section for better visualization.*

masses 900, 1100, 1300, 1500 and 1700 GeV for both electron and muon channels. The signal acceptances for both electron and muon channels remain nearly constant over the T mass range considered in the analysis.

These trigger requirements are applied to both data and simulation. Any mismatch between data and simulation is corrected by trigger efficiency scale factors. These scale factors are obtained as a function of lepton p_{T} and η by using the *tag-and-probe method*[105]. They are applied to simulated samples.

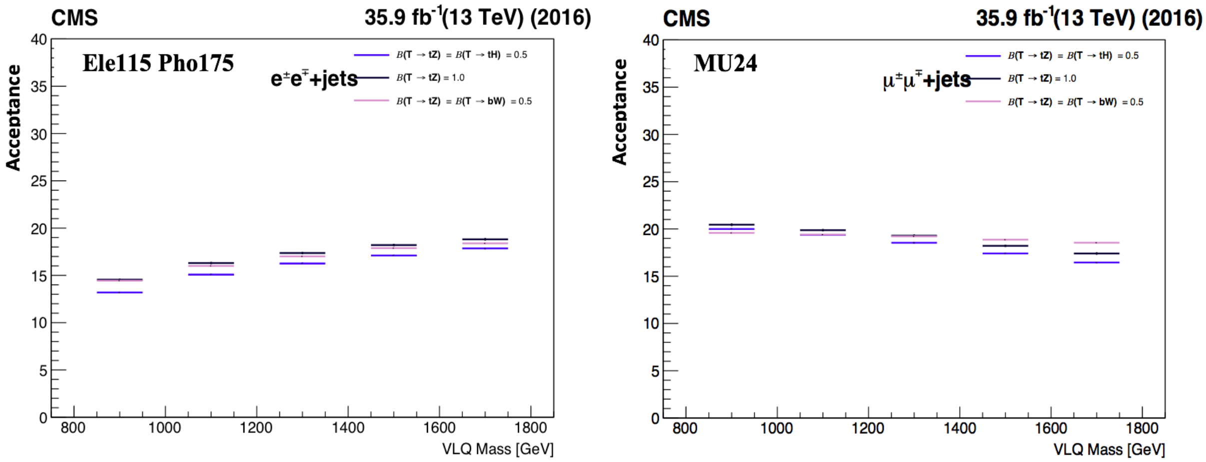


Figure 5.5: *Signal acceptance obtained for **Ele115 Pho175** (left) and **Mu24** (right) triggers as a function of T mass. Branching ratios shown in the legend represent the three benchmark $\mathcal{B}$s mentioned in Sec. 5.1.*

5.3.2 Dilepton invariant mass

Two electrons (muons) are selected with $p_{\mathrm{T}} \geq 120$, 25 GeV ($p_{\mathrm{T}} \geq 45$, 25 GeV) as mentioned in Sec. 5.3.1, that have opposite charges, should satisfy the dilepton invariant mass selection, $75 < M(l^+l^-) < 105$ GeV to assure that they are originating from a Z boson decay. The Z boson candidate is required to have $p_{\mathrm{T}}(\mathrm{Z}) > 100$ GeV. Events that have exactly one such Z boson are selected in this analysis.

5.3.3 Number of jets

The rest of the particles originated from T quarks decay hadronically and reconstructed as jets. Hence, events are required to have ≥ 3 AK4 jets (see Sec. 4.2.6). Furthermore, the jets are expected to have a fair amount of Lorentz boost as they are originated from heavy T quarks. The variable H_{T} defined as, $H_{\mathrm{T}} = \sum_{Ak4jets}(p_{\mathrm{T}})$ is required to be ≥ 200 GeV to select boosted jets.

5.3.4 Event pre-selection

The events selected according to the trigger and dilepton invariant mass selections, and number of AK4 jets described in Sections 5.3.1, 5.3.2 and 5.3.3 are considered in this analysis. This selection criterion is known as event *pre-selection* and is summarized in Table 5.3.

Table 5.3: *Summary of event pre-selection.*

Variable	Selection
Trigger selection	As described in Sec. 5.3.1
Z boson mass	$75 < M(Z) < 105$ GeV
$N(\mathrm{Z} \to \mathrm{l}^{+}\mathrm{l}^{-})$	$= 1$
$p_{\mathrm{T}}(\mathrm{Z} \to \mathrm{l}^{+}\mathrm{l}^{-})$	> 100 GeV
N(AK4 jets)	≥ 3
H_{T}	≥ 200 GeV

5.3.5 Definition of the signal regionl

As the decay products are originated from a heavy T quark, a variable that is sensitive to p_{T} has more discriminating power to discover the new signal (see Sec. 5.1). Hence, the S_{T} variable defined as $S_{\mathrm{T}} = H_{\mathrm{T}} + p_{\mathrm{T}}(\mathrm{Z}) + \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ is used. In general, it has a good discriminating power to distinguish signal and SM background as can be seen in Fig. 5.3.

In order to preserve the signal efficiency (especially for $\mathrm{m_T} < 1300$ GeV), signal region for the search is defined as events having $S_{\mathrm{T}} \geq 1000$ GeV. Furthermore, the leading and sub leading AK4 jets are required to have $p_{\mathrm{T}} > 100$ and > 50 GeV respectively plus at least one b jet is required. This is defined as *signal selection* and summarized in Table 5.4.

Table 5.4: *Summary of signal selection.*

Variable	Selection
Leading AK4 jet p_{T}	> 100 GeV
2nd leading AK4 jet p_{T}	> 50 GeV
N(b jets)	≥ 1
S_{T}	≥ 1000 GeV

The *control regions* are used to validate the background MC simulations and are discussed in Sec. 5.4 in detail. These regions are defined such that they are orthogonal to signal region and, where the contribution from $\mathrm{T\overline{T}}$ signal is negligible. One of the requirements of the signal selection listed in Table 5.4 is inverted to define these regions. For example, requiring $S_{\mathrm{T}} < 1000$ GeV or $N(\mathrm{b\ jets}) = 0$ defines control regions, where main contribution comes from the SM background processes. These control regions can be used to study the modeling of the SM backgrounds in simulation.

5.3.6 Event cutflow

Fig. 5.6 and Table. 5.5 present the event cut flow for both data and simulation after event pre-selection and subsequent signal selection steps for both $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ channels. The data distributions are shown with black points with vertical and horizontal bars. The background distributions are presented as shaded histograms, while the signal distributions are shown in solid lines. The signal samples shown here are normalized to their expected cross sections given in Table 4.1. The hashed band shows the statistical and systematic uncertainties on the total background contribution.

There is a good agreement between data and background simulation within the total uncertainty after event pre-selection and for subsequent selections including the signal region, which is the last bin in Fig. 5.6.

5.3.7 Kinematics after event pre-selection

Fig. 5.7 shows the invariant mass distribution of the two opposite sign leptons $\mathrm{M}(\mathrm{Z} \rightarrow l^+l^-)$, the p_{T} of the Z boson candidate $p_{\mathrm{T}}(\mathrm{Z} \rightarrow l^+l^-)$ and the separation between the two leptons $\Delta R(l^+, l^-)$ after the event pre-selection. The $\mathrm{M}(\mathrm{Z} \rightarrow l^+l^-)$ distribution has a peak around Z boson mass of 91.2 GeV, confirming that two opposite sign leptons are originated from a Z boson. The mismodeling seen in both dilepton invariant mass and the tail of Z boson candidate p_{T} distributions in the $Z \rightarrow e^+e^-$ channel is due to the electron p_{T} resolution. Most of the bins in these distributions are within the systematic uncertainty. In addition, the p_{T} distribution of the two leptons are shown in Fig. 5.8, which has a good agreement

Table 5.5: *Event counts after the pre-selection and signal selection for $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels. Signal MC event yields are for a T quark with mass 1000 GeV for there decays* $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$, $\mathrm{T\bar{T}} \to \mathrm{tZtH}$, *and* $\mathrm{T\bar{T}} \to \mathrm{tZbW}$.

$Z \to e^+e^-$ **channel**								
Cuts	Data	DY	$\mathrm{t\bar{t}}$	$\mathrm{t\bar{t}Z}$	other backgrounds	$M(\mathrm{T})$1000 GeV		
						TT $\to$ tZtZ	TT $\to$ tZtH	TT $\to$ tZbW
Preselection	14315	14295.4 ± 54.2	221.8 ± 6.5	94.4 ± 0.5	361.8 ± 12.7	44.7 ± 0.9	22.1 ± 0.5	22.5 ± 0.5
1st jet $p_{\mathrm{T}} > 100$ GeV	13325	13346.9 ± 51.9	199.9 ± 6.2	84.8 ± 0.5	332.5 ± 12.2	44.6 ± 0.9	22.1 ± 0.5	22.5 ± 0.5
2nd jet $p_{\mathrm{T}} > 50$ GeV	11702	11843.5 ± 48.0	187.6 ± 6.0	83.0 ± 0.5	295.3 ± 11.5	44.5 ± 0.9	22.1 ± 0.5	22.5 ± 0.5
N(b jet) ≥ 1	1708	1499.2 ± 17.0	131.8 ± 5.0	57.5 ± 0.4	55.2 ± 4.6	35.1 ± 0.8	19.0 ± 0.4	16.9 ± 0.4
$S_{\mathrm{T}} > 1000 GeV$	308	233.4 ± 4.0	28.6 ± 2.4	18.6 ± 0.2	10.2 ± 2.0	35.8 ± 0.8	18.6 ± 0.4	17.1 ± 0.4
$Z \to \mu^+\mu^-$ **channel**								
Cuts	Data	DY	$\mathrm{t\bar{t}}$	$\mathrm{t\bar{t}Z}$	other backgrounds	$M(\mathrm{T}) = 1000$ GeV		
						TT $\to$ tZtZ	TT $\to$ tZtH	TT $\to$ tZbW
Preselection	44947	41410.7 ± 100.0	1531.2 ± 17.7	222.9 ± 0.8	938.3 ± 20.6	55.0 ± 1.0	29.9 ± 0.5	29.0 ± 0.5
1st jet $p_{\mathrm{T}} > 100$ GeV	39738	37243.5 ± 94.3	1268.7 ± 16.1	188.5 ± 0.7	834.5 ± 19.5	54.9 ± 1.0	29.9 ± 0.5	29.0 ± 0.5
2nd jet $p_{\mathrm{T}} > 50$ GeV	34858	32948.8 ± 87.4	1184.8 ± 15.6	183.9 ± 0.7	728.2 ± 18.2	54.7 ± 1.0	29.9 ± 0.5	29.0 ± 0.5
N(b jet) ≥ 1	5299	4106.7 ± 30.9	827.5 ± 12.8	126.4 ± 0.6	115.2 ± 6.4	43.6 ± 0.9	26.0 ± 0.5	21.5 ± 0.4
$S_{\mathrm{T}} > 1000$ GeV	571	432.3 ± 5.5	77.7 ± 3.9	29.0 ± 0.3	9.3 ± 1.6	44.2 ± 0.9	25.5 ± 0.5	21.8 ± 0.5

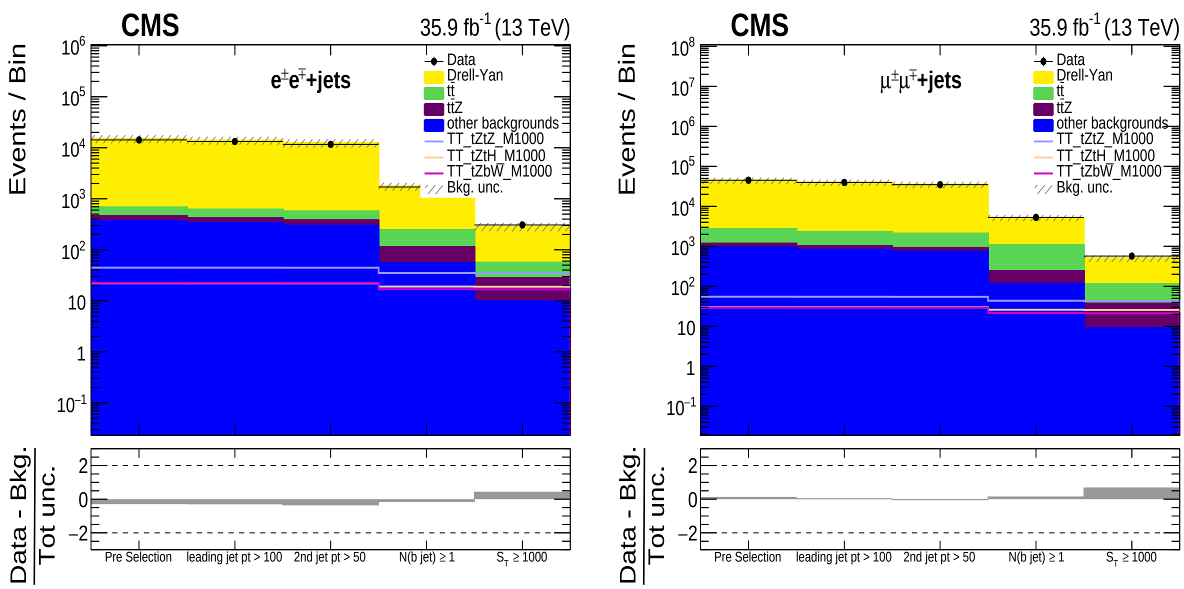


Figure 5.6: *The event cutflow after preselection and subsequent signal selection requirements for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) channels. The hashed band shows the statistical and systematic uncertainties.*

between data and background MC simulation within the total uncertainty.

Fig. 5.9 shows the distributions for E_T^{miss}, AK4 jets multiplicity, H_T and S_T for $Z \to e^+e^-$and $Z \to \mu^+\mu^-$channels. All of these distributions have a good agreement between data and background MC simulation.

5.4 Validation of MC simulation in Control regions

Control regions orthogonal to the signal region are defined to validate the data modeling of background MC simulation and to derive any scale factors to correct the MC simulation if necessary. The main background contribution comes from the DY+jets process, where Z boson decays into two opposite-sign lepton pair (e^+e^- or $\mu^+\mu^-$). The other background contributions such as $t\bar{t}$, $t\bar{t}Z$, ZZ, WZ, WW, tZq, tWZ, and $t\bar{t}W$ are small. The contribution from the standard model single top quark, W+jets and QCD multi jet production processes are found to be negligible. The background yields for the analysis are estimated based on

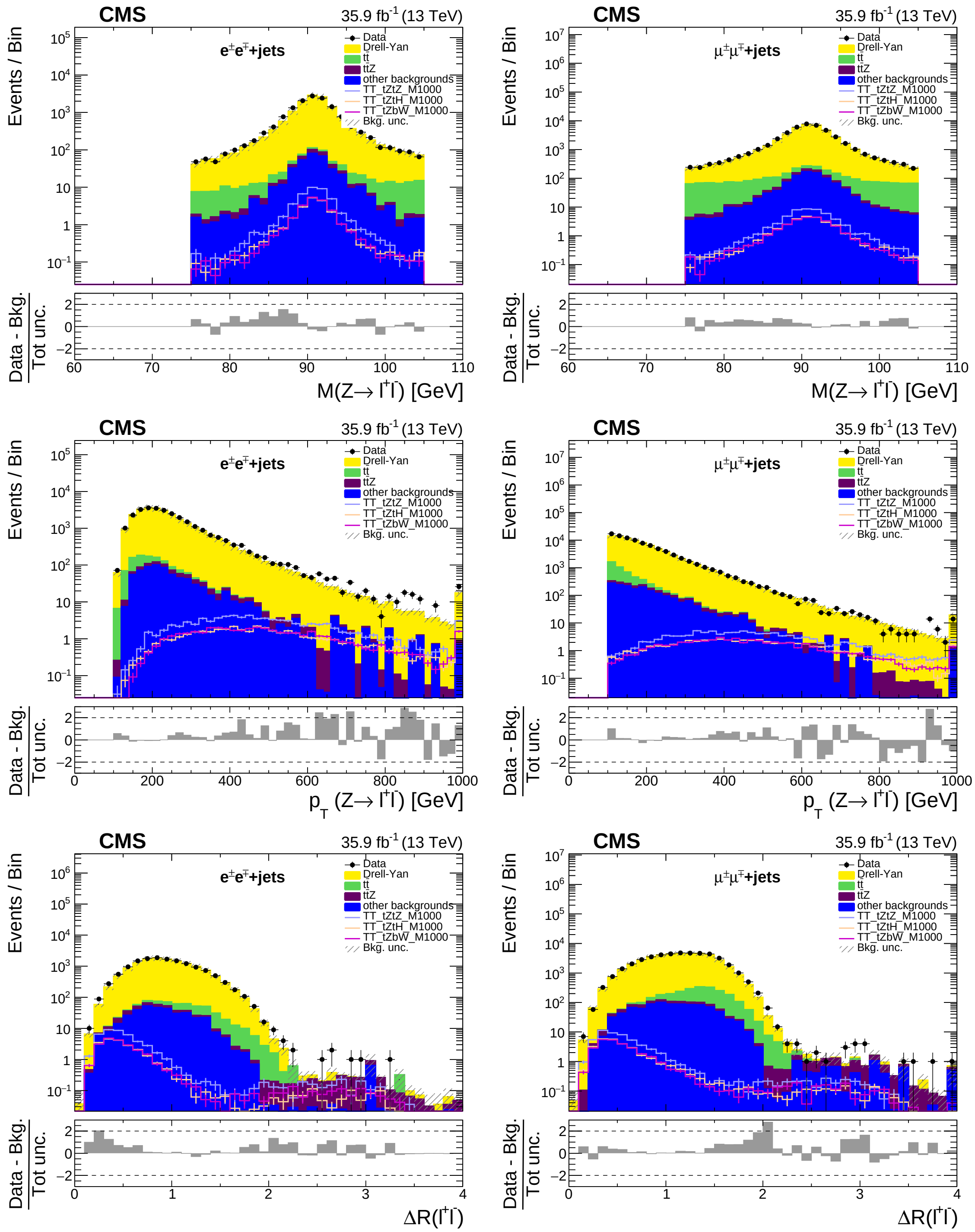


Figure 5.7: *Distributions for dilepton invariant mass, p_{T} of the Z boson candidate and $\Delta R(l,l)$ for $Z \to e^+e^-$(left) and $Z \to \mu^+\mu^-$(right) channels after event pre-selection. The hashed band shows the statistical and systematic uncertainties.*

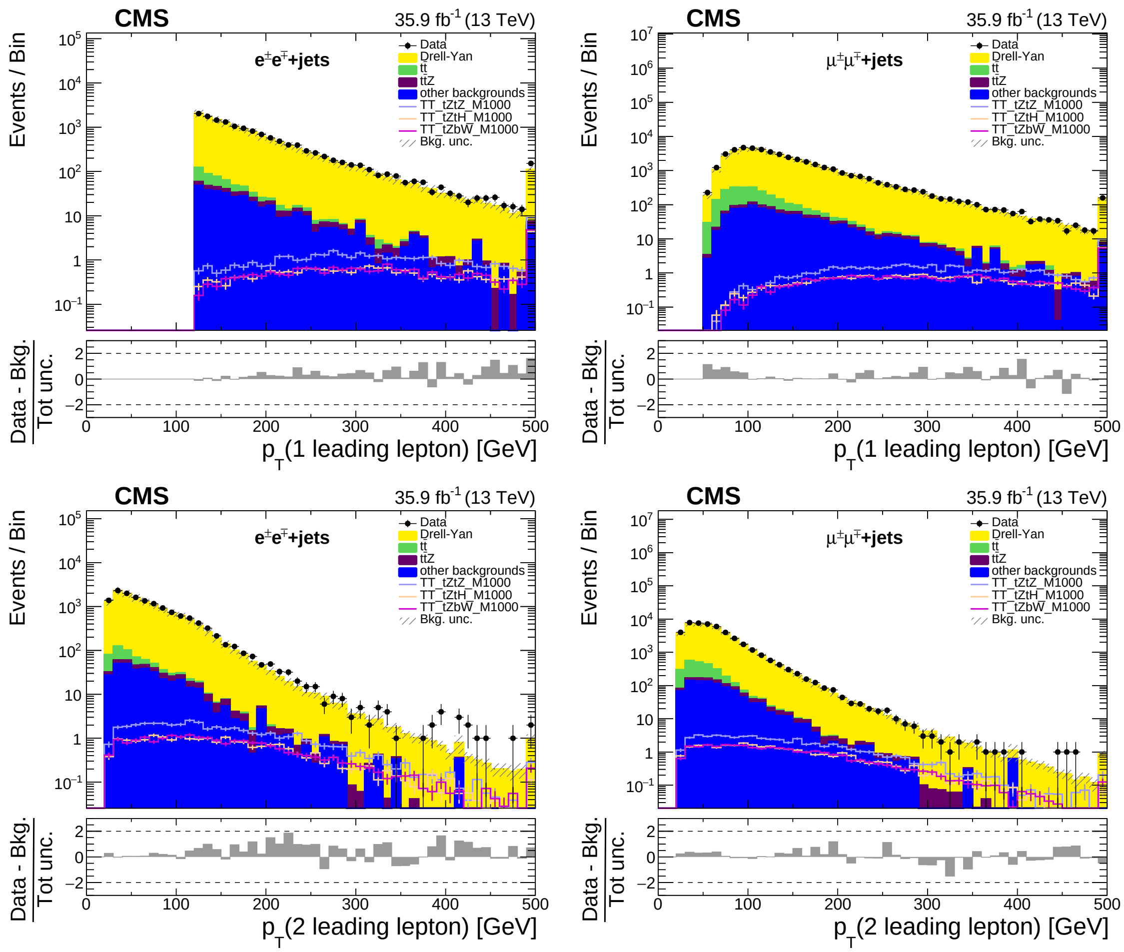


Figure 5.8: *Distributions of leading (first row) and second leading (second row) lepton* p_{T} *in* $Z \rightarrow e^{+}e^{-}$ *(left) and* $Z \rightarrow \mu^{+}\mu^{-}$ *(right) channels after event pre-selection. The hashed band shows the statistical and systematic uncertainties.*

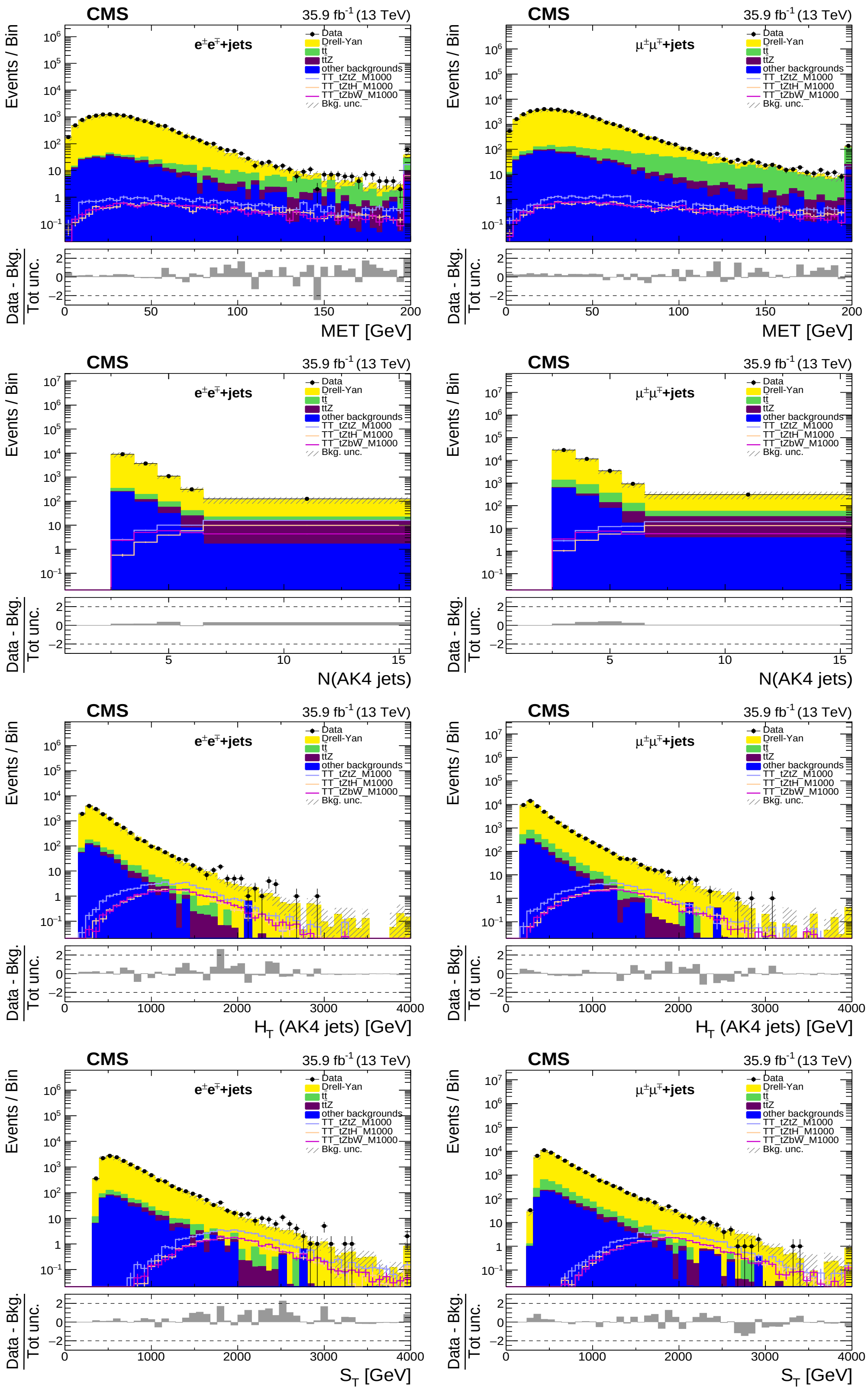


Figure 5.9: *Distributions for* E_{T}^{miss}*, AK4 jets multiplicity,* H_{T} *and* S_{T} *for* $Z \to e^{+}e^{-}$ *(left) and* $Z \to \mu^{+}\mu^{-}$ *(right) channels after event pre-selection. The hashed band shows the statistical and systematic uncertainties.*

the MC predictions.

