A search for pair produced vector-like T quarks in dilepton and multi-jet final states at $\sqrt{s}$ = 13 TeV

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Irene Dutta
Reg. No. : 20121076

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune-411008, INDIA.

May, 2017

Supervisor: Dr. Seema Sharma

Certificate

This is to certify that this dissertation entitled **A search for pair produced vector-like T quarks in dilepton and multi-jet final states at $\sqrt{s}$ = 13 TeV** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Irene Dutta at Indian Institute of Science Education and Research under the supervision of Dr. Seema Sharma, Assistant Professor, Department of Physics, during the academic year 2016-2017.

Irene Dutta 27/4/2017

Dr. Seema Sharma 27/04/2017

Committee:

Dr. Seema Sharma

Dr. Devdatta Majumder

This thesis is dedicated to my parents.

Declaration

I hereby declare that the matter embodied in the report entitled **A search for pair produced vector-like T quarks in dilepton and multi-jet final states at $\sqrt{s}$ = 13 TeV** , are the results of the work carried out by me at the Department of Physics, IISER Pune, under the supervision of Dr. Seema Sharma and the same has not been submitted elsewhere for any other degree.

27/04/2017

Dr. Seema Sharma

Pune 27/4/2017

Irene Dutta

Acknowledgements

I would like to thank my supervisor, Prof. Seema Sharma, for her support and encouragement throughout the duration of the project. She helped me to develop the right attitude for doing research. I would also like to thank my TAC member, Dr. Devdatta Majumder, for providing me with insightful comments on my thesis work.

I'm grateful to Prof. Sourabh Dube, for all the intellectually stimulating discussions on various particles physics topics. Apart from this, I would like to thank Varun Srivastava, Prachi Atmasiddha and Shubham Pandey, for all the thesis related discussions and also for their support and friendship. I want to also thank the other members of our EHEP group who are Shubhanshu Chauhan, Anshul Kapoor, Vinay Hegde, Aditee Rane and Kunal Kothekar, for all the help and guidance they provided me at various stages of the thesis. I'm also grateful to the IT department at IISER Pune, for their dedicated efforts towards the installation and maintenance of the IISER Pune CMS cluster and also for promptly resolving our internet issues. I would like to extend a special thanks to Mr. Brij Jashal, TIFR T2 Grid System Admin, who has resolved our countless complaints and issues relating to grid computing. I also want to acknowledge the DST-INSPIRE Scholarship for Higher Education (SHE) for providing me with a fellowship throughout my undergraduate study.

And last but not the least, I want to thank my parents for their unconditional love and support.

Abstract

The Standard Model fermions are known to be chiral in nature. Several new theories like composite Higgs models, Randall-Sundrum model, GUTs etc. however predict the existence of new heavy vector-like fermions. In this thesis, we explore the possibility of finding heavy vector-like partners of the top quark, namely the Tprime (T) quark, in pp collisions at the LHC at a centre of mass energy of 13 TeV. We have assumed that the T decays to a top quark and a Z boson with 100 % branching ratio (T$\rightarrow$tZ). We look for pair produced T quarks decaying to a dilepton and a multi-jet final state.

The T quark, if found in nature, is expected to be very massive. Its decay products will, therefore, have a considerable boost. We have used the properties of boosted decay topologies in designing a χ^2 algorithm that chooses the best possible candidates for the decay products of the T quark in an event. Using candidates selected by the χ^2 algorithm, we reconstruct an invariant mass for the T quark.We set expected and observed 95% CL upper limits on the TT production cross-section using the Asymptotic CLs method. We were able to set an expected 95 % CL lower limit of 1170 GeV on the T quark mass. We also used the χ^2 algorithm on the data (integrated luminosity of 35.9 fb^{-1}) and were able to exclude T quark masses below 1095 GeV with 95% CL.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Standard Model (SM) of particle physics is a theory of elementary particles and the laws that govern the interactions among these particles. In 1967, Abdus Salam and Steven Weinberg had incorporated the Higgs boson into the electroweak theory. In 1974, the discovery of the J/ψ meson (along with other experiments) confirmed that hadrons are composed of quarks, thus incorporating the strong force into the SM. The model had correctly predicted the existence of $W^{\pm}$ and Z bosons along with the top quark [1–4]. The discovery of the Higgs boson in 2012 at the LHC (CERN) further vindicated the model [5].
Fig. 1.1 shows the known three generations of matter for leptons and quarks, along with the gauge bosons and one Higgs boson. The leptons and quarks make up matter, whereas the gauge bosons mediate the interactions among them. All massive fundamental particles in the Standard Model get their masses through the Brout-Englert-Higgs (BEH) mechanism [6, 7].