There are three main control regions defined as CR2, CR3 and CR4 for the current analysis as mentioned in Table 5.6 to validate the modeling of data by simulation.

Table 5.6: *Selection criteria for the control regions CR2, CR3 and CR4.*

Selection	CR2	CR3	CR4
Pre-selection	Yes	Yes	Yes
N(bjets)	$= 0$	$= 0$	≥ 1
S_{T}	≥ 0 GeV	≥ 1000 GeV	≤ 1000 GeV

5.4.1 Control region CR2

Control region CR2 is enriched with DY background. It is defined by requiring $N(\text{bjets}) = 0$ and loosening the S_{T} requirement ($S_{\mathrm{T}} \geq 0$ GeV). This region is mainly used to check the DY modeling in data. AK4 jet multiplicity distribution shows a mismatch between data and DY simulation. This difference is largest at higher jet multiplicities. Correction factors for DY simulation are obtained by using this distribution according to the description provided below. These correction factors are applied to the DY simulation as an event weight depending on the number of AK4 jets present in an event.

Correcting DY simulation using data

A mis-modeling in background simulation is observed in the distribution of AK4 jet multiplicity. The mis-modeling is prominent at higher jet multiplicity bins, as they are not modeled well by the MC generator. Fig. 5.10 shows the AK4 jet multiplicity distributions after pre-selection (first row) and in the control region CR4 (see Sec. 5.4.3) (second row) for $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels. It can be seen from this figure that the $\mathrm{T\overline{T}}$ signal tends to occupy the higher jet multiplicity bins. Since this analysis relies on the MC predictions, it is important to correct the DY simulation by using this distribution.

These correction factors are derived by taking the ratio between data ($\mathrm{t\bar{t}}$, $\mathrm{t\bar{t}Z}$ and other background yields are subtracted from data) and DY simulation using the AK4 jet multiplicity distribution. The top row of Fig. 5.11 shows the AK4 jet multiplicity distribution in CR2

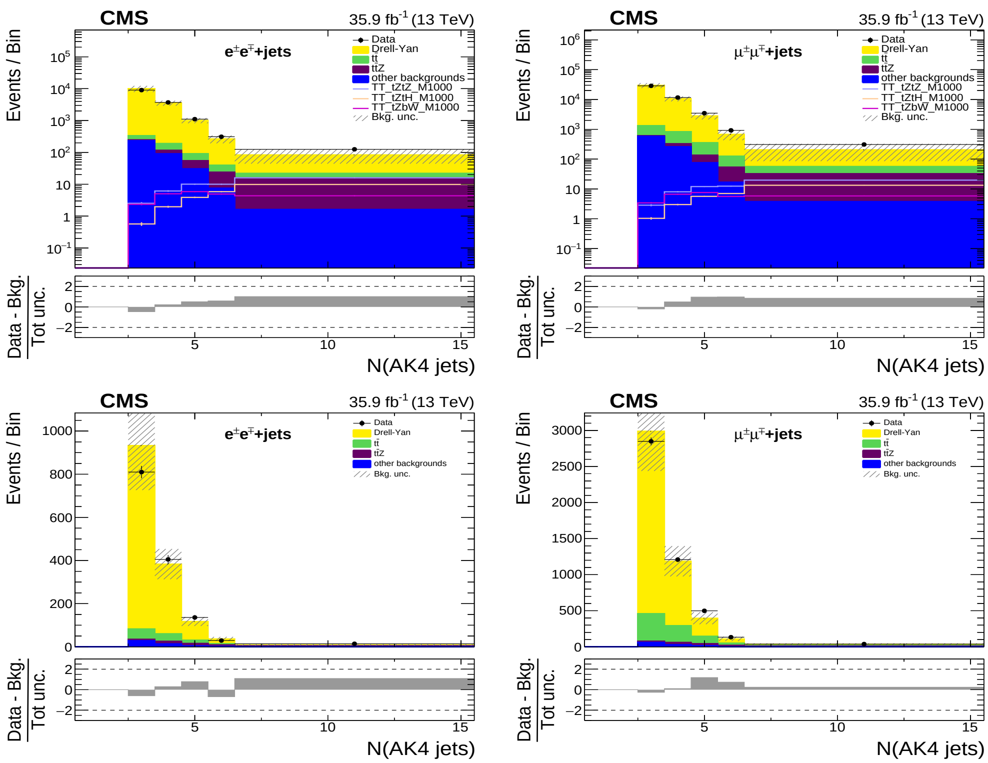


Figure 5.10: *The $N_{\text{AK4−jets}}$ spectrum after event pre-selection (first row) and in control region CR4 (second row) for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) channels. The hashed band shows the statistical and systematic uncertainties.*

and the bottom row shows the ratio between data ($t\bar{t}$, $t\bar{t}Z$ and other background yields are subtracted) and DY simulation for both $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) channels. Obtained scale factors per bin of jet multiplicity distribution are given in Table. 5.7 and are shown in the bottom row of Fig. 5.11. These scale factors are applied to the DY simulated events as a function of the number of AK4 jets reconstructed within an event.

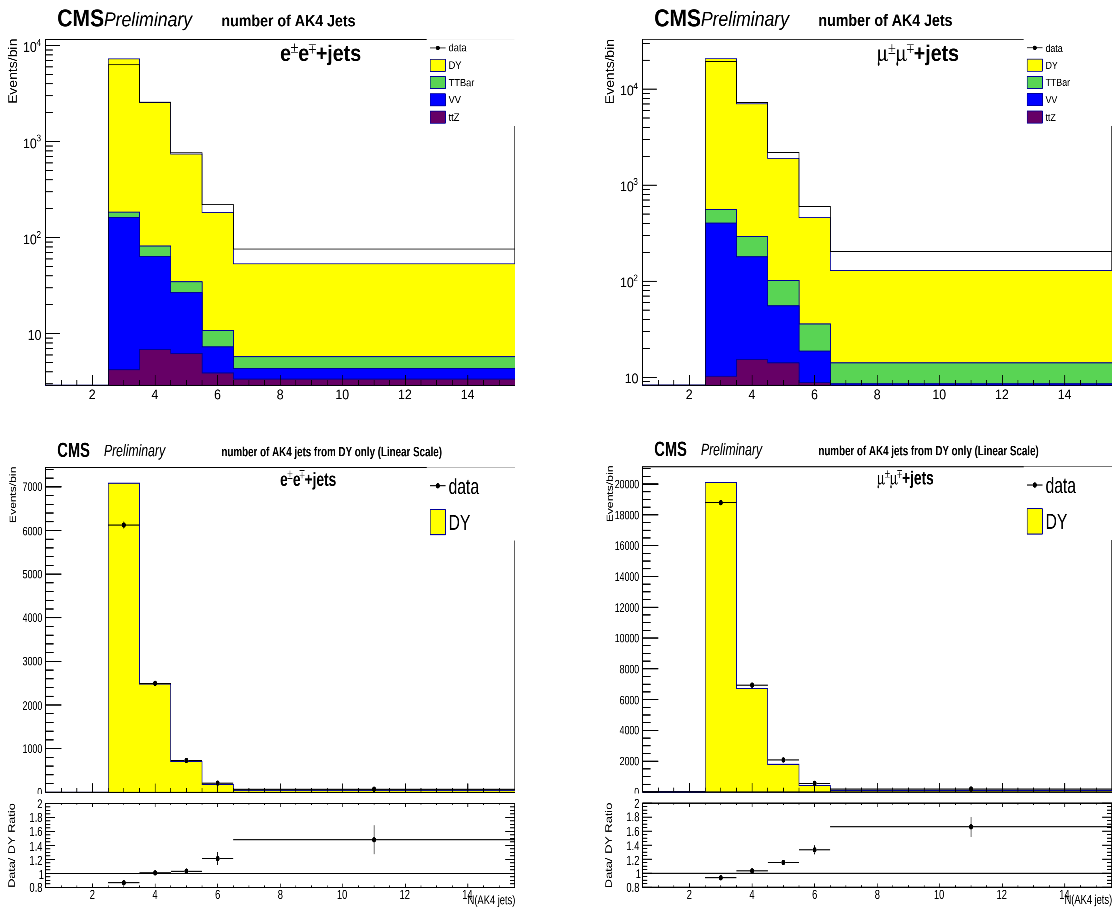


Figure 5.11: *The upper plots show the jet multiplicity distribution in CR2 and the lower plots show the ratio between background subtracted data and DY simulation for both* $Z \to e^+e^-$ *(left) and* $Z \to \mu^+\mu^-$ *(right) channels.*

The Fig. 5.12 and Fig. 5.13 show the effect of applying DY scale factors on the NAK4 and S_T distributions after event pre-selection and in control region CR4 for both $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels respectively. Left (right) plots of Fig. 5.13 show the kinematic distributions before (after) applying DY scale factors on DY simulation. It can be seen

Table 5.7: *Calculated DY correction factors for ee+jets and $\mu\mu$+jets channel.*

Number of AK4 Jets	$Z \to e^+e^-$channel	$Z \to \mu^+\mu^-$channel
$= 3$	0.86 ± 0.01	0.93 ± 0.01
$= 4$	1.01 ± 0.02	1.03 ± 0.01
$= 5$	1.03 ± 0.04	1.15 ± 0.03
$= 6$	1.21 ± 0.10	1.33 ± 0.06
≥ 7	1.47 ± 0.22	1.66 ± 0.14

from these distributions that applying the DY scale factors improves the overall data to MC modeling.

5.4.2 Control region CR3

Control region CR3 is defined, where events that are pre-selected should satisfy $N(\text{bjets}) = 0$ and $S_{\text{T}} \geq 1000$ GeV. This region is used to verify an agreement between data and background MC simulation and validate modeling of the S_{T} variable. Fig. 5.14 shows the S_{T} distributions for both $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels in this region.

SM backgrounds do not have any boosted bosons in this region. Therefore, this region is used to determine the mistag rate scale factors for jet substructure, W/Z-, H- and top-taggers and their associated uncertainties as mentioned in Sec. 4.2.7. The scale factors are applied to background MC simulation per tagged jet in an event.

5.4.3 Control region CR4

Control region CR4, is a low S_{T} control region, where pre-selected events are required to have $N(\text{bjets}) \geq 1$ and $S_{\text{T}} < 1000$ GeV. This region is used to validate the modeling between data and background simulation for all of the generic kinematic variables including S_{T}. Fig. 5.15 shows the distributions for dilepton invariant mass, p_{T} of the Z boson candidate and $\Delta R(l, l)$ in both $Z \to e^+e^-$and $Z \to \mu^+\mu^-$channels. The Fig. 5.16 shows the distributions for $\text{E}_{\text{T}}^{\text{miss}}$, AK4 jets multiplicity, H_{T} and S_{T} distributions in $Z \to e^+e^-$and $Z \to \mu^+\mu^-$channels. There is a good agreement seen in all of the distributions.

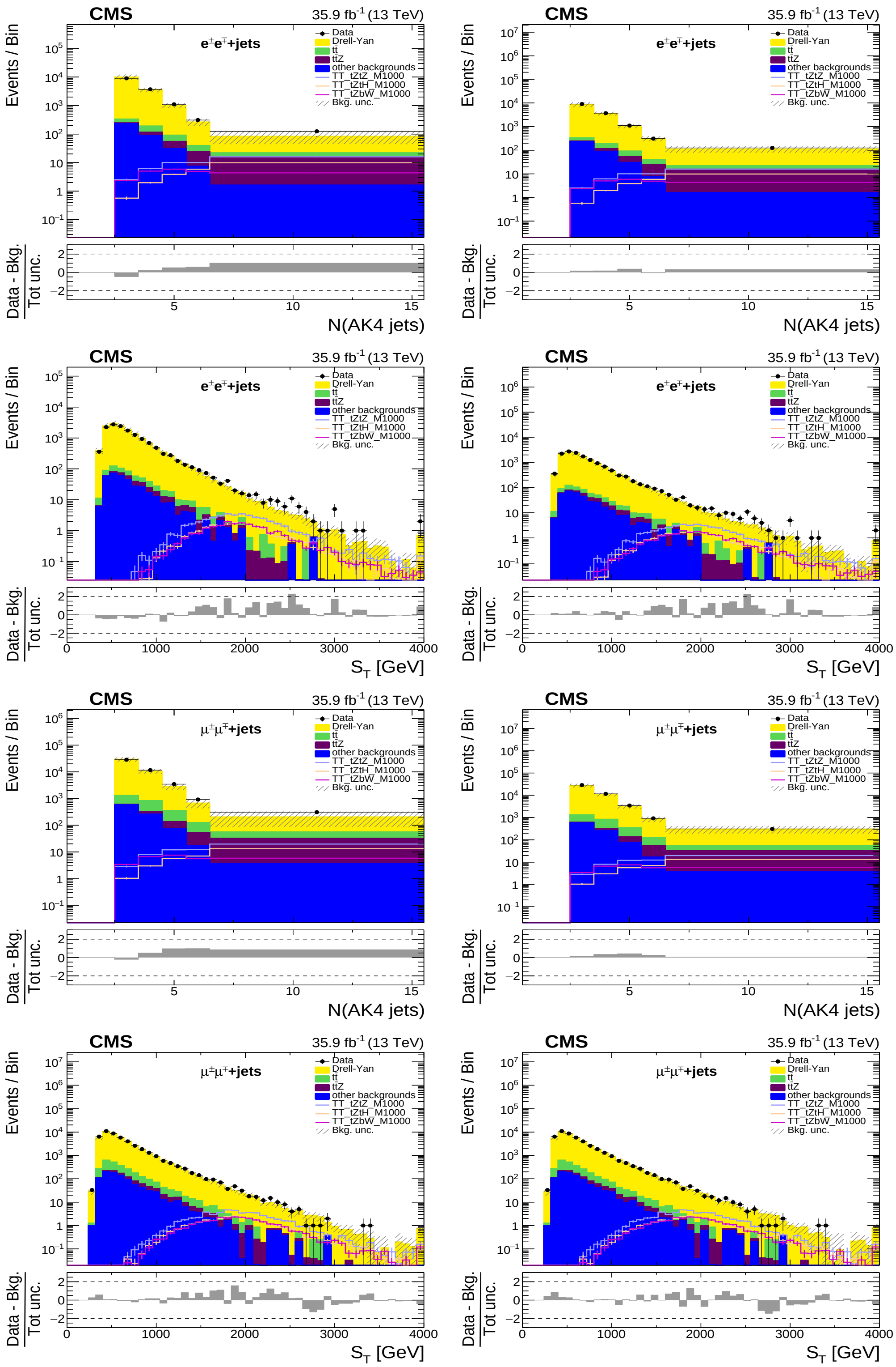


Figure 5.12: *The $N_{\text{AK4-jets}}$ and S_T spectrum for $Z \to e^+e^-$ (first and second rows) and $Z \to \mu^+\mu^-$ (third and fourth rows) channels in preselection region. The left (right) side plots are before (after) applying the DY SFs. The hashed band shows the statistical and systematic uncertainties.*

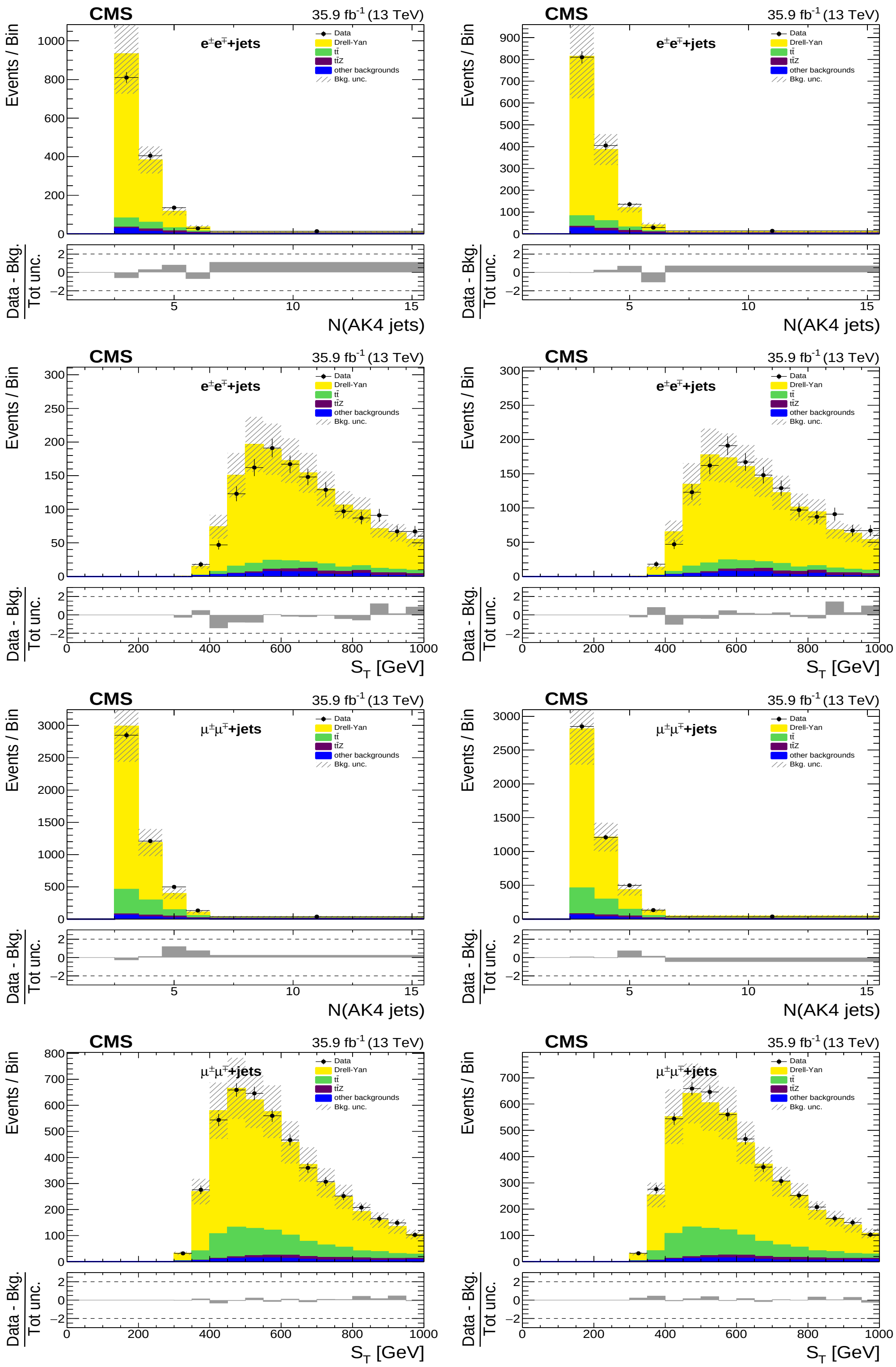


Figure 5.13: *The* $N_{\mathrm{AK4-jets}}$ *and* S_{T} *spectrums for* $Z \to e^{+}e^{-}$ *(first and second rows) and* $Z \to \mu^{+}\mu^{-}$ *(third and fourth rows) channel in CR4 control region. The left (right) side plots are before (after) applying the DY SFs. The hashed band shows the statistical and systematic uncertainties.*

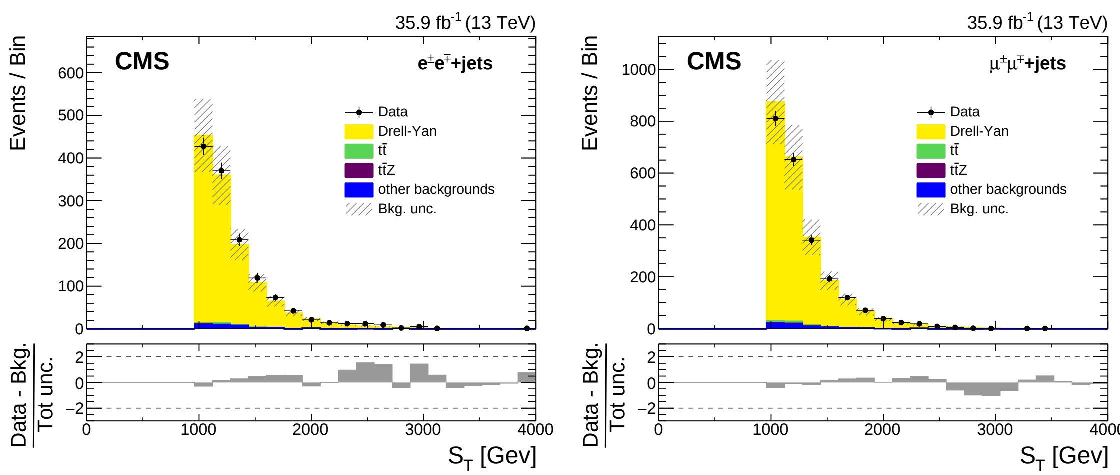


Figure 5.14: *S_{T} distributions in control region CR3 for $Z \rightarrow e^+e^-$(left) and $Z \rightarrow \mu^+\mu^-$(right) channels. The hashed band shows the statistical and systematic uncertainties.*

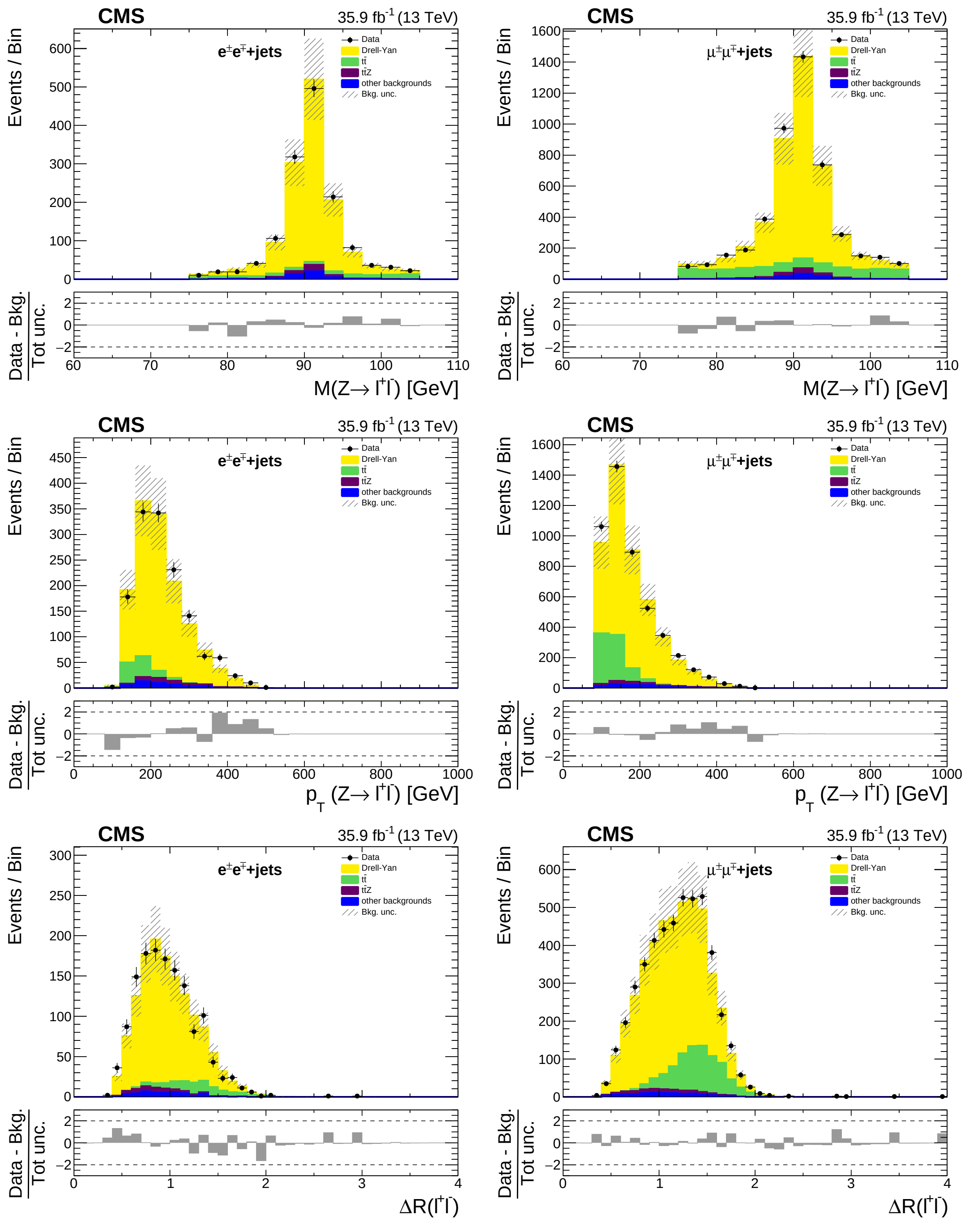


Figure 5.15: *Distributions for dilepton invariant mass, p_{T} of the Z boson candidate and $\Delta R(l,l)$ in $Z \to e^+e^-$(left) and $Z \to \mu^+\mu^-$(right) channels in control region CR4. The DY correction factors are applied. The hashed band shows the statistical and systematic uncertainties.*

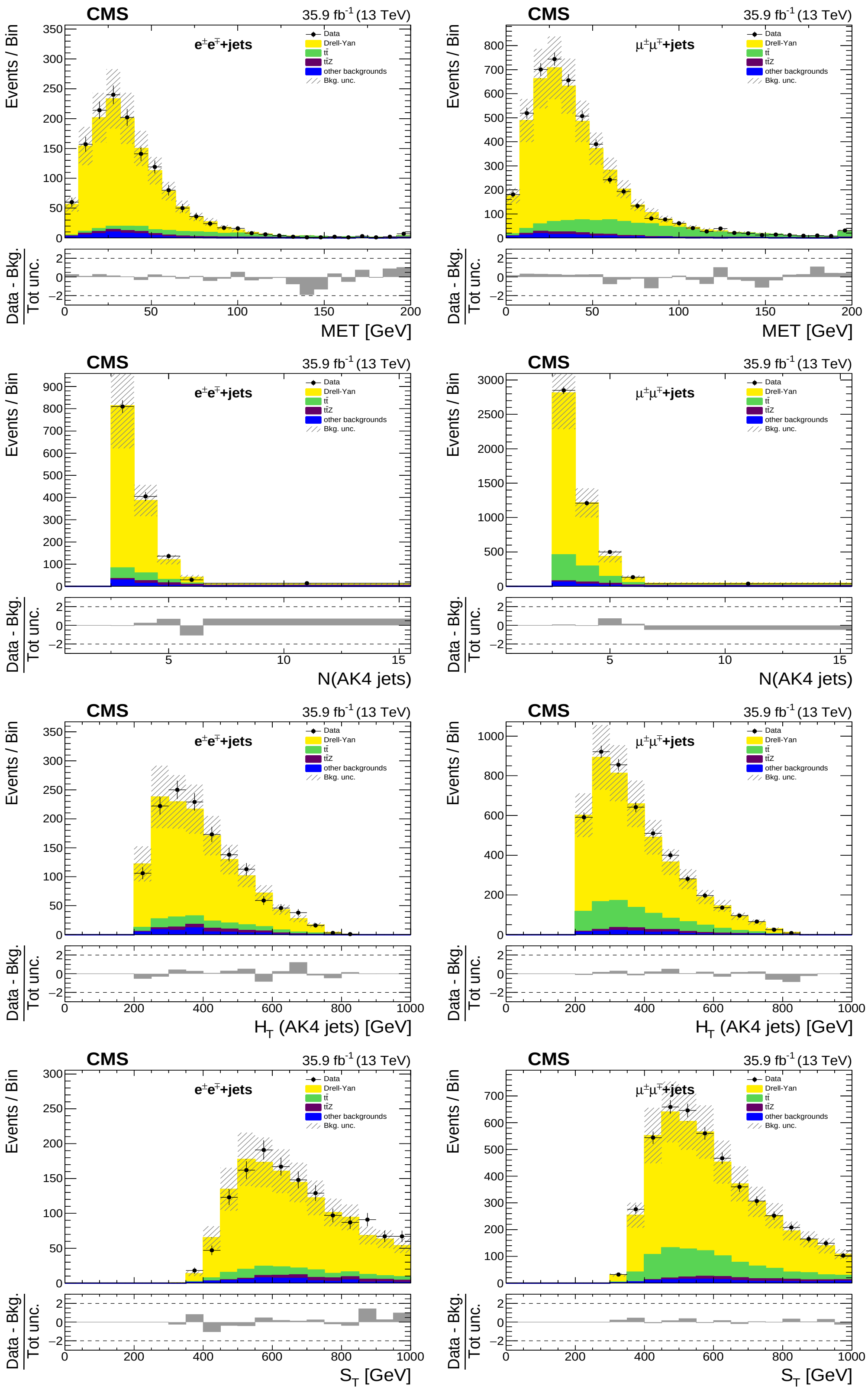


Figure 5.16: *Distributions for* E_T^{miss}, *AK4 jets multiplicity,* H_T *and* S_T *in* $Z \to e^+e^-$ *(left) and* $Z \to \mu^+\mu^-$ *(right) channels in control region CR4. The DY correction factors are applied. The hashed band shows the statistical and systematic uncertainties.*

Chapter 6

Search strategy and optimization

Fig. 6.1 and Table 6.1 present the S_{T} distributions and the expected number of events after full event selection for both $Z \rightarrow e^{+}e^{-}$ and $Z \rightarrow \mu^{+}\mu^{-}$ channels. The signal processes shown in Fig. 6.1 correspond to $\mathrm{T\bar{T}} \rightarrow$ tZtZ, $\mathrm{T\bar{T}} \rightarrow$ tZtH and $\mathrm{T\bar{T}} \rightarrow$ tZbW decay channels, where the samples are normalized to 1 pb cross section for better visibility. According to Fig. 6.1, it can be seen that the contribution from the background MC events decreases with increasing S_{T} and the $\mathrm{T\overline{T}}$ signal contribution peaks at higher S_{T} values as expected.

The selected events are further categorized according to the number of W/Z, H bosons and top and bottom quark candidates reconstructed within an event as described in Sec. 6.1. The purpose of this event categorization is to further purify the signal and reduce the SM background contributions.

6.1 $\mathrm{T\overline{T}}$ event reconstruction

According to the Feynman diagram shown in Fig. 5.1, the final states of $\mathrm{T\overline{T}}$ decay consist of many hadronically decaying objects such as W/Z and H bosons, top and bottom quarks. Identifying these hadronically decaying candidates after full event selection is important to optimize the $\mathrm{T\overline{T}}$ search. Signal events are expected have a larger number of such reconstructed candidates. First, all the hadronically decaying W/Z and H bosons, top and

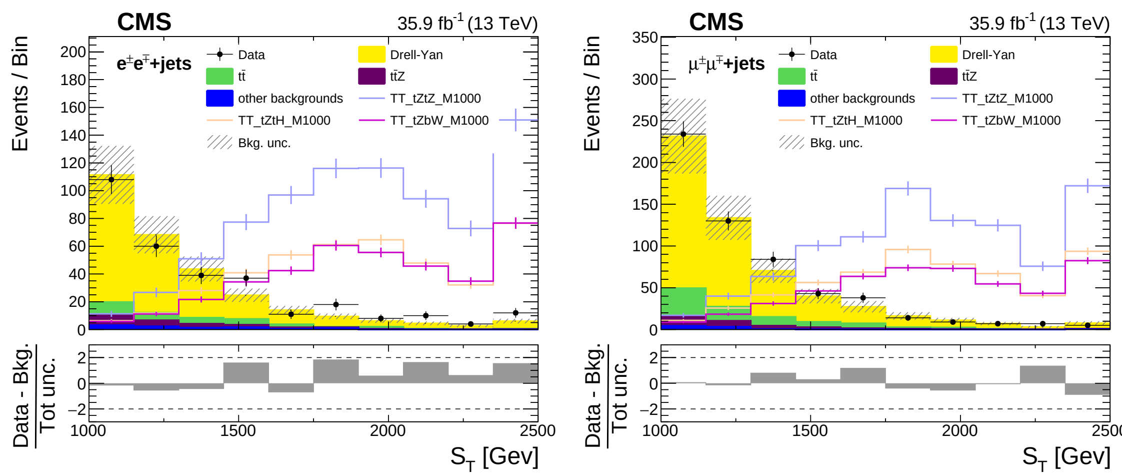


Figure 6.1: *S_{T} distributions in $Z \to e^{+}e^{-}$ (left) and $Z \to \mu^{+}\mu^{-}$ (right) channels after full event selection. The signal MC are normalized using a cross section of 1 pb for better visibility. The DY correction factors are applied. The hashed band shows the statistical and systematic uncertainties.*

Table 6.1: *Expected number of events after the full event selection. The signal MC yields are given for the three decay modes,* $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$, $\mathrm{T\bar{T}} \to \mathrm{tZbW}$ *and* $\mathrm{T\bar{T}} \to \mathrm{tZtH}$ *for T quark mass of 1000* GeV. *The background uncertainties include both statistical and systematic contributions.*

Sample	$Z \to e^{+}e^{-}$channel	$Z \to \mu^{+}\mu^{-}$channel
Data	308	571
DY+jets	233.4 ± 24.8	432.3 ± 48.4
$\mathrm{t\bar{t}}$	28.6 ± 5.6	77.7 ± 16.5
$\mathrm{t\bar{t}Z}$	18.6 ± 1.5	29.0 ± 2.4
other backgrounds	10.2 ± 2.5	9.3 ± 2.2
Total Bkg.	290.9 ± 27.2	548.3 ± 55.4
$\mathrm{T\bar{T}} \to \mathrm{tZtZ}$	35.8 ± 0.8	44.2 ± 0.9
$\mathrm{T\bar{T}} \to \mathrm{tZbW}$	17.1 ± 0.4	21.8 ± 0.5
$\mathrm{T\bar{T}} \to \mathrm{tZtH}$	18.6 ± 0.4	25.5 ± 0.5

bottom quark candidates are identified within an event. This is done as follows.

The hadronically decaying objects can be identified in three different ways defined as *resolved*, *partially merged* or *fully merged* candidates. Resolved candidates are identified, where the W/Z and H bosons, and top quarks originated from the T quarks are not boosted. Therefore, any subsequent decays from these objects can be reconstructed as separate AK4 jets. Partially or fully merged candidates are identified, where few or all of the W/Z and H bosons, and top quarks originated from the T quarks are boosted. In such cases, these objects can be reconstructed as AK8 jets and jet substructure techniques mentioned in Sec. 4.2.6.7 can be employed to identify them. The resolved and merged (partially or fully merged) candidates correspond to different p_{T} regimes of the W/Z and H bosons, and top quarks.

Since the reconstructed AK8 jets have a cone size parameter (R=0.8) greater than the AK4 jets (R=0.4), there can be instances, where AK4 jets overlap with AK8 jets. Such overlaps between AK4 and AK8 jets are identified if $\Delta\mathrm{R}(\mathrm{AK4\ jet, AK8\ jet}) < 0.8$ and corresponding AK4 jet is removed from the AK4 jet collection to avoid double counting.

The hadronically decaying candidates can be identified as:

- **Top quark candidates** - A top quark decays into a W boson and a b quark. As the top quarks in this analysis are originated from heavy T quarks, they can be identified as either resolved, partially merged or fully merged top quarks as described below.

  - **Resolved top quark**- Resolved top quark candidates are identified when a top quark decays via, $\mathrm{t} \to \mathrm{Wb} \to \mathrm{q'\bar{q}b}$ as shown in the left diagram of Fig 6.4, and all three quarks $\mathrm{q'}, \bar{\mathrm{q}}$ and b are reconstructed as three separate AK4 jets with invariant mass of the three jets consistent with the top quark mass.

    The selections to define the resolved top quark candidates are derived using the generator level information from the $\mathrm{T\overline{T}}$ signal MC samples. Two AK4 jets matched to two quarks that are originated from a hadronically decaying W boson ($W \to q'\bar{q}$) are combined with an AK4 jet that is matched to a b quark from the same top quark decay, are used to identify resolved top quark candidates. The matching is done by applying $\Delta\mathrm{R}(\mathrm{generator\ level\ particle,\ reco\ AK4\ jet}) < 0.3$

requirement. The invariant mass and the vector sum p_{T} distributions of the three selected AK4 jets are shown in Fig. 6.2. The signal processes shown here correspond to three different T quark masses, $\mathrm{m_T} = 800,\ 1200$ and 1600 GeV in the $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtZ}$ decay channel.

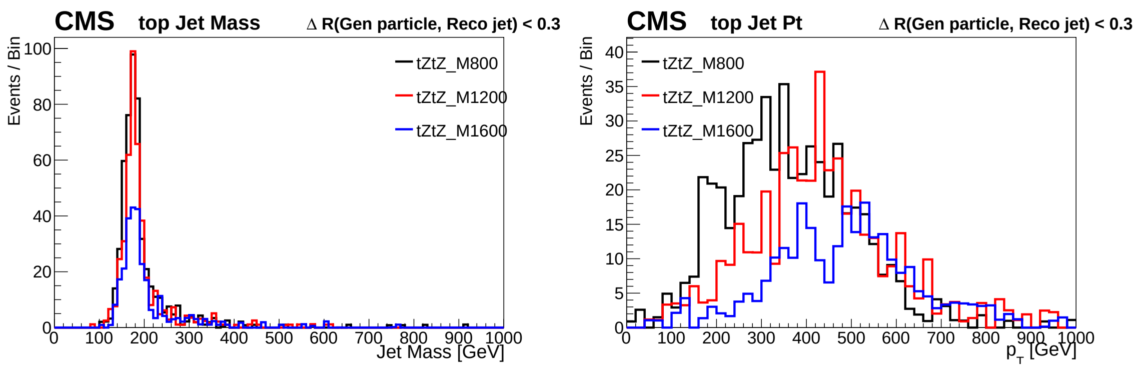


Figure 6.2: *The mass and* p_{T} *distributions of the three reconstructed AK4 jets corresponding to decay products of a top quark. The signal processes shown here correspond to three different T quark masses,* $\mathrm{m_T} = 800, 1200$ *and* 1600 GeV *in the* $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtZ}$ *decay channel.*

As can be seen in Fig. 6.2, the invariant mass of the three jet system peaks around the top quark mass. Hence, the invariant mass of the three jet system, $120 < \mathrm{M}(j_1, j_2, j_3) < 240$ GeV consistent with the top quark mass and vector sum p_{T} of the three jet system, $\sum p_{\mathrm{T}}(j_1, j_2, j_3) > 100$ GeV are used to identify *resolved top quark candidates*. More than 98% of the signal is preserved by using these selections.

– **Partially merged top quark** - Partially merged top quark candidates can be identified when a top quark decays to a W boson and a b quark via $\mathrm{t} \rightarrow \mathrm{Wb}$, where W boson is boosted as shown in the middle diagram of Fig 6.4 . In this case, W boson is reconstructed as a W tagged AK8 jet as mentioned in Sec. 4.2.6.12.

The selections on partially merged top quark candidates are derived by using the generator level information from the $\mathrm{T\overline{T}}$ signal MC samples. W tagged AK8 jet matched to a hadronically decaying W boson is combined with an AK4 jet that is matched to a b quark, are used to identify partially merged top quark candidates. The matching is done by applying $\Delta\mathrm{R}(\text{generator level b quark},\ \text{reco AK4 jet}) <$

0.3 and ΔR(generator level W boson, reco AK8 jet) < 0.8 requirements. The invariant mass and the vector sum p_{T} distributions of the selected AK4 jet and W tagged AK8 jet are shown in Fig. 6.3. The signal processes shown here correspond to three different T quark masses, $\mathrm{m_T} = 800,\ 1200$ and 1600 GeV in the $\mathrm{T\bar{T}} \rightarrow$ tZtZ decay channel.

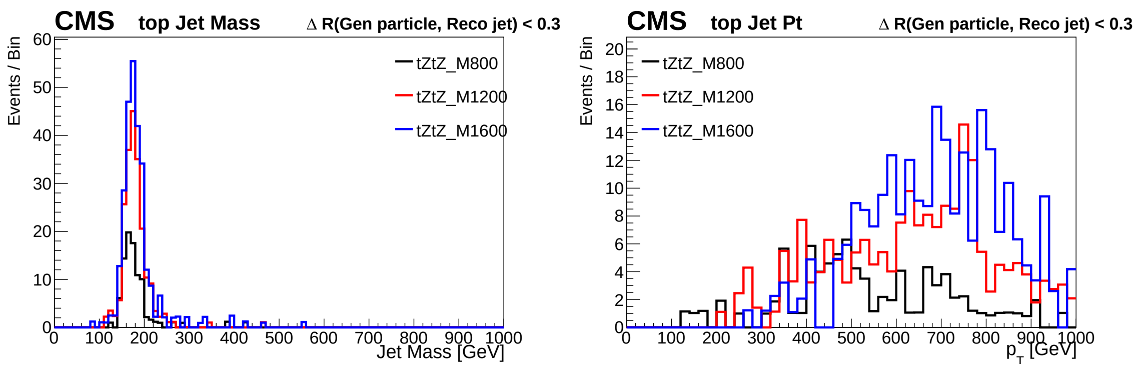


Figure 6.3: *The mass and p_{T} distributions of the reconstructed AK4 and W tagged AK8 jets corresponding to decay products of a top quark. The signal processes shown here correspond to three different T quark masses,* $\mathrm{m_T} = 800, 1200$ *and* 1600 GeV *in the* $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtZ}$ *decay channel.*

According to Fig. 6.3, the invariant mass of the two jet system W tagged AK8 jet and b jet, peaks around the top quark mass. Hence, the invariant mass of the two jet system, $120 < \mathrm{M}(j_1, j_2) < 240$ GeV consistent with the top quark mass and vector sum p_{T} of the two jet system, $\sum p_{\mathrm{T}}(j_1, j_2) > 150$ GeV are used to identify *partially merged top quark candidates.* More than 98% of the signal is preserved by using these selections.

- **Fully merged top quark** - Fully merged top quark candidates can be identified when a top quark is boosted as shown in right side digram of Fig. 6.4. Decay products from the top quark are collinear and hence reconstructed using the boosted top quark tagger mentioned in Sec. 4.2.6.14, using the softdrop algorithm. According to the requirements of this tagger, a boosted top quark is identified if the reconstructed AK8 jet has only three subjets with a softdrop jet mass ($\mathrm{M_{softdrop}}$) that satisfy $105 < \mathrm{M}_{softdrop} < 220$ GeV. In addition, $\tau_3/\tau_2 < 0.81$ and

$p_{\mathrm{T}} > 400$ GeV are required.

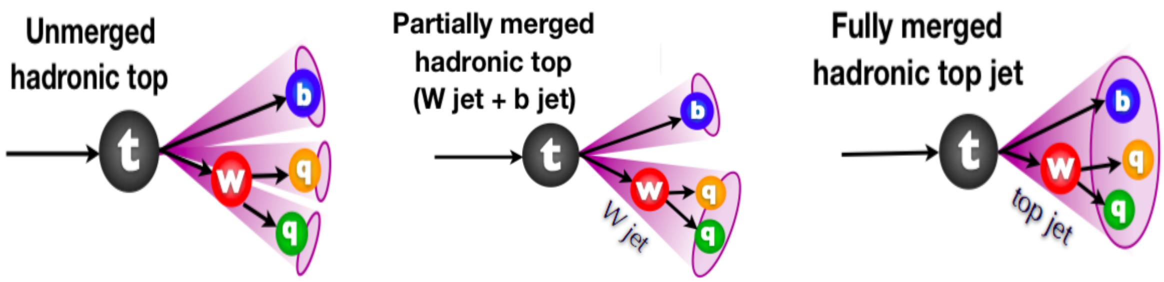


Figure 6.4: *The resolved(left), partially merged (middle) and fully merged (right) top quark jets*[121]

- **W/Z (V) boson candidates**
  - **Resolved V boson** - Resolved V bosons originated from a T quark can decay into two quarks via, $\mathrm{W} \rightarrow \mathrm{q}'\bar{\mathrm{q}}$ or $\mathrm{Z} \rightarrow \mathrm{q}\bar{\mathrm{q}}$. The two quarks are reconstructed as two separate AK4 jets.

    The selections on resolved V boson candidates are derived using the generator level information from the $\mathrm{T}\overline{\mathrm{T}}$ signal MC samples. Two AK4 jets matched to two quarks originated from a hadronically decaying V boson ($V \rightarrow q\bar{q}$) are used to identify resolved V boson candidates. The matching is done by applying $\Delta\mathrm{R}(\text{generator level particle, reco AK4 jet}) < 0.3$ requirement. The invariant mass and the vector sum p_{T} distributions of the two selected AK4 jets are shown in Fig. 6.5. The signal processes shown here correspond to three different T quark masses, $\mathrm{m_T} = 800,\ 1200$ and 1600 GeV in the $\mathrm{T}\bar{\mathrm{T}} \rightarrow \mathrm{tZtZ}$ decay channel.

    According to Fig. 6.5, the invariant mass of the two jet system peaks around the V boson mass. Hence, the invariant mass of the two jet system, $70 < \mathrm{M}(j_1, j_2) < 120$ GeV consistent with the V boson mass and vector sum p_{T} of the two jet system, $\sum p_{\mathrm{T}}(j_1, j_2) > 100$ GeV are used to identify *resolved V boson candidates*. More than 95% of the signal is preserved by using these selections.
  - **Fully merged V boson** - Fully merged V bosons can be identified when W or Z boson is boosted. Decay products are collinear and hence reconstructed using

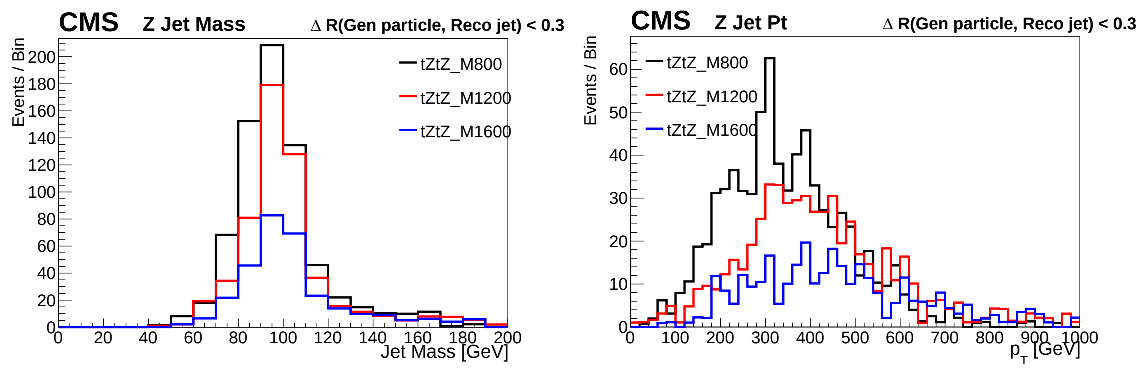


Figure 6.5: *The mass and vector sum p_{T} distributions of the two reconstructed AK4 jets that are originated from a V boson. The signal processes shown here correspond to three different T quark masses,* $\mathrm{m_T} = 800, 1200$ *and* 1600 GeV *in the* $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtZ}$ *decay.*

the boosted W/Z boson tagger mentioned in Sec. 4.2.6.12, using the pruning algorithm. According to the requirements of this tagger, a boosted V boson is identified if the reconstructed AK8 jet has only two subjets with a pruned jet mass ($\mathrm{M_{pruned}}$) that satisfy $65 < \mathrm{M}_{pruned} < 105$ GeV. No selection is applied on τ_2/τ_1 variable (i.e $\tau_2/\tau_1 < 1.0$) and $p_{\mathrm{T}} > 200$ GeV is requried.

- **H boson candidates**

  - **Resolved H boson** - Resolved H bosons originated from a T quark can decay into two quarks via, $\mathrm{H} \rightarrow \mathrm{q\bar{q}}$. The two quarks are reconstructed as two separate AK4 jets where at least one of them should be b tagged as mentioned in Sec. 4.2.6.6.

    The identification criteria for resolved H boson candidates are derived using the generator level information from the $\mathrm{T\overline{T}}$ signal MC samples. Two AK4 jets are matched to two quarks originated from a hadronically decaying H boson ($H \rightarrow q\bar{q}$), where at least one of them is a b jet. The matching is done by applying ΔR(generator level particle, reco AK4 jet) < 0.3 requirement. The invariant mass and the vector sum p_{T} distributions of the two selected AK4 jets are shown in Fig. 6.6. The signal processes shown here correspond to three different T quark masses, $\mathrm{m_T} = 800,\ 1200$ and 1600 GeV in the $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtH}$ decay channel.

    According to Fig. 6.6, the invariant mass of the two jet system peaks around the

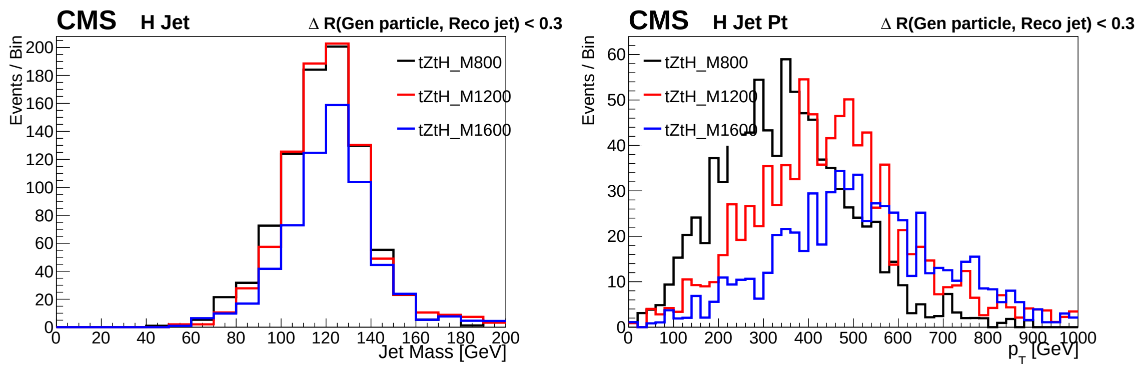


Figure 6.6: *The mass and vector sum* p_{T} *distributions of the two reconstructed AK4 jets originated from a H boson where at least one of them is a b jet. The signal processes shown here correspond to three different T quark masses,* $\mathrm{m_T} = 800, 1200$ *and* 1600 GeV *in the* $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtH}$ *decay channel.*

H boson mass. Hence, the invariant mass of the two jet system, $80 < \mathrm{M}(j_1, j_2) < 160$ GeV consistent with the H boson mass and vector sum p_{T} of the two jet system, $\sum p_{\mathrm{T}}(j_1, j_2) > 100$ GeV are used to identify *resolved H boson candidates*. More than 95% of the signal is preserved by using these selections.

– **Fully merged H boson** - Fully merged H bosons can be identified when H boson is boosted. Decay products originated from the H bosons have a fair amount of Lorentz boost and are collinear with each other. Hence, they cannot be reconstructed as separate AK4 jets. The boosted H boson tagger mentioned in Sec. 4.2.6.13 is used in such cases with the pruning algorithm. According to the requirements of this tagger, a boosted H boson is identified if the reconstructed AK8 jet has only two subjets with a pruned jet mass ($\mathrm{M_{pruned}}$) that satisfy $105 < \mathrm{M}_{pruned} < 135$ GeV. No selection is applied on τ_2/τ_1 variable (i.e $\tau_2/\tau_1 < 1.0$) and $p_{\mathrm{T}} > 300$ GeV is required.

These selections are summarized in Table. 6.2. The resolved, partially merged and fully merged objects within an event are identified according to these selections. These reconstructed objects are used to optimize the signal search. As mentioned earlier, $\mathrm{T\overline{T}}$ signal events should have a higher number of such reconstructed objects than the SM background events. The multiplicity of these objects can be used to reduce SM background contribution. The selected events after full event selection are categorized according to number of identified candidates and number of b jets, present within an event.

Table 6.2: *Selection summary for hadronically decaying top, V and H candidates*

Category	**top**	**V**	**H**
Resolved	$\sum p_\mathrm{T}(j_1, j_2, j_3) > 100$ GeV $120 < \mathrm{M}(j_1, j_2, j_3) < 240$ GeV	$\sum p_\mathrm{T}(j_1, j_2) > 100$ GeV $70 < \mathrm{M}(j_1, j_2) < 120$ GeV	$\sum p_\mathrm{T}(j_1, j_2) > 100$ GeV $80 < \mathrm{M}(j_1, j_2) < 160$ GeV n(b jets) ≥ 1
Partially merged	V tagged AK8 jet $65 < \mathrm{M}_{pruned} < 105$ GeV $p_\mathrm{T} > 200$ GeV $\tau_2/\tau_1 < 1.0$		
	$\sum p_\mathrm{T}(j_1, j_2) > 150$ GeV $120 < \mathrm{M}(j_1, j_2) < 240$ GeV		
Fully merged	top tagged AK8 jet $105 < \mathrm{M}_{softdrop} < 220$ GeV $p_\mathrm{T} > 400$ GeV $\tau_3/\tau_2 < 0.81$	V tagged AK8 jet $65 < \mathrm{M}_{pruned} < 105$ GeV $p_\mathrm{T} > 200$ GeV $\tau_2/\tau_1 < 1.0$	H tagged AK8 jet $105 < \mathrm{M}_{pruned} < 135$ GeV $p_\mathrm{T} > 300$ GeV $\tau_2/\tau_1 < 1.0$

The reconstructed AK4 and AK8 jets within an event are used to identify the resolved, partially merged and fully merged candidates for hadronically decaying V and H bosons, and top quarks according to the selection criteria mentioned in Table 6.2. The total number of hadronically decaying objects (N_X) for each type (V, H or top) is then obtained by summing the number of resolved ($N_{Resolved}$), partially merged ($N_{Partially\ merged}$) and fully merged ($N_{Fully\ merged}$) candidates identified for each object per event as given by Eq. 6.1.

$$N_X = N_{Resolved(X)} + N_{Partially\ merged(X)} + N_{Fully\ merged(X)} \tag{6.1}$$

Where X = top/ V/ H. For example, the total number of hadronically decaying top quarks (N_t) in an event is obtained by summing the number of resolved, partially merged and fully merged top candidates identified using the AK4 and AK8 jets according to selection criteria mentioned in Table 6.2. The total numbers of hadronically decaying V bosons (N_V) and H bosons (N_H) are counted in the same way. Then, events can be categorized according the number of such identified objects N_{t}, N_{V}, N_{H} and number of b jets (N_{b}). Sixteen event categories are constructed according to $N_{\mathrm{t}} = 0$ or ≥ 1, $N_{\mathrm{V}} = 0$ or ≥ 1, $N_{\mathrm{H}} = 0$ or ≥ 1 and $N_{\mathrm{b}} = 1$ or ≥ 2 as shown in Table 6.3.

Table 6.3: *Description of sixteen categories depending on* N_{t}, N_{V}, N_{H} *and* N_{b}.

	Category	N_{t}	N_{V}	N_{H}	N_{b}
1	t1Z1H1b1	≥ 1	≥ 1	≥ 1	=1
2	t1Z1H1b2	≥ 1	≥ 1	≥ 1	≥ 2
3	t1Z1H0b1	≥ 1	≥ 1	=0	=1
4	t1Z1H1b2	≥ 1	≥ 1	=0	≥ 2
5	t0Z1H1b1	=0	≥ 1	≥ 1	=1
6	t0Z1H1b2	=0	≥ 1	≥ 1	≥ 2
7	t0Z1H0b1	=0	≥ 1	=0	=1
8	t0Z1H1b2	=0	≥ 1	=0	≥ 2
9	t1Z0H1b1	≥ 1	=0	≥ 1	=1
10	t1Z0H1b2	≥ 1	=0	≥ 1	≥ 2
11	t1Z0H0b1	≥ 1	=0	=0	=1
12	t1Z0H1b2	≥ 1	=0	=0	≥ 2
13	t0Z0H1b1	=0	=0	≥ 1	=1
14	t0Z0H1b2	=0	=0	≥ 1	≥ 2
15	t0Z0H0b1	=0	=0	=0	=1
16	t0Z0H1b2	=0	=0	=0	≥ 2

The counting of t, V, H and b candidates is done independently. All possible AK4 jet permutations are considered, which satisfy the resolved top, H and V candidate requirements mentioned in Table 6.2. Therefore, the same AK4 jet can be used by more than one resolved tagger. These possible overlaps between different types of candidates within an event are allowed to simplify the analysis.

According to strategy of the analysis, if a candidate is identified and counted as both a top candidate and a H candidate, the event is counted in "t1H1" category only. Despite the overlaps between candidates, each event falls into only one of the sixteen categories mentioned in Table 6.3, depending on the number of candidates identified. So the sixteen event categories are mutually exclusive.

The small overlapping mass window of top and H taggers mentioned in Sections 4.2.6.14 and 4.2.6.13 allows the same AK8 jet to be tagged as both a top quark and a H boson candidate. In such cases the MC truth information is used to identify true top and Higgs candidates originated from the hard scattering process as mentioned in Sec. 4.2.7. The appropriate efficiency or mistag scale factors are applied as an event weight based on these information.

Fig. 6.7 and Table 6.4 show the distribution of data, background and signal events among the sixteen categories mentioned in Table 6.3, in $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels in signal region. The signal event distributions among sixteen categories are shown separately in Fig. 6.8 for $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$, $\mathrm{T\bar{T}} \to \mathrm{tZtH}$ and $\mathrm{T\bar{T}} \to \mathrm{tZbW}$ decay channels, for the benchmark T quark mass of 1000 GeV.

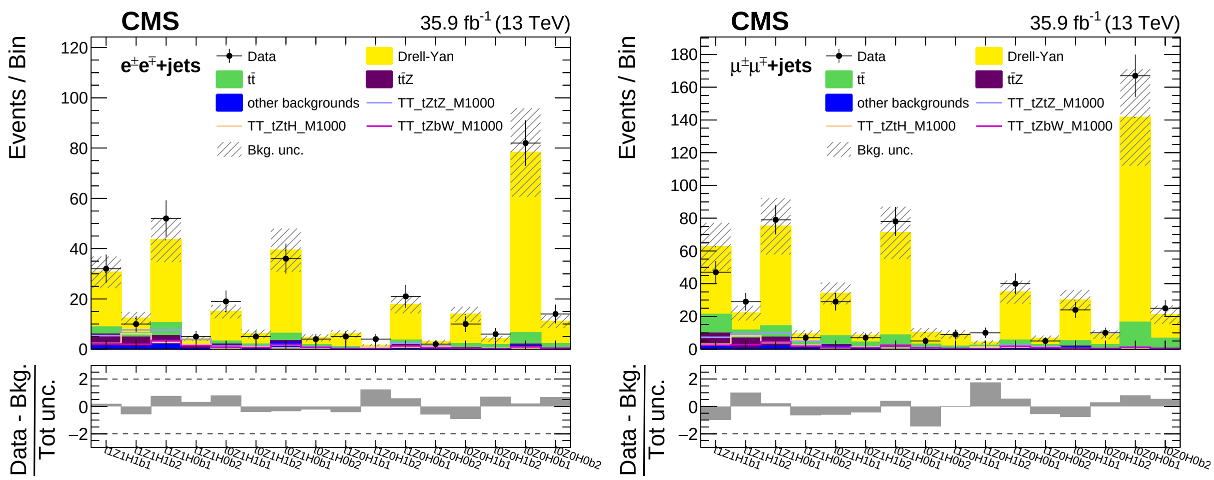


Figure 6.7: *Event counts in all the 16 mutually exclusive event categories based on the number of t, V, H and b candidates identified in* $Z \to e^+e^-$ *(left) and the* $Z \to \mu^+\mu^-$ *(right) channels in signal region. The hashed band shows the statistical and systematic uncertainties. The DY correction factors are applied.*

According to Fig. 6.8, most of the signal events mainly fall into the categories t1Z1H1b1, t1Z1H1b2, t1Z1H0b1 and t1Z1H0b2. This is expected for the $\mathrm{T\overline{T}}$ signal, as these are the categories that have many boosted top quarks, H or V bosons within an event. It can be seen from Fig. 6.7, that most of the background events fall into the t0Z0H0b1 category. This is because, most of the background MC events in the signal region already have ≥ 1 b jet (according to final event selection), but are not expected to have boosted top quarks, H or V bosons.

In order to use these categories for the $\mathrm{T\overline{T}}$ signal optimization, they are validated for

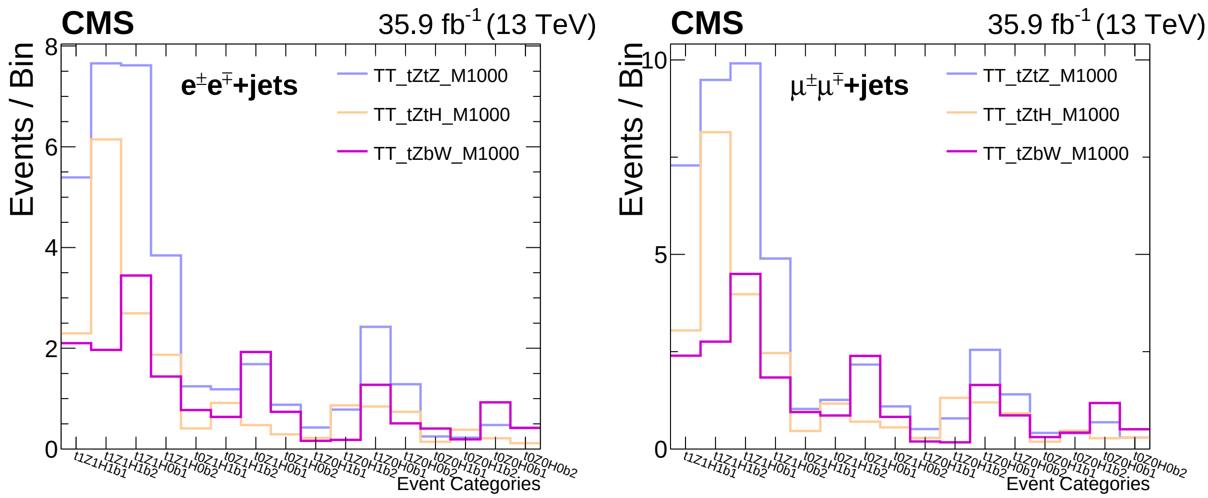


Figure 6.8: *Event counts in all the 16 mutually exclusive event categories based on the number of t, V, H and b candidates identified in* $Z \to e^+e^-$ *(left) and the* $Z \to \mu^+\mu^-$ *(right) channels in signal region for* $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$, $\mathrm{T\bar{T}} \to \mathrm{tZtH}$ *and* $\mathrm{T\bar{T}} \to \mathrm{tZbW}$ *decay channels.*

the agreement between data and background simulation in control region CR4 (mentioned in Sec. 5.4.3) as shown in Fig. 6.9. All sixteen categories have a good agreement between data and background simulation in control region CR4 for both $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels, and hence further optimization is possible.

6.2 $\mathrm{T\overline{T}}$ signal optimization

The search for the new physics signal is performed using the Bayesian likelihood technique, which is described in Sec. 8.1.2. The fit to the observed S_{T} distribution is performed using the combination of background and signal S_{T} distributions. This is done by using the THETA framework[146] and computing 95% confidence level limits[1].

Not all sixteen categories described earlier are sensitive to the $\mathrm{T\overline{T}}$ signal. Only some of these categories are selected based on their significance. Significance is defined as the ratio between number of signal events and $\sqrt{\text{number of background events}}$ ($S/\sqrt{B}$) present in each category. The event categories with the highest significance are the most sensitive

[1]The statistical concepts and a brief introduction to THETA framework is given in Sec. 8.1

Table 6.4: *Expected number of events in the 16 mutually exclusive event categories based on N_{t}, N_{V}, N_{H} and N_{b} in $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels. The "t" , "Z" , "H" and "b" in each category name denotes "top quark", "V boson", "H boson" and "b jet" respectively. The upper half of the table represents the $Z \to e^+e^-$ channel and lower half represents the $\mu\mu$+jets channel.*

Electron channel: ee+jets

	Category	Data	DY	$t\bar{t}$	$t\bar{t}Z$	other backgrounds	T → tZ, M800 GeV			T → tZ, M1000 GeV			T → tZ, M1200 GeV		
							T → tZ	T → tH	T → bW	T → tZ	T → tH	T → bW	T → tZ	T → tH	T → bW
1	t1Z1H1b1	32.0	21.9 ± 1.3	2.8 ±0.7	4.8 ± 0.1	1.1 ±0.6	27.9 ±1.6	11.6 ±0.7	8.4 ±0.6	5.4 ± 0.3	2.3 ±0.1	2.1 ±0.1	1.4 ± 0.1	0.6 ±0.0	0.5 ± 0.0
2	t1Z1H1b2	10.0	5.9 ± 0.7	1.7 ±0.6	3.8 ± 0.1	0.9 ±0.6	31.1 ±1.7	24.6 ±1.0	8.3 ±0.6	7.7 ± 0.4	6.1 ±0.2	2.0 ±0.1	1.8 ± 0.1	1.5 ±0.1	0.5 ± 0.0
3	t1Z1H0b1	52.0	33.1 ± 1.6	5.1 ±1.0	3.4 ± 0.1	1.9 ±0.8	27.1 ±1.5	10.0 ±0.7	13.0 ±0.8	7.6 ± 0.4	2.7 ±0.2	3.4 ±0.2	2.2 ± 0.1	0.8 ±0.0	0.9 ± 0.0
4	t1Z1H0b2	5.0	3.2 ± 0.5	0.0 ±0.0	1.0 ± 0.1	0.1 ±0.0	12.5 ±1.1	6.5 ±0.5	5.7 ±0.5	3.8 ± 0.3	1.9 ±0.1	1.4 ±0.1	1.5 ± 0.1	0.5 ±0.0	0.4 ± 0.0
5	t0Z1H1b1	20.0	11.9 ± 0.9	1.2 ±0.5	0.6 ± 0.0	1.3 ±0.7	4.3 ±0.6	1.4 ±0.3	3.0 ±0.4	1.2 ± 0.2	0.4 ±0.1	0.8 ±0.1	0.2 ± 0.0	0.1 ±0.0	0.2 ± 0.0
6	t0Z1H1b2	5.0	4.1 ± 0.5	1.4 ±0.5	0.5 ± 0.0	0.0 ±0.0	3.8 ±0.6	2.9 ±0.4	2.9 ±0.4	1.2 ± 0.2	0.9 ±0.1	0.6 ±0.1	0.3 ± 0.0	0.2 ±0.0	0.2 ± 0.0
7	t0Z1H0b1	36.0	33.2 ± 1.5	2.8 ±0.7	0.7 ± 0.0	2.7 ±1.2	4.7 ±0.6	2.4 ±0.3	6.3 ±0.5	1.7 ± 0.2	0.5 ±0.1	1.9 ±0.1	0.6 ± 0.1	0.1 ±0.0	0.6 ± 0.0
8	t0Z1H0b2	4.0	2.8 ± 0.5	1.4 ±0.5	0.2 ± 0.0	0.1 ±0.1	2.2 ±0.4	1.3 ±0.2	2.7 ±0.3	0.9 ± 0.1	0.3 ±0.1	0.7 ±0.1	0.3 ± 0.0	0.1 ±0.0	0.2 ± 0.0
9	t1Z0H1b1	5.0	5.2 ± 0.6	0.4 ±0.3	0.4 ± 0.0	0.0 ±0.0	2.1 ±0.4	0.8 ±0.2	0.4 ±0.1	0.4 ± 0.1	0.2 ±0.0	0.2 ±0.0	0.1 ± 0.0	0.1 ±0.0	0.0 ± 0.0
10	t1Z0H1b2	4.0	0.7 ± 0.3	0.4 ±0.3	0.3 ± 0.0	0.0 ±0.0	2.1 ±0.4	2.9 ±0.4	0.7 ±0.2	0.8 ± 0.1	0.9 ±0.1	0.2 ±0.0	0.1 ± 0.0	0.3 ±0.0	0.1 ± 0.0
11	t1Z0H0b1	21.0	14.2 ± 0.9	1.9 ±0.6	1.1 ± 0.1	0.5 ±0.3	7.3 ±0.8	2.2 ±0.3	3.2 ±0.4	2.4 ± 0.2	0.8 ±0.1	1.3 ±0.1	0.6 ± 0.1	0.3 ±0.0	0.3 ± 0.0
12	t1Z0H0b2	2.0	2.4 ± 0.3	0.0 ±0.0	0.5 ± 0.0	0.1 ±0.0	3.7 ±0.6	2.3 ±0.3	1.8 ±0.3	1.3 ± 0.2	0.7 ±0.1	0.5 ±0.1	0.3 ± 0.0	0.2 ±0.0	0.2 ± 0.0
13	t0Z0H1b1	10.0	11.8 ± 0.9	1.6 ±0.6	0.4 ± 0.0	0.1 ±0.0	1.5 ±0.4	0.4 ±0.1	1.3 ±0.2	0.2 ± 0.1	0.1 ±0.0	0.4 ±0.1	0.1 ± 0.0	0.0 ±0.0	0.1 ± 0.0
14	t0Z0H1b2	6.0	2.4 ± 0.4	1.4 ±0.5	0.3 ± 0.0	0.0 ±0.0	0.8 ±0.3	0.8 ±0.2	0.7 ±0.2	0.2 ± 0.1	0.4 ±0.1	0.2 ±0.0	0.1 ± 0.0	0.1 ±0.0	0.1 ± 0.0
15	t0Z0H0b1	82.0	71.8 ± 2.1	4.6 ±0.9	0.5 ± 0.0	1.4 ±0.8	1.5 ±0.4	0.7 ±0.2	4.2 ±0.4	0.5 ± 0.1	0.2 ±0.0	0.9 ±0.1	0.2 ± 0.0	0.1 ±0.0	0.3 ± 0.0
16	t0Z0H0b2	14.0	8.9 ± 0.8	1.8 ±0.6	0.2 ± 0.0	0.1 ±0.0	0.5 ±0.2	0.3 ±0.1	1.3 ±0.2	0.4 ± 0.1	0.1 ±0.0	0.4 ±0.1	0.0 ± 0.0	0.0 ±0.0	0.2 ± 0.0
	Total	308.0	233.4 ± 4.0	28.6 ±2.3	18.6 ± 0.2	10.2 ±2.0	133.1±3.4	71.3 ±1.8	64.0 ±1.7	35.8 ± 0.8	18.6 ±0.4	17.1 ±0.4	9.9 ± 0.2	5.1 ±0.1	4.8 ± 0.1

Muon channel: $\mu\mu$+jets

	Category	Data	DY	$t\bar{t}$	$t\bar{t}Z$	other backgrounds	T → tZ, M800 GeV			T → tZ, M1000 GeV			T → tZ, M1200 GeV		
							T → tZ	T → tH	T → bW	T → tZ	T → tH	T → bW	T → tZ	T → tH	T → bW
1	t1Z1H1b1	47.0	41.5 ± 1.9	11.5 ±1.5	7.1 ± 0.1	2.5 ±0.9	37.4 ±1.8	15.0 ±0.8	12.5 ±0.7	7.3 ± 0.4	3.0 ±0.2	2.4 ±0.1	1.8 ± 0.1	0.7 ±0.0	0.5 ± 0.0
2	t1Z1H1b2	29.0	10.6 ± 0.9	4.8 ±1.0	6.2 ± 0.1	0.4 ±0.0	47.0 ±2.0	38.7 ±1.3	12.8 ±0.7	9.5 ± 0.4	8.1 ±0.3	2.8 ±0.2	2.2 ± 0.1	1.7 ±0.1	0.5 ± 0.0
3	t1Z1H0b1	79.0	61.0 ± 2.2	6.9 ±1.2	5.4 ± 0.1	1.7 ±0.6	35.4 ±1.8	13.8 ±0.8	20.8 ±1.0	9.9 ± 0.4	4.0 ±0.2	4.5 ±0.2	2.6 ± 0.1	1.1 ±0.1	1.2 ± 0.1
4	t1Z1H0b2	7.0	5.3 ± 0.7	2.0 ±0.6	1.5 ± 0.1	0.5 ±0.3	18.3 ±1.3	7.9 ±0.6	6.8 ±0.6	4.9 ± 0.3	2.5 ±0.2	1.8 ±0.1	1.5 ± 0.1	0.7 ±0.0	0.5 ± 0.0
5	t0Z1H1b1	29.0	26.1 ± 1.4	5.4 ±1.0	1.0 ± 0.1	1.6 ±0.8	4.9 ±0.6	1.7 ±0.3	3.8 ±0.4	1.0 ± 0.1	0.5 ±0.1	0.9 ±0.1	0.4 ± 0.0	0.1 ±0.0	0.3 ± 0.0
6	t0Z1H1b2	7.0	4.3 ± 0.5	3.2 ±0.8	0.8 ± 0.0	0.1 ±0.0	5.5 ±0.7	4.3 ±0.4	3.8 ±0.4	1.3 ± 0.2	1.2 ±0.1	0.9 ±0.1	0.4 ± 0.0	0.3 ±0.0	0.2 ± 0.0
7	t0Z1H0b1	78.0	62.6 ± 2.2	7.0 ±1.2	1.1 ± 0.1	0.4 ±0.0	7.8 ±0.8	2.1 ±0.3	9.1 ±0.6	2.2 ± 0.2	0.7 ±0.1	2.4 ±0.1	0.7 ± 0.1	0.3 ±0.0	0.8 ± 0.0
8	t0Z1H0b2	5.0	7.3 ± 0.8	2.5 ±0.7	0.3 ± 0.0	0.1 ±0.0	3.4 ±0.5	1.9 ±0.3	4.0 ±0.4	1.1 ± 0.1	0.6 ±0.1	0.8 ±0.1	0.3 ± 0.0	0.1 ±0.0	0.3 ± 0.0
9	t1Z0H1b1	9.0	7.2 ± 0.8	1.2 ±0.5	0.5 ± 0.0	0.1 ±0.0	1.7 ±0.4	1.1 ±0.2	1.3 ±0.2	0.5 ± 0.1	0.3 ±0.1	0.2 ±0.0	0.2 ± 0.0	0.1 ±0.0	0.0 ± 0.0
10	t1Z0H1b2	10.0	2.0 ± 0.4	1.5 ±0.5	0.6 ± 0.0	0.0 ±0.0	1.9 ±0.4	3.3 ±0.4	1.0 ±0.2	0.8 ± 0.1	1.3 ±0.1	0.2 ±0.0	0.2 ± 0.0	0.3 ±0.0	0.1 ± 0.0
11	t1Z0H0b1	40.0	29.5 ± 1.4	3.2 ±0.8	1.7 ± 0.1	0.4 ±0.2	7.6 ±0.8	3.3 ±0.4	4.5 ±0.4	2.5 ± 0.2	1.2 ±0.1	1.6 ±0.1	0.9 ± 0.1	0.4 ±0.0	0.5 ± 0.0
12	t1Z0H0b2	5.0	4.0 ± 0.6	1.5 ±0.5	0.9 ± 0.1	0.1 ±0.0	4.4 ±0.6	2.9 ±0.4	3.2 ±0.4	1.4 ± 0.2	0.9 ±0.1	0.9 ±0.1	0.4 ± 0.0	0.4 ±0.0	0.2 ± 0.0
13	t0Z0H1b1	24.0	25.0 ± 1.3	3.3 ±0.8	0.5 ± 0.0	1.2 ±0.7	1.7 ±0.4	0.6 ±0.2	1.3 ±0.2	0.4 ± 0.1	0.2 ±0.0	0.3 ±0.1	0.1 ± 0.0	0.1 ±0.0	0.1 ± 0.0
14	t0Z0H1b2	10.0	6.3 ± 0.7	2.1 ±0.6	0.4 ± 0.0	0.1 ±0.0	1.1 ±0.3	2.2 ±0.3	1.7 ±0.3	0.4 ± 0.1	0.5 ±0.1	0.4 ±0.1	0.1 ± 0.0	0.1 ±0.0	0.1 ± 0.0
15	t0Z0H0b1	167.0	125.3 ± 3.0	15.3 ±1.7	0.7 ± 0.0	0.2 ±0.0	1.8 ±0.4	0.7 ±0.2	4.9 ±0.5	0.7 ± 0.1	0.3 ±0.1	1.2 ±0.1	0.2 ± 0.0	0.1 ±0.0	0.4 ± 0.0
16	t0Z0H0b2	25.0	14.6 ± 1.0	6.1 ±1.1	0.2 ± 0.0	0.1 ±0.0	0.8 ±0.3	1.0 ±0.2	2.1 ±0.3	0.3 ± 0.1	0.3 ±0.1	0.5 ±0.1	0.1 ± 0.0	0.1 ±0.0	0.1 ± 0.0
	Total	571.0	432.3 ± 5.8	77.7 ±3.9	29.0 ± 0.3	9.3 ±1.5	180.7±3.9	100.6±2.1	93.5 ±2.0	44.2 ± 0.9	25.4 ±0.5	21.8 ±0.5	12.1 ± 0.2	6.6 ±0.1	5.8 ± 0.1

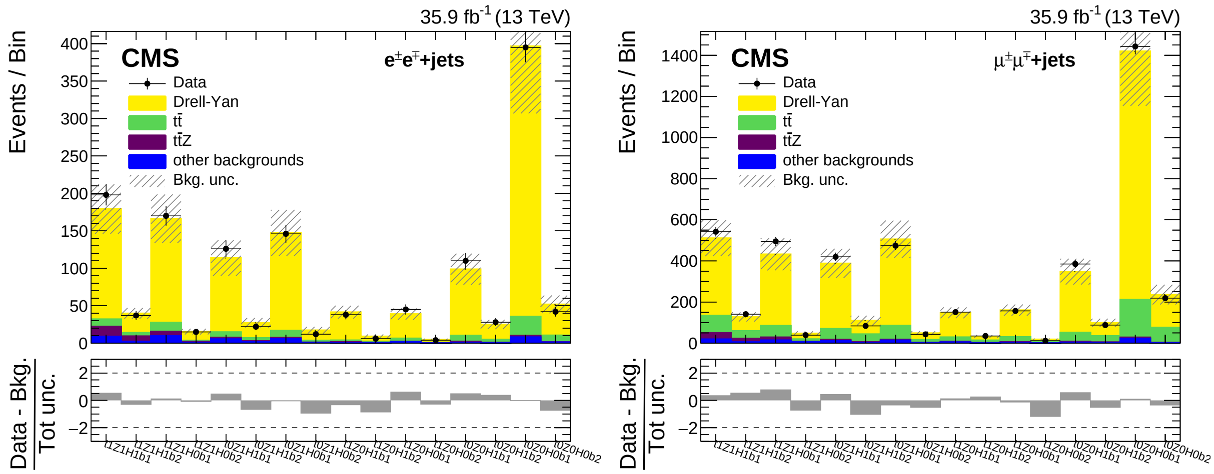


Figure 6.9: *Event counts in all the 16 mutually exclusive event categories based on the number of t, V, H and b candidates identified in $Z \to e^+e^-$ (left) and the $Z \to \mu^+\mu^-$ (right) channels in control region CR4. The hashed band shows the statistical and systematic uncertainties. The DY correction factors are applied.*

to $\mathrm{T\overline{T}}$ signal. Fig. 6.10 shows the significance for each event category using $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$, $\mathrm{T\bar{T}} \to \mathrm{tZtH}$ and $\mathrm{T\bar{T}} \to \mathrm{tZbW}$ decay channels for the benchmark T quark mass of 1000 GeV.

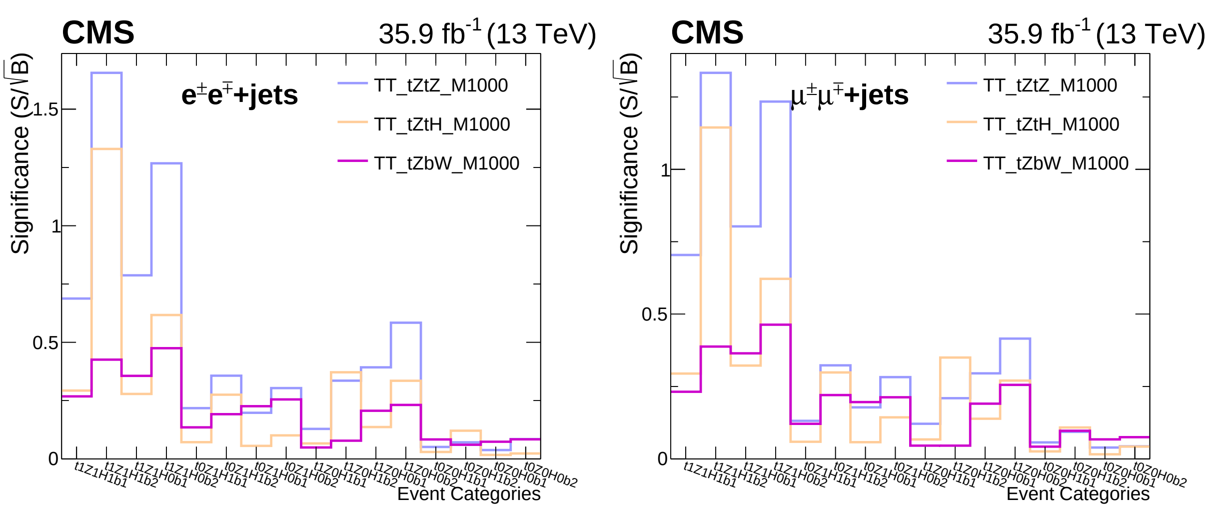


Figure 6.10: *Significances $(S/\sqrt{B})$ for sixteen event categories in $Z \to e^+e^-$ (left) and the $Z \to \mu^+\mu^-$ (right) channels in the signal region for* $\mathrm{T\bar{T}} \to \mathrm{tZtZ}$, $\mathrm{T\bar{T}} \to \mathrm{tZtH}$ *and* $\mathrm{T\bar{T}} \to \mathrm{tZbW}$ *decay channels with* $\mathrm{m_T} = 1000$ GeV.

The Table 6.5 shows the distribution of background and signal events for combined $Z \to$

e^+e^- and $Z \rightarrow \mu^+\mu^-$ channels in all sixteen categories. The signal processes shown here correspond to $\mathrm{T\bar{T}} \rightarrow$ tZtZ, $\mathrm{T\bar{T}} \rightarrow$ tZtH and $\mathrm{T\bar{T}} \rightarrow$ tZbW decays assuming $m_{\mathrm{T}} = 1000$ GeV. Significance of different event categories for each decay mode are given in columns 5, 8 and 11 of Table 6.5.

The categories that have the highest significance are more sensitive to the $\mathrm{T\overline{T}}$ signal. The 95% confidence level expected limit computed using the Bayesian likelihood technique (described in Sec. 8.1.2) can be used to select the event categories that are most sensitive to signal. This is done as follows.

First, the categories given in Table 6.5 are sorted according to their significance in the $\mathrm{T\bar{T}} \rightarrow$ tZtZ decay channel in descending order. The category with the highest significance (t1Z1H1b2) is tested for the expected limit [2]by considering it as an individual channel. The next category in the sorted list is then added as another individual channel and the expected limit is computed. If the latter category has a similar significance compared to the former category, the expected limit is also computed by merging them together to form one channel. If the addition does not improve the expected limits, the latter category is kept as a separate channel and the next category is tested by repeating the same procedure. The merging of the different event categories from the sorted list is done in this way until no further improvement is observed in the expected limit. It is observed that merging the event categories with similar significance improves the expected limit because one can use finer binning in the S_{T} distributions to capture the shape difference between signal and background.

Four event groups A, B, C and D are formed by merging the event categories t1Z1H1b2, t1Z1H0b2, t1Z1H0b1, t1Z1H1b1, t1Z0H0b2, t1Z0H0b1, t1Z0H1b2, t0Z1H1b2 and t0Z1H0b2, as they provide the highest expected limit. Merging the rest of the event categories does not improve the expected limit and hence they are discarded. The four event groups selected for the analysis are shown in Table 6.6. According Table 6.6, groups A and B are the most sensitive channels to the $\mathrm{T\overline{T}}$ signal as they capture events that have boosted top quarks, H

[2]Expected limit is calculated by using the S_{T} distributions from the background and signal simulations only.

Table 6.5: *The expected number of signal events for the benchmark point* $T\bar{T} \rightarrow tZtZ$, $T\bar{T} \rightarrow tZtH$, $T\bar{T} \rightarrow tZbW$, $M_T = 1000$ GeV, *and total standard model background events per each category of events, along with respective significance for the combined ee+jets and* $\mu\mu$*+jets channels.*

	Category	$T\bar{T} \rightarrow tZtZ$, $m_T = 1000$ GeV			$T\bar{T} \rightarrow tZtH$, $M_T = 1000$ GeV			$T\bar{T} \rightarrow tZbW$, $M_T = 1000$ GeV		
		Exp. Signal	Exp. Background	Significance	Exp. Signal	Exp. Background	Significance	Exp. Signal	Exp. Background	Significance
1	t1Z1H1b1	12.7 ± 0.7	93.1 ± 7.2	1.31	5.3 ± 0.3	93.1 ± 7.2	0.55	4.5 ± 0.3	93.1 ± 7.2	0.47
2	t1Z1H1b2	17.1 ± 0.8	34.3 ± 4.1	2.93	14.3 ± 0.5	34.3 ± 4.1	2.44	4.7 ± 0.3	34.3 ± 4.1	0.81
3	t1Z1H0b1	17.5 ± 0.8	118.5 ± 7.7	1.61	6.7 ± 0.4	118.5 ± 7.7	0.61	7.9 ± 0.4	118.5 ± 7.7	0.73
4	t1Z1H0b2	8.7 ± 0.6	13.5 ± 2.3	2.38	4.3 ± 0.3	13.5 ± 2.3	1.18	3.3 ± 0.3	13.5 ± 2.3	0.89
5	t0Z1H1b1	2.3 ± 0.3	49.0 ± 5.4	0.32	0.9 ± 0.1	49.0 ± 5.4	0.12	1.7 ± 0.2	49.0 ± 5.4	0.25
6	t0Z1H1b2	2.4 ± 0.3	14.5 ± 2.5	0.64	2.1 ± 0.2	14.5 ± 2.5	0.55	1.5 ± 0.2	14.5 ± 2.5	0.39
7	t0Z1H0b1	3.9 ± 0.4	110.4 ± 7.0	0.37	1.2 ± 0.1	110.4 ± 7.0	0.11	4.3 ± 0.3	110.4 ± 7.0	0.41
8	t0Z1H0b2	2.0 ± 0.3	14.6 ± 2.6	0.52	0.8 ± 0.1	14.6 ± 2.6	0.22	1.6 ± 0.2	14.6 ± 2.6	0.41
9	t1Z0H1b1	0.9 ± 0.2	15.0 ± 2.3	0.24	0.5 ± 0.1	15.0 ± 2.3	0.13	0.4 ± 0.1	15.0 ± 2.3	0.09
10	t1Z0H1b2	1.6 ± 0.2	5.6 ± 1.5	0.66	2.2 ± 0.2	5.6 ± 1.5	0.92	0.4 ± 0.1	5.6 ± 1.5	0.15
11	t1Z0H0b1	5.0 ± 0.4	52.6 ± 4.4	0.69	2.0 ± 0.2	52.6 ± 4.4	0.28	2.9 ± 0.2	52.6 ± 4.4	0.40
12	t1Z0H0b2	2.7 ± 0.3	9.4 ± 1.6	0.88	1.7 ± 0.2	9.4 ± 1.6	0.54	1.4 ± 0.2	9.4 ± 1.6	0.45
13	t0Z0H1b1	0.7 ± 0.2	43.8 ± 4.3	0.10	0.3 ± 0.1	43.8 ± 4.3	0.05	0.7 ± 0.1	43.8 ± 4.3	0.11
14	t0Z0H1b2	0.6 ± 0.2	13.0 ± 2.4	0.17	0.9 ± 0.1	13.0 ± 2.4	0.24	0.6 ± 0.1	13.0 ± 2.4	0.17
15	t0Z0H0b1	1.2 ± 0.2	219.8 ± 8.8	0.08	0.5 ± 0.1	219.8 ± 8.8	0.03	2.1 ± 0.2	219.8 ± 8.8	0.14
16	t0Z0H0b2	0.7 ± 0.2	32.0 ± 3.6	0.13	0.4 ± 0.1	32.0 ± 3.6	0.07	0.9 ± 0.1	32.0 ± 3.6	0.16

,V bosons and b quarks. Groups C and D are not sensitive as group A and B but, they tend to capture considerable fraction of signal events in the $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtH}$ and $\mathrm{T\bar{T}} \rightarrow \mathrm{tZbW}$ decay channels. The S_{T} distributions corresponding to these four event groups are employed, when calculating the 95% CL limits using the THETA framework as described in Chapter 8. The expected limit is improved by ~ 70 GeV by using A, B, C and D event groups instead of all sixteen categories.

Table 6.6: *Event groups that provide the highest significance for the* $\mathrm{T\overline{T}}$ *search.*

Group	N_{b}	N_{V}	N_{H}	N_{t}	Significance in the $\mathrm{T\bar{T}} \rightarrow \mathrm{tZtZ}$ channel
A	$= 1$	≥ 1	$= 0$	≥ 1	1.61
	$= 1$	≥ 1	≥ 1	≥ 1	1.31
B	≥ 2	≥ 1	$= 0$	≥ 1	2.38
	≥ 2	≥ 1	≥ 1	≥ 1	2.93
C	$= 1$	$= 0$	$= 0$	≥ 1	0.69
	≥ 2	≥ 1	$= 0$	$= 0$	0.52
D	≥ 2	≥ 1	≥ 1	$= 0$	0.64
	≥ 2	$= 0$	≥ 1	≥ 1	0.66
	≥ 2	$= 0$	$= 0$	≥ 1	0.88

Fig 6.11 shows the S_{T} distributions in final four event groups A, B, C and D in signal region for both $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ channels. The signal processes are normalized to 1 pb cross section for better visibility. It can be seen from Fig 6.11, that $\mathrm{T\overline{T}}$ signal peaks in the high-S_{T} regime ($\sim S_{\mathrm{T}} > 1500$ GeV) while the majority of the SM background contributions fall into the low-S_{T} regime ($\sim S_{\mathrm{T}} < 1500$ GeV).

As this analysis relies on the predicted background simulation, these final four event groups should be validated in the control region CR4 to account any mismodeling between data and background simulation. Fig. 6.12 shows the S_{T} distributions in final four event groups A, B, C and D in control region CR4. The hashed band shows the statistical and systematic uncertainties. All the distributions are in good agreement within the total uncertainty.

The events that fall into the final four event groups, A, B, C and D are used with their corresponding S_{T} distributions for the final search as described in Chapter 8.

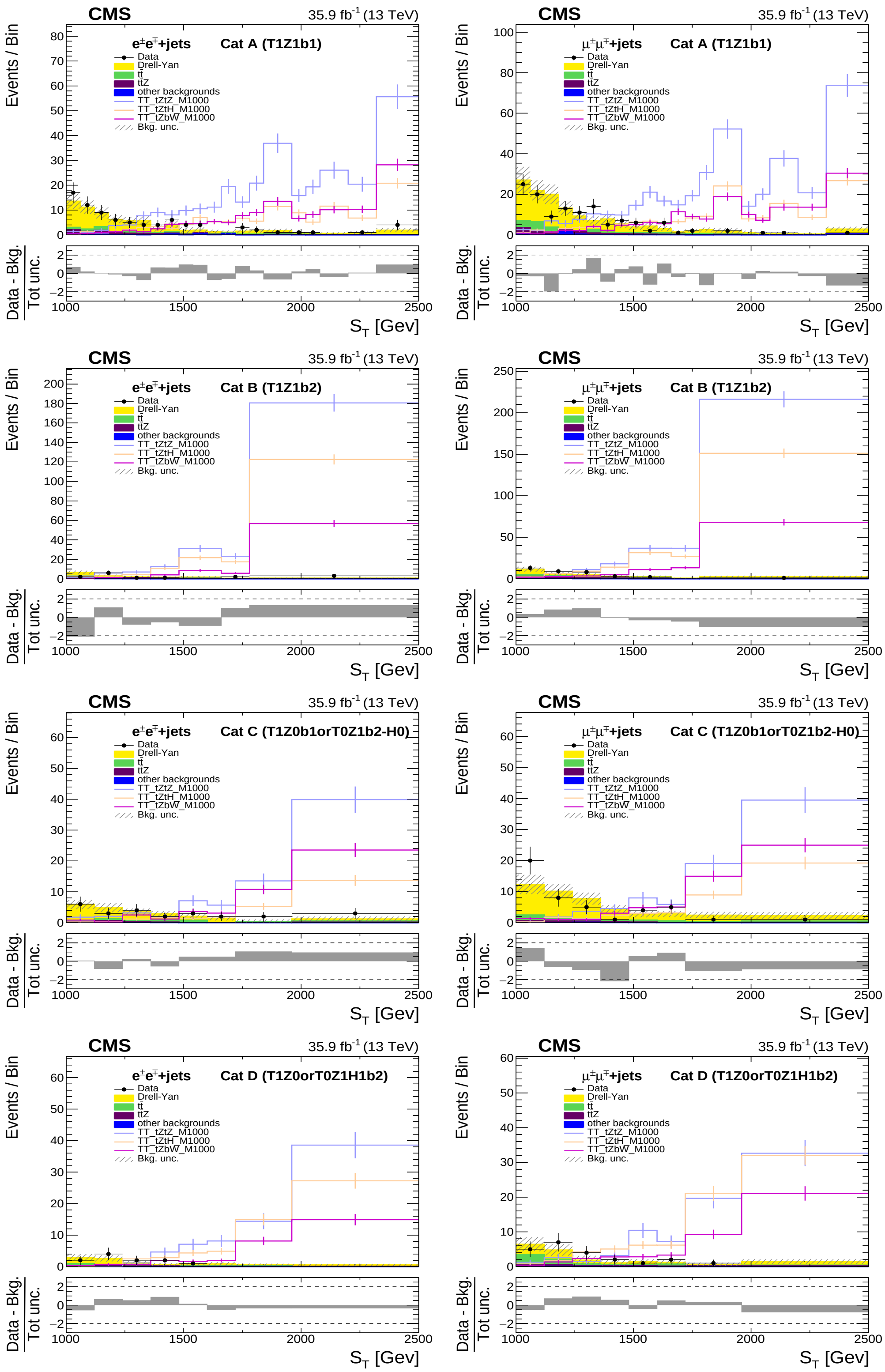


Figure 6.11: *S_{T} distributions in A, B, C and D groups for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) channels in the signal region. For visibility, signal contributions are normalized to 1 pb cross section. The DY correction factors are applied. The hashed band shows the statistical and systematic uncertainties.*

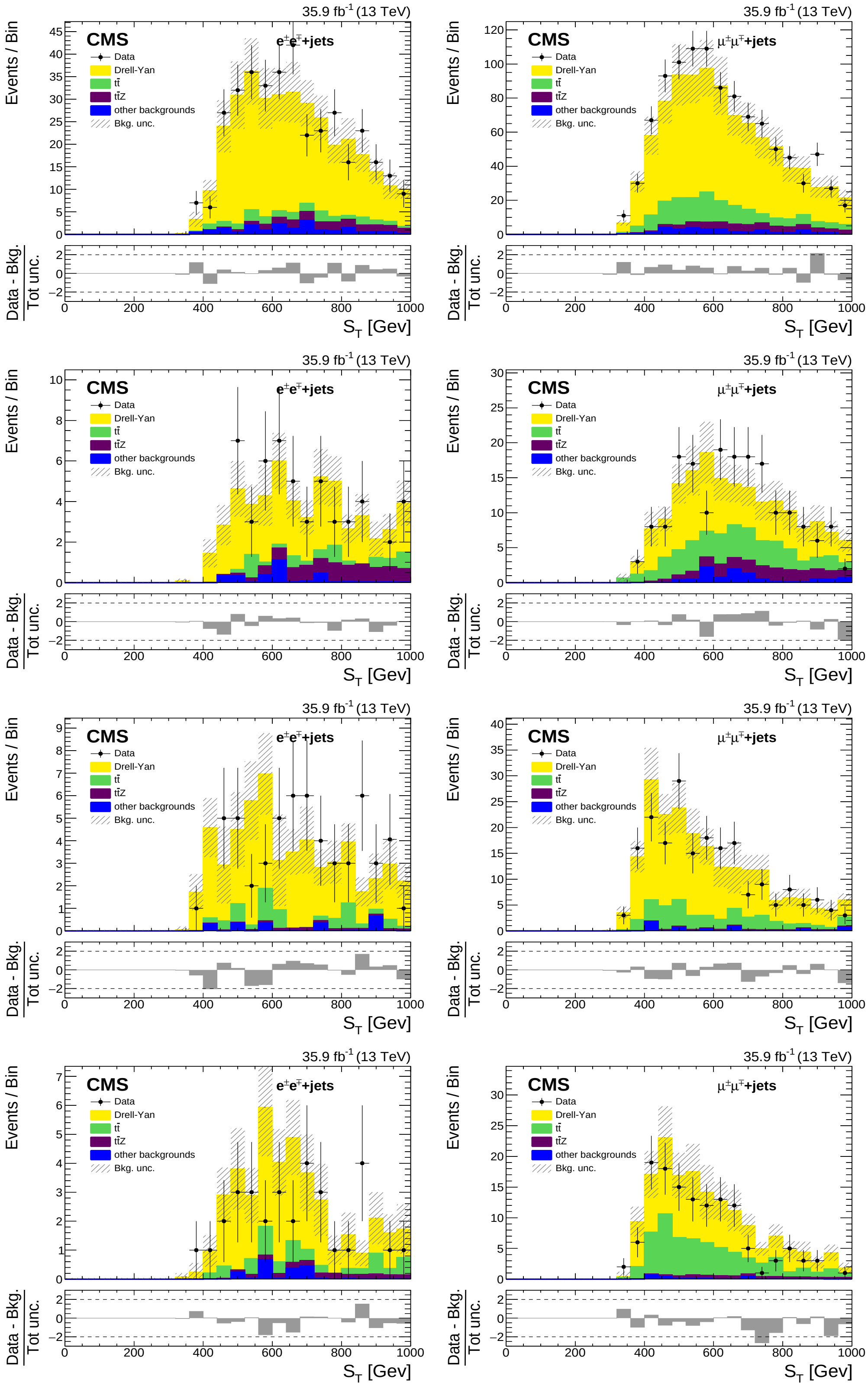


Figure 6.12: *S_{T} distributions in A, B, C and D groups for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) channels in CR4 control region. The DY correction factors are applied. The hashed band shows the statistical and systematic uncertainties.*

Chapter 7

Systematic uncertainties

The predicted rates and shapes of the physical observables obtained from the MC simulation depend on many parameters. For example, the b tagging efficiency scale factors (described in Sec. 4.2.6.6) applied to MC simulation affect both the rate and shape of the observable. There can be uncertainties associated with these parameters, which are known as *systematic uncertainties*.

The contribution from these uncertainties should also be considered together with statistical uncertainties. The total uncertainty band shown in the kinematic distributions in the previous sections includes contributions from statistical and systematic uncertainties added in quadrature. This helps to determine the agreement between data and MC simulation. Also, the systematic uncertainties are used in the final statistical analysis as described in Sec. 8.1.

This section describes the systematic uncertainties considered in the analysis, in the cross section $\times$ branching ratio calculations. The systematic uncertainties can be divided into two main categories as *normalization* and *shape* uncertainties.

7.1 Normalization uncertainties

These uncertainties only affect the normalization (rate) of the distribution. These include:

- **Luminosity uncertainty** - The uncertainty on integrated luminosity is measured to be 2.5 %[147]. This affects both background and signal yields.

- **SM background measurement uncertainty** - The production cross-sections for the main SM background processes DY+jets, $t\bar{t}$ + jets, $t\bar{t}Z$ and dibosons were measured at center-of-mass energies $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV. The uncertainties correspond to these measurements were estimated to be 15 % each[148–150].

- **Lepton identification uncertainty** - The identification of electrons and muons for the analysis is described in Sections 4.2.4 and 4.2.5. Efficiency scale factors for lepton identification are applied to MC simulation. The uncertainty on lepton identification efficiency scale factors is measured to be 3%.

- **Trigger efficiency uncertainty** - The uncertainty on trigger efficiency scale factors is measured to be 1%.

7.2 Shape uncertainties

These uncertainties affect both normalization and shape of the distribution. The contribution from the shape uncertainties is accounted for a given parameter by producing two templates, which are further referred as *up-* and *down-* templates. In order to produce these templates, the nominal template is varied up and down separately moving the respective parameter by + and $-$ 1σ.

The produced up and down templates represent one standard deviation of the corresponding systematic uncertainty. They are used with the nominal template in the statistical model described in Sec. 8.1 for the limit calculations, where the method is commonly known as *vertical template morphing*. The shape uncertainties include:

- **Renormalization(μ_R) and factorization (μ_F) scale uncertainties** - The renormalization(μ_R) and factorization (μ_F) scales are used in the cross section measurements described in Sec. 4.1. μ_R is a measure of energy scale at which the partonic scattering process occurred, while μ_F is a measure of the threshold between perturbative and non-perturbative regimes. The values of μ_R^2 and μ_F^2 are chosen such that they match to energy scale of Q^2(momentum transfer of the hard scattering process)[6]. Since there are no prior constraints assigned on these two parameters, possible variations of them should be considered as systematic uncertainties[6].

The μ_R and μ_F scale uncertainties provide a considerable contribution to the systematic uncertainties. This contribution is evaluated by considering six variations correspond to weights 2 (up variations) and 1/2 (down variations) for both μ_R and μ_F parameters. Two of these distributions are obtained by varying both μ_R and μ_F up or down simultaneously, while other four distributions are obtained by varying only one of them at a time by fixing the other one. The final up and down variations are obtained by calculating the maximum bin by bin variation from all six variations with respect to nominal[6].

- **PDF uncertainty** - As described in Sec. 4.1, PDFs are used to model the momentum transfer of the scattered partons in the proton-proton scattering process. The uncertainties associated with the PDFs used in the analysis should be considered. Event weights that correspond to 100 replicas are stored in the MC samples for this purpose. These replicas are obtained by performing the fit to the nominal PDF distribution 100 times using different random numbers for the initialization when minimizing the fit[6].

  This is described in PDF4LHC procedure[84] in detail, where the systematic uncertainty is calculated using these replicas. The observable is re-computed using the specified weight for each replica. The up and down variations that correspond to one standard deviation are obtained by taking the bin by bin root mean square values of all hundred distributions[6].

- **Pileup uncertainty** - An inelastic cross-section of 69.2 mb is used to perform the pileup reweighting as described in Sec. 4.2.8. The uncertainty related to this cross-section measurement is estimated to be 5%. The uncertainty related to the pileup reweighting procedure is taken into account by producing up and down variations that correspond to one standard deviation by varying the nominal proton-proton inelastic cross section up and down by 5%[145].

- **DY scale factor uncertainty** - The DY scale factors were derived to correct the AK4 jet multiplicity distribution mainly in the higher AK4 jet multiplicity bins. The scale factors listed in Table 5.7 are applied to DY MC simulation by default, depending on the number of AK4 jets reconstructed within an event. The corresponding uncertainty is obtained by producing the up and down templates without applying the DY scale factors to DY MC simulation. Since, both up and down templates are equivalent in this case, they are known as *one sided* systematic uncertainties.

- **b tagging uncertainty** - b tagging plays an important role in the current analysis as it provides a better discrimination between the signal and SM background events. Efficiency scale factors for identified heavy and light flavor jets are applied on simulation as described in Sec. 4.2.6.6. The uncertainty on the b tagging efficiency scale factors is evaluated by varying the nominal scale factor for identified b-, c- and light-flavor jets up and down by their uncertainty and producing templates corresponding to one standard deviation[6]. Two different methods are used to evaluate the tagging efficiency scale factors for heavy (b and c) and light flavors. So the uncertainties measured for them are considered as uncorrelated[6].

- **Jet energy scale and jet energy resolution uncertainty** - The observed differences between data and simulation, for both jet energy scale and resolution of the reconstructed AK4 and AK8 jets, are corrected by using dedicated measurements as described in Sections 4.2.6.1 and 4.2.6.5.

  The systematic uncertainty associated with these measurements are evaluated by simultaneously varying four vectors of AK4 and AK8 jets with provided uncertainties up and down and producing the templates correspond to one standard deviation[126;151].

- **H/W/Z tagging uncertainty** - W/Z and H tagged AK8 jets carry additional sources of systematic uncertainties apart from the jet energy scale and resolution uncertainties mentioned earlier. These are due to the pruned jet mass and τ_2/τ_1 selections described in Sections 4.2.6.12 and 4.2.6.13. The efficiency scale factors for these selections are derived using the difference between the peaks of the pruned jet mass distributions obtained from data and MC simulations[6]. There are two sources of systematic uncertainties associated with pruned jet mass selection known as *jet mass scale* and *jet mass resolution*. They are estimated to be 2.3% and 18%. The corresponding uncertainty on τ_2/τ_1 selection is estimated to be 8%.

  The up and down templates correspond to one standard deviation of the uncertainty are obtained by varying the jet mass scale up and down by 2.3% simultaneously with the jet energy scale and by varying the jet mass resolution up and down by 18% simultaneously with the jet energy resolution according to Eq. 7.1[6].

$$m_{\text{smeared jet}} = m_{\text{generator level jet}} + (1 \pm 0.18)(m_{\text{reconstructed jet}} - m_{\text{generator level jet}}) \tag{7.1}$$

  Similarly, up and down templates corresponding to one standard deviation on the uncertainty of the τ_2/τ_1 selection are obtained by varying the corresponding τ_2/τ_1 efficiency scale factor up and down by 8%[6]. An additional uncertainty of $0.041 \times ln(p_{\mathrm{T}}/200)$ GeV is applied on high p_{T} ($p_{\mathrm{T}} > 200$ GeV) W/Z and H tagged AK8 jets as a function of AK8 jet p_{T}. For H tagging, the uncertainty on b-tagging is considered on the b-tagged subjets[6].

- **top tagging uncertainty** - top tagged AK8 jets carry an additional source of systematic uncertainty apart from the jet energy scale and resolution uncertainties mentioned earlier. This is due to the soft-drop jet mass and τ_3/τ_2 selections described in Sec. 4.2.6.14. An efficiency scale factor is derived by using the difference between the peaks of the soft-drop mass distributions from data and MC simulation corresponds to

$\tau_3/\tau_2 = 0.81$ selection. The uncertainty corresponds to this selection is evaluated to be 9%. The systematic templates correspond to this uncertainty is obtained by varying the N-subjetiness scale factor ($\tau_{32} = 0.81$) up and down by 0.09 per top tagged AK8 jet in an event.

- **Boosted W/Z, H and top misidentification rate uncertainties** - Boosted W/Z, H and top jets can be misidentified by the jet substructure taggers in the background MC simulation as described in Sec. 4.2.7. Mistag rate scale factors are derived to apply on background MC simulation per tagged W/Z, H and top jet. The uncertainty per mistagged W/Z, H and top jets is estimated to be 5%, 14% and 7% respectively.

  The systematic templates correspond to one standard deviation of the uncertainty are obtained by varying the misidentification scale factors mentioned in Table 4.9 by their uncertainties per misidentified jet in the background MC simulation.

The summary of systematic uncertainties used in the analysis are shown in Table 7.1.

Fig 7.1 and Fig. 7.2 show the nominal (black), up (red) and down (blue) variations for the S_{T} variable for jet energy scale (JEC), jet energy resolution (JES), pileup (PU), $\mu_{R,F}$ scale (Scale), pdf (Pdf), b-tagging (BTAG) , DY scale factors (NAK4), mistag rates for W,H and top tagging (Mistag W, Mistag H, Mistag Top) for total background simulation in $Z \rightarrow e^+e^-$+jets and $Z \rightarrow \mu^+\mu^-$+jets channels in signal region. It can be seen from these figures that maximum contribution for total background comes from the μ_R and μ_F scale uncertainties.

These systematic uncertainties are used to marginalize the maximum likelihood function described in Sec. 8.1.2.

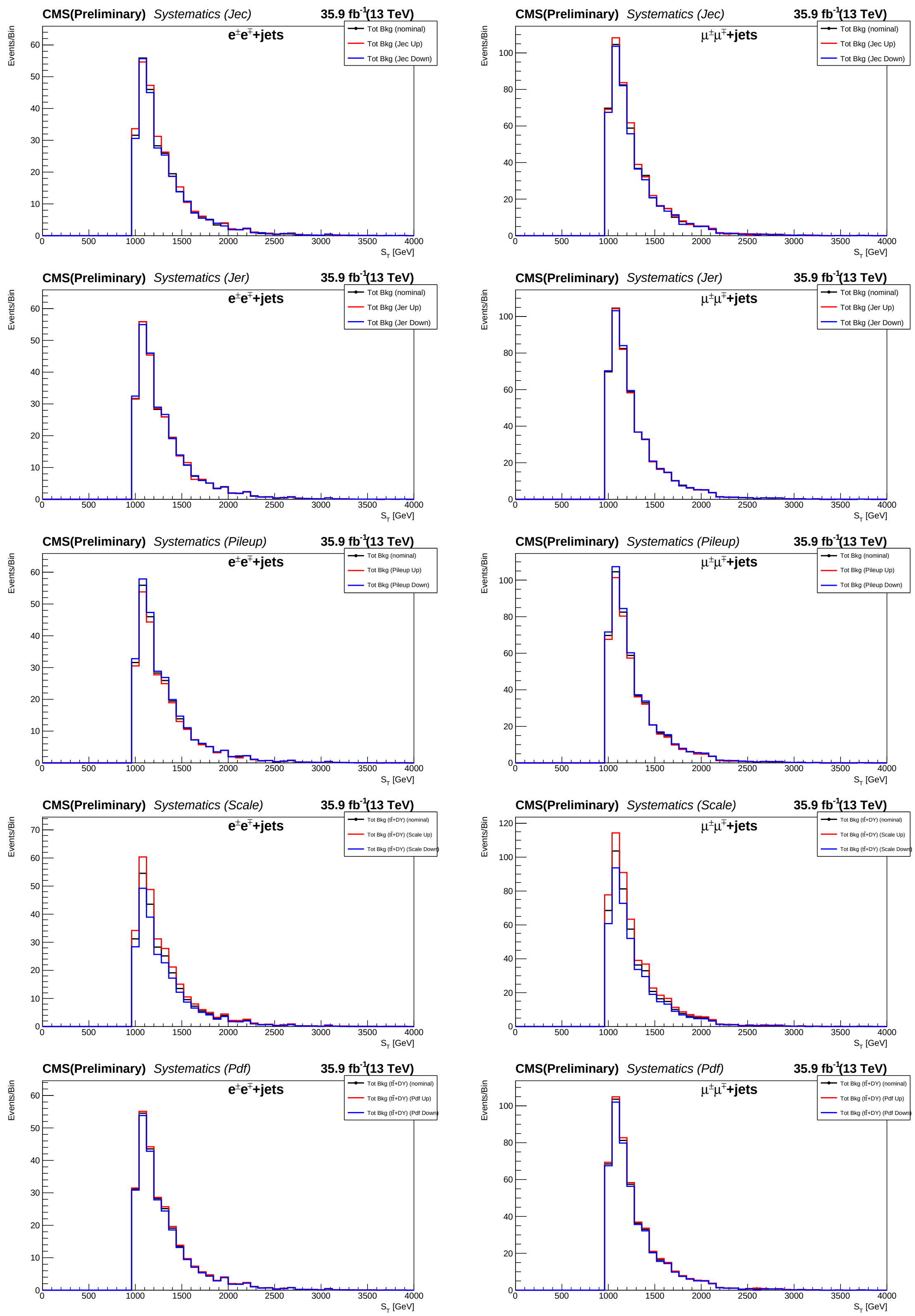


Figure 7.1: *nominal and $\pm 1\sigma$ S_{T} variations for jet energy scale (JEC), jet energy resolution (JES), pileup (PU), $\mu_{R,F}$ scale (Scale) and pdf (Pdf) for total background simulation in ee+jets (left) and $\mu\mu$+jets (right) channel in the signal region.*

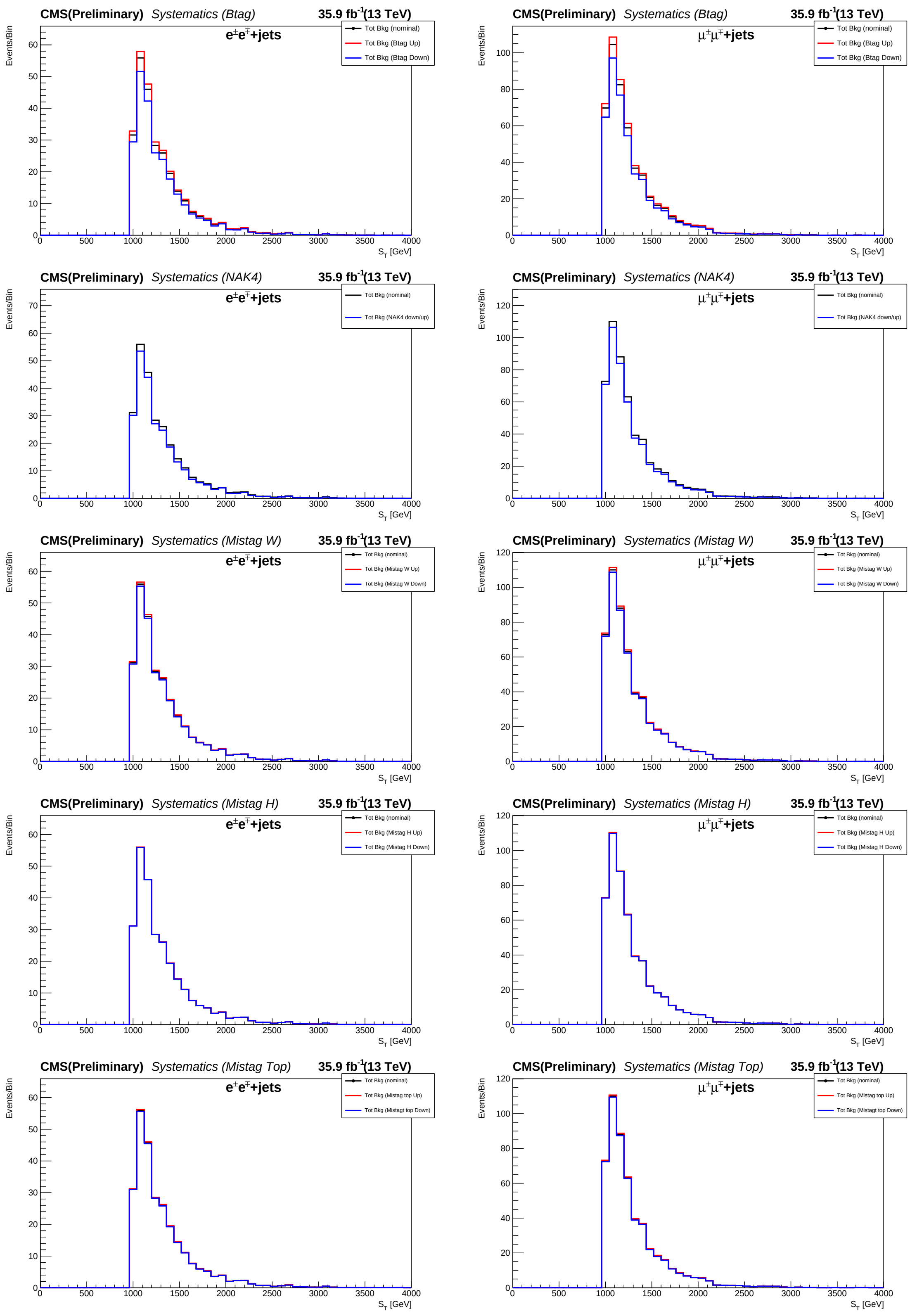


Figure 7.2: *nominal and* $\pm 1\sigma$ S_{T} *variations for b-tagging (BTAG) , DY scale factors (NAK4), mistag rates for W,H and top tagging (Mistag W, Mistag H, Mistag Top) for total background simulation in ee+jets (left) and* $\mu\mu$*+jets (right) channel in the signal region.*

Table 7.1: *Systematic uncertainties considered on background and signal simulation.*

	Source	Uncertainty (%)
Normalization	Integrated luminosity	2.5
	DY+jets measurement	15
	$t\bar{t}$ + jets measurement	15
	$t\bar{t}Z$ measurement	15
	Diboson measurement	15
	Lepton identification efficiency	3
	Trigger efficiency	1
Shape & Normalization	Jet energy scale	
	Jet energy resolution	
	Pileup	
	b tagging scale factor	
	DY correction factor	
	W/Z/H-tagging	
	Top-tagging (τ_{32})	
	subjet b-tagging for H-tag	
	PDF	
	$\mu_{R/F}$ scale	
	mistag rate SF for W-tag	
	mistag rate SF for H-tag	
	mistag rate SF for top-tag	

Chapter 8

Interpretation and results

The final four event groups, A, B, C and D described in Sec. 6.2 are used in the statistical analysis. The statistical concepts and models used for the analysis are described in Sec. 8.1. Table 8.1 and 8.2 provide the number of observed data events in each group for the $\mathrm{T\overline{T}}$ production in both $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ channels. It also presents the expected number of background yields along with the expected number of signal yields for the three benchmark $\mathcal{B}$s, $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = 100\%$ (tZtZ) , $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \rightarrow \mathrm{tH}) = 50\%$ (tZtH), and $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \rightarrow \mathrm{bW}) = 50\%$ (tZbW). The uncertainties for the expected background event yields include the contributions from both statistical and systematic uncertainties. The uncertainties for data and $\mathrm{T\overline{T}}$ signal include statistical uncertainty only. The observed number of data events and background yields agree within their uncertainties.

The S_T distributions in groups A, B, C and D in $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ channels are combined together to perform the statistical analysis. This is to increase the background MC statistics of each category and thereby to reduce the contribution from the statistical uncertainty. These combined rebinned S_T distributions are shown in Fig. 8.1. The $\mathrm{T\overline{T}}$ signal shown corresponds to the benchmark branching ratio, $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = 100\%$ with mass of the T quark, $m_\mathrm{T} = 1200$ GeV. It is multiplied by a factor of two for better visibility. No indication of new signal is present in any of the S_T distributions and hence 95% C.L exclusion limits are set on the T quark mass using the Bayesian technique as described in Sec. 8.1.

Table 8.1: *The expected and observed number of events in the final four event groups, A, B, C and D in the $Z \to e^+e^-$ channel. The expected numbers of* $\mathrm{T\overline{T}}$ *events are shown for three benchmark branching ratios,* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = 100\%$ *(tZtZ) ,* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{tH}) = 50\%$ *(tZtH), and* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{bW}) = 50\%$ *(tZbW) for T quark masses,* $m_\mathrm{T} = 800$ *and* 1200 GeV*. The background uncertainties include both statistical and systematic uncertainties added in quadrature.*

Event group	A	B	C	D
DY+jets	54.9±5.2	9.0±1.9	17.0±2.4	7.2±1.4
$\mathrm{t\bar{t}}$ + jets	7.9±1.7	1.7±0.8	3.2±1.1	1.8±0.8
$\mathrm{t\bar{t}Z}$	8.2±0.8	4.9±0.6	1.3±0.2	1.3±0.2
Other backgrounds	3.0±1.7	0.9±0.7	0.6±0.4	0.1±0.1
Total	74.1±6.2	16.5±2.5	22.2±2.9	10.4±1.8
Data	84	15	25	11
tZtZ, m_T=800 GeV	54.9±2.2	43.6±2.0	9.6±0.9	9.6±0.9
tZtH, m_T=800 GeV	24.8±1.0	26.7±0.8	4.2±0.3	6.5±0.4
tZbW, m_T=800 GeV	24.5±1.0	17.9±0.6	5.4±0.3	5.2±0.3
tZtZ, m_T=1200 GeV	3.6±0.1	3.3±0.1	0.9±0.1	0.8±0.1
tZtH, m_T=1200 GeV	1.6±0.1	1.8±0.1	0.4±0.1	0.6±0.1
tZbW, m_T=1200 GeV	1.6±0.1	1.3±0.1	0.5±0.1	0.4±0.1

Table 8.2: *The expected and observed number of events in the final four event groups, A, B, C and D in the $Z \to \mu^+\mu^-$ channel. The expected number of* $\mathrm{T\overline{T}}$ *events are shown for three benchmark branching ratios,* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = 100\%$ *(tZtZ) ,* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{tH}) = 50\%$ *(tZtH), and* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{bW}) = 50\%$ *(tZbW) for T quark masses,* $m_\mathrm{T} = 800$ *and* 1200 GeV*. The background uncertainties include both statistical and systematic uncertainties added in quadrature.*

Event group	A	B	C	D
DY+jets	102.5±10.2	15.8±3.1	36.8±4.4	10.2±2.1
$\mathrm{t\bar{t}}$ + jets	18.4±3.4	6.8±1.7	5.7±1.5	6.3±1.7
$\mathrm{t\bar{t}Z}$	12.5±1.2	7.7±1.0	2.0±0.3	2.3±0.3
Other backgrounds	4.2±1.3	0.9±0.4	0.5±0.3	0.3±0.1
Total	137.6±11.6	31.2±4.5	45.0±5.0	19.1±3.2
Data	126	36	45	22
tZtZ, m_T=800 GeV	72.8±2.5	65.4±2.4	10.9±1.0	11.9±1.0
tZtH, m_T=800 GeV	33.0±0.8	40.0±0.9	5.5±0.3	8.4±0.4
tZbW, m_T=800 GeV	34.9±0.9	26.2±0.8	7.0±0.4	7.0±0.4
tZtZ, m_T=1200 GeV	4.4±0.1	3.7±0.1	1.2±0.1	1.0±0.1
tZtH, m_T=1200 GeV	2.0±0.1	2.2±0.1	0.6±0.1	0.8±0.1
tZbW, m_T=1200 GeV	1.9±0.1	1.4±0.1	0.7±0.1	0.5±0.1

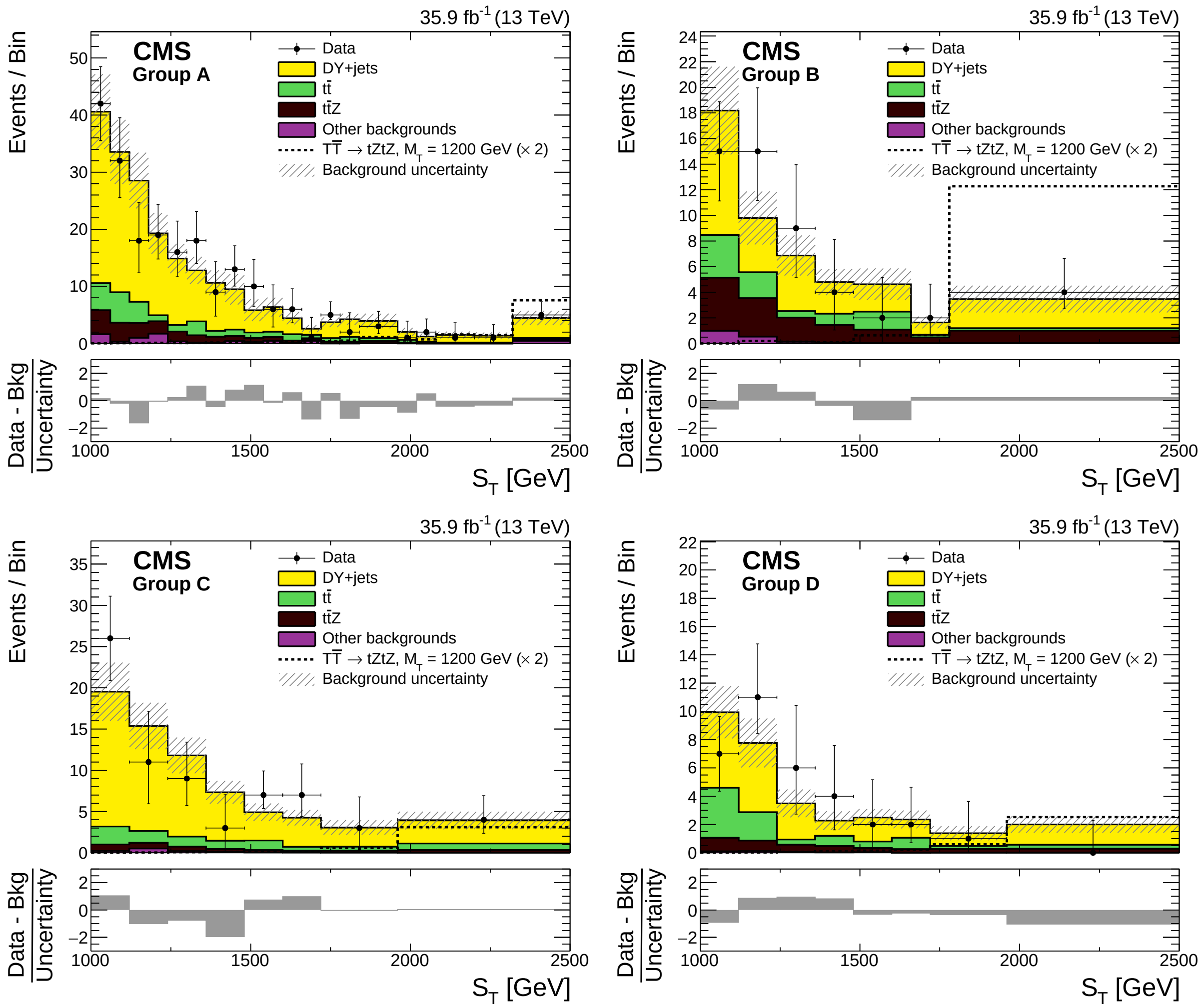


Figure 8.1: *The S_{T} distributions for the final groups A, B, C and D with combined $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels. The points with vertical and horizontal bars represent the data distributions where vertical bars show the central 68% CL intervals for Poisson-distributed data and horizontal bars show the bin widths. The colored shaded histograms represent the expected background distributions. The dotted lines show the $\mathrm{T}\overline{\mathrm{T}}$ signal for benchmark branching ratio, $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = 100\%$ with $m_{\mathrm{T}} = 1200$ GeV scaled by a factor of 2 for visibility. The grey hatched bands represent the statistical and systematic uncertainties in the total background simulation added in quadrature.*

8.1 Statistical methods

This section describes the statistical concepts used for the analysis including the Bayesian technique. It also describes the THETA framework, which is used to compute the 95% C.L expected and observed limits.

8.1.1 Maximum-likelihood method

Maximum-likelihood method can be employed to analyze the distribution of a measured data sample. It is capable of providing the true values for unknown parameters of the data distribution. The relationship between the measured data and the the unknown parameters is given by the maximum-likelihood function $L(x|\theta)$ given by Eq. 8.1 [6].

$$L(x|\theta) = \prod_{i=1}^{N} f(x_i|\theta) \tag{8.1}$$

Here, x represents the measured data set that consists of N measured values. The data set $x = x_1, x_2, x_3, ..., x_i, ..., x_N$, can be thought as a distribution with N bins where x_i represents the data yield in i^{th} bin. θ represents the unknown parameters that should be measured. θ defines how the data is distributed within the so-called *model*. Then, $f(x_i|\theta)$ defines how the data x is distributed in relation to θ. $f(x_i|\theta)$ is known as the probability density function (pdf) [6].

It is assumed that a discrete probability distribution follows a Poisson distribution. The probability of observing k events is given by the Poisson distribution as shown in Eq. 8.2 where, λ is the rate parameter.

$$\text{Probability}(\text{k}, \lambda) = \frac{\text{e}^{-\lambda}\lambda^{\text{k}}}{\text{k}!} \tag{8.2}$$

So $f(x_i|\theta)$ can be written in terms of Poisson distribution as given by Eq. 8.3.

$$f(x_i|\lambda_i(\theta)) = \frac{\lambda_i(\theta)^{x_i} e^{-\lambda_i(\theta)}}{x_i!} \tag{8.3}$$

The parameter dependent expectation value for each bin i is given by $\lambda_i(\theta)$. Parameters θ can be estimated by maximizing the likelihood function, that satisfy the relation given by Eq. 8.4[6].

$$\frac{\partial lnL(x|\theta)}{\partial\theta} = 0 \tag{8.4}$$

If M systematic uncertainties are involved with parameters θ, they can be incorporated to maximum-likelihood function by expanding $L(x|\theta)$ using an additional term $D(\nu) = \prod_{j=1}^{M} d(\nu_j|\hat{\nu}_j, \sigma_{\nu_j})$ as given by Eq. 8.5[6].

$$L(x|\theta(\nu),\nu) = \prod_{i=1}^{N} f(x_i|\theta(\nu))D(\nu) = \prod_{i=1}^{N}\prod_{j=1}^{M} f(x_i|\theta(\nu_j))d(\nu_j|\hat{\nu}_j, \sigma_{\nu_j}) \tag{8.5}$$

These new parameters ν are known as *nuisance parameters* such that individual systematic uncertainty contribution is represented by free parameter ν_j. ν_j influences the $\theta(\nu)$ when maximizing the likelihood function. $\hat{\nu}_j$ and σ_{ν_j} represent the mean and width of $D(\nu)$. ν_j is chosen such that, $\hat{\nu}_j = 0$ and $\sigma_{\nu_j} = 1$. So, nominal value for θ is given by when $\nu_j = 0$ and ±1 standard deviation of systematic uncertainty j is given by when $\nu_j = \pm1$ respectively[6].

There are two type of systematic contributions so-called *rate* and *shape* uncertainties as described in Chapter 7. Rate uncertainties only affect the normalization while shape uncertainties affect both normalization and shape of the distributions of parameters θ. *Log-normal* and *Gaussian* terms are included in the term $d(\nu_j|\hat{\nu}_j, \sigma_{\nu_j})$ to account for the contributions coming from rate and shape uncertainties respectively[6]. The maximum-likelihood function $L(x|\theta(\nu),\nu)$ given by Eq. 8.5 has a wide range of applications. It is used in the Bayesian technique mentioned Sec 8.1.2 to perform the limit calculations.

This analysis uses Bayesian likelihood based technique to compute 95% confidence level limits. In comparison to the method discussed earlier, $\theta(\nu)$ for this analysis correspond to the predicted standard model (SM) background yields for a given variable[6]. These backgrounds (θ) are dependent on a set of systematic uncertainties given by parameter ν[6].

These background MC contributions are predicted by using the MC generators mentioned in Sec. 4.1.4. As these predictions are obtained using a finite number of MC events, there is an additional source of uncertainty; called *statistical uncertainty* contributions for the modeling of θ that depends on MC statistics[6]. The variations of systematic uncertainties for a given distribution are correlated between the bins of the distribution and hence there is only one parameter ν_j, per systematic uncertainty. Conversely, statistical uncertainties are un-correlated between bins such that there is one parameter ν_j per each bin of the distribution[6]. These statistical uncertainties are also treated in the same way as systematic uncertainties.

When there are large enough event statistics, the statistical uncertainty can also be treated as a Gaussian distribution according to *Barlow-Beeston light* method[152;153]. This is accomplished by choosing proper binning for the variable (S_{T} in this case) such that the statistical uncertainty per each bin is $< 20\%$[6].

The response function can be derived to both types of systematic uncertainties. For rate uncertainties, the variation is created by correlated shift of the corresponding parameter ν_j across all bins. For shape uncertainties, the variations corresponding to $\nu_j = \pm 1$ are created by varying the given uncertainties by one standard deviation up and down from the nominal value. This procedure is know as *vertical template morphing*[6]. The final response function for parameters considering systematic templates ($\theta(\nu_j)$) is calculated by interpolating the templates for $\nu_j = 0, \pm 1$ with a cubic function[6].

For the current analysis, all the systematic uncertainties for the maximum likelihood function are created for S_{T} variable for the final event groups A, B, C and D using vertical template morphing.

8.1.2 Bayesian method

A quantitative test should be performed to predict the existence of the the new physics signal after full event selection. This is done by constructing a hypothesis $H(\mu)$ that depends on the strength of the new physics signal μ. Then, $\mu > 0$ and $\mu = 0$ corresponds to the *presence of signal hypothesis* and *null (or background only) hypothesis* respectively. Presence of signal is usually proportional to its production cross section. Using the statistical inference, one can either reject the hypothesis (null hypothesis) or derive the *confidence intervals* for which μ has true values at a certain *confidence level*(CL)[6]. There are two statistical models called *Frequentist* and *Bayesian*, in order to test the hypothesis constructed. Current search uses Bayesian approach to obtain the confidence intervals.

Bayesian concept is based on the Bayesian theorem given in Eq. 8.6 with $P(B) \neq 0$ and A, B are events of the experiment[6].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{8.6}$$

$P(A|B)$ represents the *conditional probability* where likelihood of event A to occur given B is true. $P(A)$ and $P(B)$ are independent probabilities of observing events A and B.

In order to validate the parameter μ defined earlier, Eq. 8.6 can be written as shown in Eq. 8.7[6].

$$P(\mu|x) = \frac{L(x|\mu)\pi(\mu)}{\int L(x|\mu')\pi(\mu')d\mu'} \tag{8.7}$$

Here, $P(\mu|x)$ is known as the *posterior probability* of having *signal strength* μ, given a set of data x. $L(x|\mu)$ is the likelihood function mentioned in Eq. 8.1 and it depends on signal strength μ. $\pi(\mu)$ is known as the prior probability that varies according to μ[6]. Usually, $\pi(\mu)$ is assigned values as shown in Eq. 8.8 and known as having a *flat prior* distribution[6].

$$\pi(\mu) = 0 \text{ for } \mu < 0 \text{ and } \pi(\mu) = 1 \text{ for } \mu \geq 0 \tag{8.8}$$

The denominator of Eq 8.7 is normalized to obtain probability values between 0 and 1. The probability $P(\mu|x)$ gives the *degree of belief* for the parameter μ to have a specific value. The likelihood function $L(x|\mu)$ can be expanded using background processes with their corresponding systematic uncertainties as given by Eq. 8.9[6].

$$L(x|\mu) = \int L(x|\mu(\nu), \theta(\nu))\ \pi(\nu)d\nu \tag{8.9}$$

Then, $L(x|\mu)$ is known as the *marginal likelihood* function. ν is the set of systematic uncertainties (nuisance parameters) that affect both background ($\theta(\nu)$) and signal strength ($\mu(\nu)$) parameters. The prior probability distribution $\pi(\nu)$ of the ν systematic parameters can be assumed in a similar way according to the definition of $D(\nu)$ as given in Eq. 8.5. *Barlow-Beeston light* method described in Sec. 8.1.1 is employed to incorporate the statistical uncertainties associated with the parameters $\theta(\nu)$[6].

The confidence interval related to Bayesian model can be calculated according to Eq. 8.10, where $1-\alpha$ represent the confidence level (CL).

$$1-\alpha = \int_{0}^{\mu_u} p(\mu|x)d\mu \tag{8.10}$$

The upper limit on the production cross section is calculated at 95% CL (i.e $1-\alpha = 0.95$). μ_u represents a free parameter that represents the upper bound. The corresponding signal model can be excluded at 95% CL if, $\mu_u <$ theoretical cross-section predicted at the corresponding signal strength μ[6].

The μ_u can be calculated using either measured data or MC simulation. The upper limit computed using measured data (MC simulation) is known as *observed* (*expected*) limit[6]. Expected limit can be used as a relative parameter to optimize the analysis strategy. For current analysis, it is used to merge the most optimal event categories described in Sec. 6.2. The combinations that provide the highest expected limit are chosen as the final four event groups A, B, C and D. The limit calculation is performed by using the THETA package[146;154] described in Sec. 8.1.3.

8.1.3 THETA framework

The THETA framework is used to compute the 95% CL *observed* and *expected* limits described in Sec. 8.1.2. This framework consists of many python scripts, programs and libraries[146;154]. It is a template-based framework that takes a set of root files as input. For current analysis four separate root files for each event group A, B, C and D are provided as the input.

Each root file consists of S_{T} distributions for data, SM background processes (DY+jets, $\mathrm{t\bar{t}}+\mathrm{jets}$, $\mathrm{t\bar{t}Z}$ and other backgrounds) and signal processes for T quark masses in the range 800 - 1800 GeV. The Signal distributions are produced by considering different $\mathcal{B}$ hypothesis described in Sec. 4.1.3, where only one $\mathcal{B}$ is considered at a time. In addition, the root file also contains the up and down templates corresponding to one standard deviation of the systematic uncertainty for the possible systematic uncertainties described in Chapter 7, for both background and signal MC processes. Different naming conventions are used to identify the data, SM background, signal and systematic uncertainty distributions.

The integration given by the Eq. 8.10 is performed numerically to compute the limits. The provided S_{T} distributions are used as the input. The *Markov Chain Monte Carlo* (MCMC) method is used to calculate this integral. In this method the integral is approximated by using a large number of iteration. In each iteration, a point in the parameter space of signal strength (μ) is randomly selected and the likelihood function is calculated. The calculated value is added to the integral. This process is repeated until all the points in the parameter space are considered according a probability density function [6;154].

The complete process is known as a *Markov Chain* and consists of 10000 iterations. Since, an approximation of the integral given by Eq. 8.10 is obtained per Markov Chain, ten such chains are run in the calculation of observed limit and the median is taken as the 95% CL observed limit. In case of calculating the expected limit, 1000 such chains are run and the median is taken as the 95% CL expected limit. In addition, the confidence intervals at 68% and 95% CLs are also provided [6].

8.2 95 % CL exclusion limits

The Bayesian likelihood technique (described in Sec. 8.1.2) [155] is employed using THETA framwork (described in Sec. 8.1.3) [154] to compute 95% confidence level (CL) upper limits on the $\mathrm{T\overline{T}}$ cross section. This is done by performing a simultaneous maximum likelihood fit to the rebinned S_{T} distributions correspond to event groups A, B, C and D, shown in Fig. 8.1

for different T quark masses and for different branching ratios mentioned in Sec. 5.1. The S_{T} distributions are rebinned to make sure that statistical uncertainty on the total background in each bin of the S_{T} distribution is $< 20\%$. The systematic uncertainties described in Chapter 7 are used as nuisance parameters to marginalize the maximum likelihood fit. The normalization uncertainties are taken as log-normal priors and shape uncertainties are taken as Gaussian-distributed priors as described in Sec. 8.1.1. Parameter for the signal cross section is taken as a uniform(flat) prior (as described in Sec. 8.1.1).

Fig. 8.2 shows the 95% CL upper limits on the $\mathrm{T\overline{T}}$ cross section as a function of T quark mass (solid black line) for three benchmark branching ratios, $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = 100\%$ (tZtZ) , $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \rightarrow \mathrm{tH}) = 50\%$ (tZtH), and $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \rightarrow \mathrm{bW}) = 50\%$ (tZbW). The dotted black line shows the expected median upper limits. Inner (green) and outer (yellow) bands show the ±1 and ±2 standard deviation uncertainties in the expected median limits respectively. The theoretical signal cross section is given by the blue dotted-dashed line[1].

If the computed 95% CL upper limit on cross-section is less than the theoretical cross-section, the considered signal model can be excluded below that point. According to Fig. 8.2, these exclusion mass points for *observed* and expected cross-sections are given by the intersection of the blue dotted-dashed line (theoretical signal cross section) with the solid black line (95% CL observed cross-section) and the dashed line (95% CL expected cross-section) respectively. According to this figure, the T quarks with masses less than $1280, 1185$ and 1120 GeV are excluded for three corresponding benchmark branching ratios by comparing the observed cross section limits with theoretical cross sections. The corresponding expected upper limits are $1290, 1175$ and 1115 GeV respectively.

Apart from the three benchmark branching ratios mentioned earlier, the 95% CL observed and expected limits on the T quark mass are computed for a set of different $\mathcal{B}$s to cover the full phase space. Different $\mathcal{B}$s are considered according the relationship $\mathcal{B}(\mathrm{T} \rightarrow \mathrm{tZ}) + \mathcal{B}(\mathrm{T} \rightarrow \mathrm{tH}) + \mathcal{B}(\mathrm{T} \rightarrow \mathrm{bW}) = 1$, described in Sec. 2.3.3. The correct signal composition for each $\mathcal{B}$ combination is obtained according to the description provided in Sec. 4.1.3.

The computed 95% CL observed and expected limits on T quark mass corresponding to

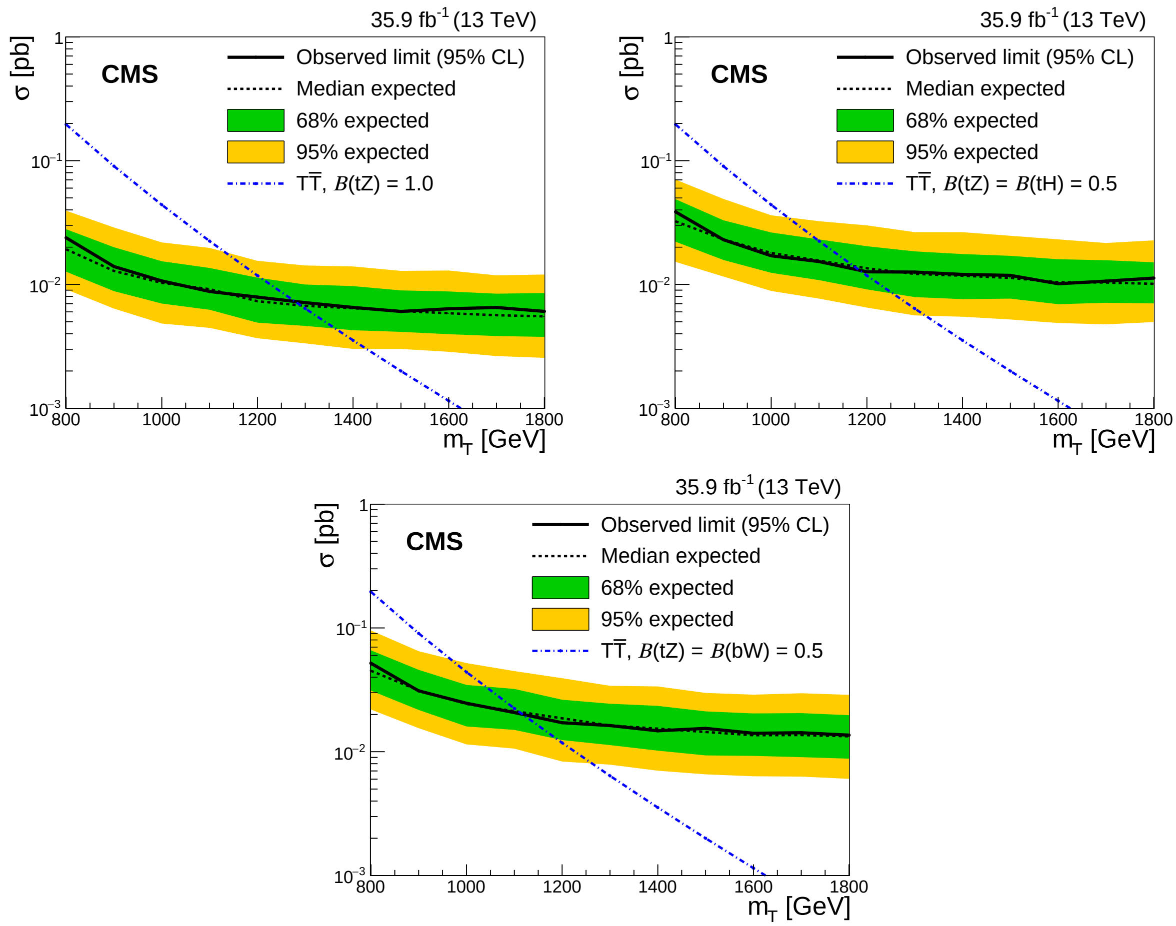


Figure 8.2: *The 95% CL observed (solid line) and expected (dashed line) upper limits on the* $\mathrm{T\overline{T}}$ *cross section as a function of the T quark mass, in* GeV. *The results are shown for three benchmark branching ratios,* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = 100\%$ *(upper left),* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{tH}) = 50\%$ *(upper right) and* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = \mathcal{B}(\mathrm{T} \to \mathrm{bW}) = 50\%$ *(lower). Inner (green) and outer (yellow) bands show the* ± 1 *and* ± 2 *standard deviation uncertainties in the expected limits respectively. The theoretical signal cross section is displayed by blue dotted-dashed line.*

different $\mathcal{B}$ scenarios are shown in Table 8.3 and Fig. 8.3.

Table 8.3: *The 95% CL observed and expected limits on T quark mass correspond to different $\mathcal{B}$ scenario.*

Combination	$\mathcal{B}(T \to tH)$	$\mathcal{B}(T \to bW)$	$\mathcal{B}(T \to tZ)$	95% CL Observed limit (GeV)	95% CL Expected limit (GeV)
1	0.0	0.0	1.0	1280	1290
2	0.2	0.0	0.8	1250	1255
3	0.4	0.0	0.6	1205	1205
4	0.6	0.0	0.4	1140	1135
5	0.8	0.0	0.2	1035	1020
6	1.0	0.0	0.0	< 800	< 800
7	0.0	0.2	0.8	1235	1230
8	0.2	0.2	0.6	1180	1185
9	0.4	0.2	0.4	1120	1120
10	0.6	0.2	0.2	1015	1005
11	0.8	0.2	0.0	< 800	< 800
12	0.0	0.4	0.6	1155	1160
13	0.2	0.4	0.4	1095	1090
14	0.4	0.4	0.2	985	975
15	0.6	0.4	0.0	< 800	< 800
16	0.0	0.6	0.4	1070	1065
17	0.2	0.6	0.2	940	945
18	0.4	0.6	0.0	< 800	< 800
19	0.0	0.8	0.2	915	910
20	0.2	0.8	0.0	< 800	< 800
21	0.0	1.0	0.0	< 800	< 800

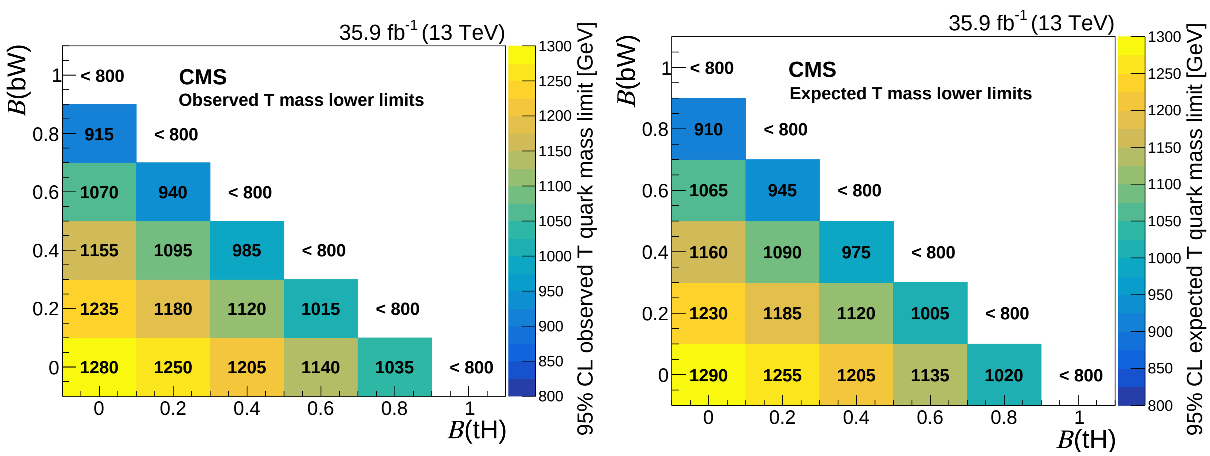


Figure 8.3: *The 95% CL observed (left) and expected (right) limits on the T quark mass, in* GeV, *for different branching fractions, where* $\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) + \mathcal{B}(\mathrm{T} \to \mathrm{tH}) + \mathcal{B}(\mathrm{T} \to \mathrm{bW}) = 1$

Chapter 9

Conclusion

A search is presented for pair production of vector-like top (T) quarks by using the proton-proton collision data collected by the CMS experiment in 2016 at a center-of-mass energy $\sqrt{s} = 13$ TeV that corresponds to an integrated luminosity of 35.9 fb^{-1}. Events are considered where one of the T quark decays via $\text{T} \rightarrow \text{tZ}$ and other T quark via $\text{T} \rightarrow \text{tZ}$, tH or bW. Events having two opposite sign leptons, e^+e^- or $\mu^+\mu^-$ consistent with originating from a Z boson decay and jets are investigated. Events are categorized according to the number of identified W, Z and Higgs bosons and top and bottom quark candidates. Search is optimized for the categories that are sensitive to the $\text{T}\overline{\text{T}}$ signal.

The fit to the S_T distributions in those categories is performed for the statistical analysis. A simultaneous binned maximum-likelihood fit is performed for observed data in those categories and 95% CL upper limits are obtained on $\text{T}\overline{\text{T}}$ production cross section by assuming different branching ratios. Theoretical cross sections predicted for $\text{T}\overline{\text{T}}$ production are compared with upper limits as a function T quark mass and 95% CL lower limits are placed on T quark mass for different branching fractions. T quark mass values below 1280 GeV are excluded in case of 100% branching fraction for $\text{T} \rightarrow \text{tZ}$.

Bibliography

[1] Micha B Czakon, Paul Fiedler, and Alexander Mitov. Total top quark pair production cross section at hadron colliders through $o(\alpha\frac{4}{S})$. *Phys. Rev. Lett.*, 110:252004, 2013. arXiv:1303.6254 ,doi:10.1103/PhysRevLett.110.252004.

[2] Wikipedia. Standard model, 2018. URL https://en.wikipedia.org/wiki/Standard_Model. online accessed : 21-Aug-2018.

[3] CMS Collaboration. Observation of a new boson with mass near 125 gev in pp collisions at $\sqrt{s} = 7$ and 8 tev. *JHEP*, 06:081, 2013. arXiv:1303.4571 ,doi:10.1007/JHEP06(2013)081.

[4] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett.*, B716:30–61, 2012. arXiv:1303.4571 ,doi:10.1007/JHEP06(2013)081.

[5] Wikipedia. boson, 2018. URL https://en.wikipedia.org/wiki/Boson. online accessed : 23-Aug-2018.

[6] J. Haller D. Nowatschin and A. Meyer. Search for vector-like quarks using jet substructure techniques with the cms experiment. CERN-THESIS-2017-142, 2017. URL http://cds.cern.ch/record/2282807.

[7] Diego Bettoni. The standard model lagrangian, 2018. URL http://www.fe.infn.it/~bettoni/particelle/Strong/SMLagrangian.pdf. online accessed : 25-Aug-2018.

[8] Wikipedia. Dirac qequation, 2018. URL https://en.wikipedia.org/wiki/Dirac_equation. online accessed : 25-Aug-2018.

[9] Wikipedia. Gell-mann matrices, 2018. URL https://en.wikipedia.org/wiki/Gell-Mann_matrices. online accessed : 26-Aug-2018.

[10] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2016.

[11] Particle data group. The particle adventure: The fundamentals of matter and force, 2014. URL http://www.particleadventure.org/quark_confinement.html. online accessed : 26-Aug-2018.

[12] Wikipedia. Asymptotic freedom, 2018. URL https://en.wikipedia.org/wiki/Asymptotic_freedom. online accessed : 26-Aug-2018.

[13] K. Kleinknecht F. J. Gilman and B. Renk. The cabibbo-kobayashi-maskawa quark-mixing matrix, 2004. URL http://pdg.lbl.gov/2004/reviews/kmmixrpp.pdf. online accessed : 27-Aug-2018.

[14] A. DJOUADI. The anatomy of electro weak symmetry breaking, tome i: The higgs boson in the standard model. *Phys. Rept.*, 457:1–216, 2008. arXiv:hep-ph/0503172 .

[15] Particle data group. Review of particle physics, 2016. doi:10.1088/1674-1137/40/10/100001 .

[16] CERN Document Server. Higgs physics, 2018. URL https://cds.cern.ch/record/1638469/plots. online accessed : 29-Aug-2018.

[17] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Broken symmetries and the Goldstone theorem. *Adv. Part. Phys.*, 2:567–708, 1968.

[18] Heiner Tholen, Alexander Schmidt, and Johannes Haller. *Search for single production of a heavy vector-like T quark decaying to a top quark and a Higgs boson with the CMS Experiment*. PhD thesis, Hamburg U, 2017. URL http://cds.cern.ch/record/2290835. presented at 2017-07-10.

[19] S. Dittmaier et al. Handbook of lhc higgs cross sections: 2. differential distributions, 2012. arXiv:1201.3084 ,doi:10.5170/CERN-2012-002.

[20] C. Quigg. Particle physics after the higgs-boson discovery : Opportunities for the large hadron collider, 2015. URL https://www.semanticscholar.org/paper/Particle-Physics-after-the-Higgs-Boson-Discovery-%3A-Quigg/5fa9bf32621925fccfec828d75f69e709abe6714.

[21] Wikipedia. Dark matter, 2018. URL https://en.wikipedia.org/wiki/Dark_matter. online accessed : 01-Oct-2018.

[22] Wikipedia. Physics beyond the Standard Model, 2018. URL https://en.wikipedia.org/wiki/Physics_beyond_the_Standard_Model. online accessed : 01-Oct-2018.

[23] Quantum Diairies. The Hierarchy Problem: why the Higgs has a snowball?s chance in hell, 2012. URL https://www.quantumdiaries.org/2012/07/01/the-hierarchy-problem-why-the-higgs-has-a-snowballs-chance-in-hell/. online accessed : 01-Oct-2018.

[24] Wikipedia. Supersymmetry, 2018. URL https://en.wikipedia.org/wiki/Supersymmetry. online accessed : 01-Oct-2018.

[25] T. Bringmann. Dark matter lecture, 2014. URL https://www.mn.uio.no/fysikk/english/people/aca/troms/teaching/DarkMatter/dm_figures_07.pdf.

[26] M. Quiros. Higgs bosons in extra dimensions. *Mod. Phys. Lett.*, A30:1540012, 2015. arXiv:1311.2824.

[27] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Prez-Victoria. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev.*, D88(9):094010, 2013. arXiv:1306.0572 ,doi:10.1103/PhysRevD.88.094010.

[28] D. Choudhury, Timothy M. P. Tait, and C. E. M. Wagner. Beautiful mirrors and precision electroweak data. *Phys. Rev.*, D65:053002, 2002. arXiv:hep-ph/0109097 ,doi:10.1103/PhysRevD.65.053002.

[29] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999. arXiv:hep-ph/9905221 ,doi:10.1103/PhysRevLett.83.3370.

[30] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson. The Littlest Higgs. *JHEP*, 07:034, 2002. arXiv:hep-ph/0206021 ,doi:10.1088/1126-6708/2002/07/034.

[31] J. A. Aguilar-Saavedra. Identifying top partners at LHC. *JHEP*, 11:030, 2009. arXiv:0907.3155 ,doi:10.1088/1126-6708/2009/11/030.

[32] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997. http://root.cern.ch/.

[33] F. del Aguila, L. Ametller, Gordon L. Kane, and J. Vidal. Vector Like Fermion and Standard Higgs Production at Hadron Colliders. *Nucl. Phys.*, B334:1–23, 1990. doi:10.1016/0550-3213(90)90655-W.

[34] ATLAS Collaboration. Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 08:105, 2015. arXiv:1505.04306 ,doi:10.1007/JHEP08(2015)105.

[35] CMS Collaboration. Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:085, 2017. arXiv:1706.03408 ,doi:10.1007/JHEP11(2017)085.

[36] CMS Collaboration. Search for pair production of vector-like quarks in the $\mathrm{bW\bar{b}W}$ channel from proton-proton collisions at $\sqrt{s} = 13$ tev. *Phys. Lett. B*, 779:82, 2018. arXiv:1710.01539 ,doi:10.1016/j.physletb.2018.01.077.

[37] CMS Collaboration. Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ tev. arXiv:1805.04758v1, 2018.

[38] ATLAS Collaboration. Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ tev

pp collisions with the atlas detector. *JHEP*, 08:052, 2017. arXiv:1705.10751 ,doi:10.1007/JHEP08(2017)052.

[39] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying to high-p_{T} W bosons and b quarks in the epton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector. *JHEP*, 10:141, 2017. arXiv:1707.03347 ,doi:10.1007/JHEP10(2017)141.

[40] ATLAS Collaboration. Search for pair production of vector-like quarks into high-p_{T} W bosons and top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector. arXiv:1806.01762, 2018.

[41] ATLAS Collaboration. Search for pair- and single-production of vector-like quarks in final states with at least one z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV. arXiv:1806.10555, 2018.

[42] CMS Collaboration. CMS Beyond-two-generations (B2G) Public Physics Results, 2018. URL `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G#B2G_Preliminary_Results`.

[43] Wikipedia. SLAC National Accelerator Laboratory, 2018. URL `https://en.wikipedia.org/wiki/SLAC_National_Accelerator_Laboratory`. online accessed : 01-Oct-2018.

[44] Wikipedia. Cyclotron, 2018. URL `https://en.wikipedia.org/wiki/Cyclotron`. online accessed : 01-Oct-2018.

[45] Wikipedia. Storage ring, 2018. URL `https://en.wikipedia.org/wiki/Storage_ring`. online accessed : 01-Oct-2018.

[46] Toda S. *MEASUREMENT OF THE $Z_{\gamma\gamma}$ PRODUCTION CROSS SECTION AT PROTON-PROTON COLLISIONS WITH THE CMS EXPERIMENT*. PhD thesis, Kansas State University, 2016.

[47] CERN accelerating science. The accelerator complex, 2018. URL https://home.cern/about/accelerators. online accessed : 18-SEP-2018.

[48] Julie Haffner. The CERN accelerator complex. complexe des cclrateurs du CERN, Oct 2013. URL https://cds.cern.ch/record/1621894. General Photo.

[49] CERN accelerating science. Cms, 2018. URL https://home.cern/about/experiments/cms. online accessed : 19-SEP-2018.

[50] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008. doi:10.1088/1748-0221/3/08/S08002.

[51] LHCb Collaboration. The LHCb detector at the LHC. *JINST*, 3:S08005, 2008. doi:10.1088/1748-0221/3/08/S08005.

[52] P. La Rocca and F. Riggi. The upgrade programme of the major experiments at the Large Hadron Collider. *J. Phys. Conf. Ser.*, 515:012012, 2014. 10.1088/1742-6596/515/1/012012.

[53] The CMS Collaboration. CMS Luminosity-Public Results, Public CMS Luminosity Information, 2018. URL https://cms-service-lumi.web.cern.ch/cms-service-lumi/publicplots/int_lumi_cumulative_pp_2.pdf. online accessed : 20-SEP-2018.

[54] Burkhard Schmidt. The High-Luminosity upgrade of the LHC: Physics and Technology Challenges for the Accelerator and the Experiments. *Journal of Physics: Conference Series*, 706:022002, 04 2016. doi:10.1088/1742-6596/706/2/022002.

[55] Spannagel S. *CMS Pixel Detector Upgrade and Top Quark Pole Mass Determination.* Berlin; New York: Springer, 2017.

[56] University of Zurich Physik-Institut. CMS wiki pages, 2017. URL https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates. online accessed : 24-SEP-2018.

[57] The CMS collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.

[58] Lucas Taylor. Superconducting magnet, 2011. URL http://cms.web.cern.ch/news/superconducting-magnet. online accessed : 24-SEP-2018.

[59] S. Norberg. Inner Tracker, 2015. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSExperiment#InnerTracker. online accessed 24-SEP-2018.

[60] J. .L. Leonard. *Z BOSON CROSS SECTION MEASUREMENT USING AT THE LHC*. PhD thesis, UNIVERSITY OF WISCONSIN - MADISON, 2011.

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

[62] CMS collaboration. CMS Technical Design Report for the Pixel Detector Upgrade, 2012. doi:10.2172/1151650.

[63] CMS collaboration. CMS tracker performance and readiness for LHC Run II. *Nucl. Instrum. Meth.*, A824:67–69, 2016. doi:10.1016/j.nima.2015.09.046.

[64] CMS Collaboration. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007. doi:10.1088/1748-0221/2/04/P04004.

[65] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*, volume Technical Design Report CMS. CERN, Geneva, 1997. URL https://cds.cern.ch/record/349375.

[66] Radziej M. *Resonant Slepton production in the two muons + jets final state at* $\sqrt{s} = 8$ *TeV*. PhD thesis, RWTH AACHEN UNIVERSITY, 2014.

[67] USCMS and ECAL/HCAL Collaborations. The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c. *Eur. Phys. J.*, C60:359–373, 2009. doi:10.1140/epjc/s10052-009-0959-5 ,doi:10.1140/epjc/s10052-009-1024-0.

[68] Ganna Dolinska. *Measurement of the Cross Sections of $t\bar{t}$ Production in the Boosted Regime with the CMS Detector at $\sqrt{s} = 8$ TeV*. PhD thesis, Hasylab, DESY, Hamburg, 2016. doi:10.3204/PUBDB-2016-03989.

[69] L. Taylor. Muon Detectors, 2011. URL http://cms.web.cern.ch/news/muon-detectors. online accessed 26-SEP-2018.

[70] Florian Bechtel. *The underlying event in proton-proton collisions*. PhD thesis, Hamburg U., 2009. doi:10.3204/DESY-THESIS-2009-015.

[71] L. Taylor. Muon Drift Tubes, 2011. URL http://cms.web.cern.ch/news/muon-drift-tubes. online accessed 26-SEP-2018.

[72] D. Acosta. Muon Detectors in CMS, 2002. URL http://www.phys.ufl.edu/~acosta/cms/cmshi.pdf. online accessed 26-SEP-2018.

[73] CMS Collaboration. The CMS trigger system. *JINST*, 12(01):P01020, 2017. arXiv:1609.02366 ,doi:10.1088/1748-0221/12/01/P01020.

[74] Thomas Lenzi. Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors. Master's thesis, U. Brussels (main), 2013. arXiv:1306.0858.

[75] E. Widl and R. Fruhwirth. Application of the Kalman alignment algorithm to the CMS tracker. *J. Phys. Conf. Ser.*, 219:032065, 2010. doi:10.1088/1742-6596/219/3/032065.

[76] CMS Collaboration. The CMS Particle Flow Algorithm. In *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pages 295–304, 2013. arXiv:1401.8155.

[77] Andy Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011. arXiv:1101.2599 ,doi:10.1016/j.physrep.2011.03.005.

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

[79] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989. arXiv:hep-ph/0409313.

[80] R. S. Thorne, A. D. Martin, W. J. Stirling, and G. Watt. Parton distributions for the LHC. In *Proceedings, 15th International Workshop on Deep-inelastic scattering and related subjects (DIS 2007). Vol. 1 and 2: Munich, Germany, April 16-20, 2007*, pages 305–308, 2007. doi: 10.3204/proc07-01/27. arXiv:0706.0456 ,doi:10.3204/proc07-01/27.

[81] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. arXiv:hep-ph/0409146 ,doi:10.1088/1126-6708/2004/11/040.

[82] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. arXiv:0709.2092 ,doi:10.1088/1126-6708/2007/11/070.

[83] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007. arXiv:0707.3088 ,doi:10.1088/1126-6708/2007/09/126.

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

[85] 19436. HERA - A Proposal for a Large Electron Proton Colliding Beam Facility at DESY, 1981.

[86] Torbjorn Sjostrand. Jet Fragmentation of Nearby Partons. *Nucl. Phys.*, B248:469–502, 1984. doi:10.1016/0550-3213(84)90607-2.

[87] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983. doi:10.1016/0370-1573(83)90080-7.

[88] G. Marchesini and B. R. Webber. Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation. *Nucl. Phys.*, B310:461–526, 1988. doi:10.1016/0550-3213(88)90089-2.

[89] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys.*, B238:492–528, 1984. doi:10.1016/0550-3213(84)90333-X.

[90] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP*, 01:013, 2007. hep-ph/0611129 ,doi:10.1088/1126-6708/2007/01/013.

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

[92] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. hep-ph/0603175 ,doi:10.1088/1126-6708/2006/05/026.

[93] Torbjrn Sjstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015. arXiv:1410.3012 ,doi:10.1016/j.cpc.2015.01.024.

[94] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008. arXiv:0803.0883 ,doi:10.1140/epjc/s10052-008-0798-9.

[95] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. arXiv:1405.0301 ,doi:10.1007/JHEP07(2014)079.

[96] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016. arXiv:1512.00815 ,doi:10.1140/epjc/s10052-016-3988-x.

[97] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002. arXiv:hep-ph/0204244 ,doi:10.1088/1126-6708/2002/06/029.

[98] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. arXiv:0710.3820 ,doi:10.1016/j.cpc.2008.01.036.

[99] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. arXiv:1002.2581 ,doi:10.1007/JHEP06(2010)043.

[100] CMS Collaboration. Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV. Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016. URL `https://cds.cern.ch/record/2235192`.

[101] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008. arXiv:0706.2569 ,doi:10.1140/epjc/s10052-007-0490-5.

[102] Pierre Billoir. Progressive track recognition with a Kalman like fitting procedure. *Comput. Phys. Commun.*, 57:390–394, 1989. doi:10.1016/0010-4655(89)90249-X.

[103] Pierre Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nucl. Instrum. Meth.*, A294:219–228, 1990. doi:10.1016/0168-9002(90)91835-Y.

[104] R. Mankel. A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. *Nucl. Instrum. Meth.*, A395:169–184, 1997. doi:10.1016/S0168-9002(97)00705-5.

[105] CMS Collaboration. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(06):P06005, 2015. arXiv:1502.02701 ,doi:10.1088/1748-0221/10/06/P06005.

[106] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998. doi:10.1109/5.726788.

[107] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *J. Phys.*, G34:N343, 2007. doi:10.1088/0954-3899/34/12/N01.

[108] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. *CMS-PAS-PFT-09-001*, 2009.

[109] CMS Collaboration. Performance of CMS Muon Reconstruction in Cosmic-Ray Events. *JINST*, 5:T03022, 2010. arXiv:0911.4994 ,doi:10.1088/1748-0221/5/03/T03022.

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

[111] CMS Collaboration. Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. *CMS-PAS-PFT-10-001*, 2010.

[112] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett.*, B659:119–126, 2008. arXiv:0707.1378 ,doi:10.1016/j.physletb.2007.09.077.

[113] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008. arXiv:0802.1188 ,doi:10.1088/1126-6708/2008/04/005.

[114] CMS collaboration. Cut based electron id for run 2, 2018. URL `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2`.

[115] CMS collaboration. Electron cut-based 80XID WPs. Scale factors for 80X, 2018. URL `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Older_Efficiencies_and_Scale_Fac`.

[116] CMS collaboration. Tight muon, 2018. URL `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Tight_Muon`.

[117] Markus Radziej. Resonant Slepton Production in the Two Muon + Jets Final State at $\sqrt{s}$ = 8 TeV. Master's thesis, RWTH Aachen U., 2014-04. URL https://web.physik.rwth-aachen.de/~hebbeker/theses/radziej_master.pdf.

[118] CMS collaboration. Results on the full 2016 data, 2018. URL https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults#Results_on_the_full_2016_data.

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

[120] Jeff Tseng and Hannah Evans. Sequential recombination algorithm for jet clustering and background subtraction. *Phys. Rev.*, D88:014044, 2013. arXiv:1304.1025 ,doi:10.1103/PhysRevD.88.014044.

[121] Jim Dolen, Justin Pilot, Alexx Perloff, Ben Kreis, Julie Hogan, Marc Osherson, and Kevin Nash. Jet Algorithms and Jet Substructure HATS, 2018. URL https://indico.cern.ch/event/726636/contributions/2990334/attachments/1666720/2672296/HATS2018_Jets1.pdf.

[122] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. arXiv:0802.1189 ,doi:10.1088/1126-6708/2008/04/063.

[123] CMS collaboration. Jet Identification for the 13 TeV data Run2016, 2018. URL https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016.

[124] CMS collaboration. Jet Analysis, 2018. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetAnalysis.

[125] CMS Collaboration. Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016, Jun 2018. URL http://cds.cern.ch/record/2622157.

[126] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. arXiv:1607.03663 ,doi:10.1088/1748-0221/12/02/P02014.

[127] Bahareh Roozbahani. Jet Energy Scale and Resolution Measurement at CMS, Aug 2017. URL `https://indico.fnal.gov/event/11999/session/15/contribution/55/material/slides/0.pdf`.

[128] Wikipedia. b-tagging, 2018. URL `https://en.wikipedia.org/wiki/B-tagging`. online accessed : 22-Oct-2018.

[129] CMS 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. URL `https://cds.cern.ch/record/2138504`.

[130] CMS Collaboration. Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV. *JHEP*, 03:136, 2011. arXiv:1102.3194 ,doi:10.1007/JHEP03(2011)136.

[131] CMS Collaboration. Performance of b-Tagging Algorithms in Proton Collisions at 13 TeV using the 2016 Data, 2018. URL `https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeVICHEP2016`.

[132] CMS collaboration. Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC, 2017. URL `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco`.

[133] CMS collaboration. Methods to apply b-tagging efficiency scale factors, 2018. URL `https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods`.

[134] Quantum Diaries. Boost!, 2005-2016. URL `https://www.quantumdiaries.org/2012/08/05/boost/`.

[135] CMS Collaboration. Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017. URL https://cds.cern.ch/record/2256875.

[136] CMS Collaboration. Top Tagging with New Approaches. Technical Report CMS-PAS-JME-15-002, CERN, Geneva, 2016.

[137] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014. doi: 10.1007/JHEP05(2014)146. arXiv:1402.2657 ,doi:10.1007/JHEP05(2014)146.

[138] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013. arXiv:1307.0007 ,doi:10.1007/JHEP09(2013)029.

[139] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev.*, D81:094023, 2010. arXiv:0912.0033 ,doi:10.1103/PhysRevD.81.094023.

[140] Jesse Thaler and Ken Van Tilburg. Maximizing Boosted Top Identification by Minimizing N-subjettiness. *JHEP*, 02:093, 2012. arXiv:1108.2701 ,doi:10.1007/JHEP02(2012)093.

[141] CMS collaboration. W/Z-tagging of Jets, 2018. URL https://twiki.cern.ch/twiki/bin/view/CMS/JetWtagging#2016_scale_factors_and_correctio.

[142] CMS Collaboration. Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B781:244–269, 2018. arXiv:1710.04960 ,doi:10.1016/j.physletb.2018.03.084.

[143] CMS collaboration. Boosted Top Jet Tagging, 2018. URL https://twiki.cern.ch/twiki/bin/view/CMS/JetTopTagging.

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

[145] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, 2016. arXiv:1610.07922 ,doi:10.23731/CYRM-2017-002.

[146] Jochen Ott. *Search for Resonant Top Quark Pair Production in the Muon+Jets Channel with the CMS Detector*. PhD thesis, Hamburg U., 2012. URL http://ekp-invenio.physik.uni-karlsruhe.de/record/48211.

[147] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017. URL https://cds.cern.ch/record/2257069.

[148] CMS Collaboration. Measurement of differential cross sections for Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78(11):965, 2018. arXiv:1804.05252 ,doi:10.1140/epjc/s10052-018-6373-0.

[149] 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. *Phys. Rev.*, D95(9):092001, 2017. arXiv:1610.04191 ,doi:10.1103/PhysRevD.95.092001.

[150] CMS Collaboration. Measurements of the pp $\rightarrow$ ZZ production cross section and the Z $\rightarrow 4\ell$ branching fraction, and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78:165, 2018. arXiv:1709.08601 ,doi:10.1140/epjc/s10052-018-5567-9.

[151] CMS Collaboration. Jet energy scale and resolution performances with 13TeV data, Jun 2016. URL https://cds.cern.ch/record/2160347.

[152] Roger J. Barlow and Christine Beeston. Fitting using finite Monte Carlo samples. *Comput. Phys. Commun.*, 77:219–228, 1993. doi:10.1016/0010-4655(93)90005-W.

[153] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, pages 115–120, 2011. arXiv:1103.0354.

[154] Jochen Ott. Theta—A framework for template-based modeling and inference, 2010. URL http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto.

[155] Particle Data Group. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi:10.1088/1674-1137/40/10/100001.