1.1 Inadequacies of the Standard Model

Although the SM very accurately describes phenomena within its domain, it is still not the complete picture. It does not address several issues, some of which are mentioned below:

- *Dark matter:* Dark matter makes up for about 24 % of our universe [9–11]. It interacts only gravitationally and perhaps weakly with other known particles. The SM does not have any candidates to explain dark matter.

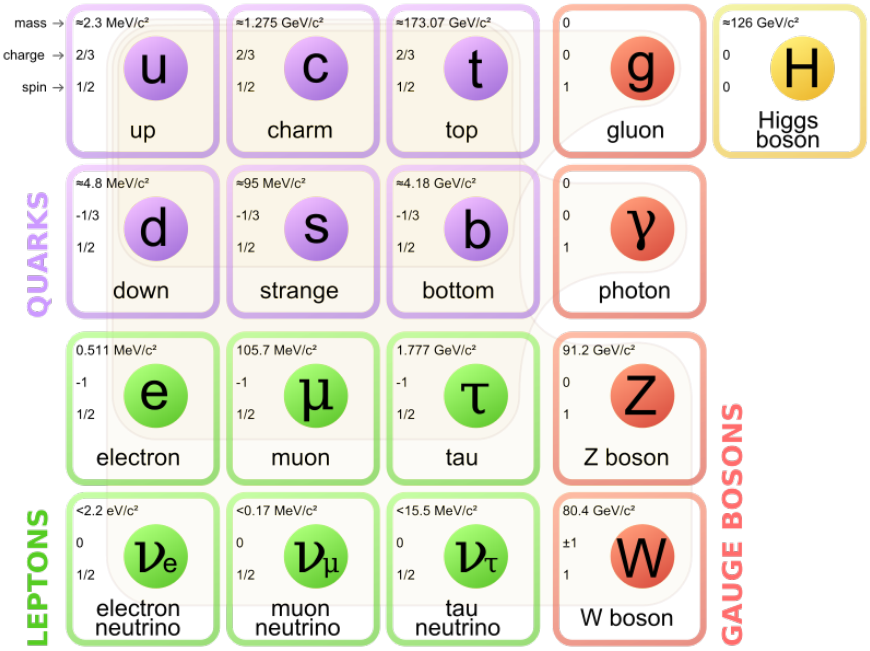


Figure 1.1: The Standard Model particles. [8]

- *Matter-antimatter asymmetry:* The universe is mostly made up of matter. There is a small CP violation seen in the SM but it is insufficient to account for the huge disproportion between matter and anti-matter that we see today.

- *Hierarchy problem:* The theoretical quantum loop corrections to the Higgs mass are divergent in nature and can force the Higgs mass to be very large, even to the order of the Planck scale (10^{18} GeV). However, we know that the CMS and ATLAS experiments at the Large Hadron Collider (LHC), CERN discovered the Higgs mass to be around 125 GeV [5]. This suggests that an unnatural cancellation of the divergent terms gives a Higgs mass close to 125 GeV. This *fine tuning* is somewhat unacceptable theoretically and there is no a priori reason explained by the SM for why things cancel so precisely.

- *Neutrino masses:* The SM considers neutrinos to be massless. However, the 2015 Nobel Prize in Physics was awarded for showing that neutrinos oscillate between different flavours which in turn enforces that they have mass [12].

- *Mass Hierarchy Problem:* The masses of the SM particles in different generations is very different. For example, the electron has a mass of about half a MeV and the tau lepton has a mass of about 1.7 GeV. It is yet unclear as to why their masses are so disparate.

- *Gravitational interaction:* Gravity, which is one of the four fundamental forces, is not incorporated into the SM. The theory also fails to explain why gravity is so much weaker compared to other forces in nature.

All these problems indicate that the SM, although so precisely validated by many collider and non-collider experiments, is still incomplete and this calls for the presence of new physics.

1.2 Beyond the Standard Model

Many theories have been formulated to extend the SM to address some of the issues discussed above. Some of these models are supersymmetry (SUSY) [13], see-saw model for right-handed neutrinos [14], models that predict vector-like quarks and extra dimensions [15]. Noticeably, all these extensions of the SM predict new particles. The SUSY model predicts the existence of a supersymmetric partner for every SM particle differing in spin by $\frac{1}{2}$ units. This implies that for every fermion, there is a supersymmetric boson partner and vice versa. The SUSY model provides a dark matter candidate and also predicts the unification of the electroweak and strong interactions. The see-saw model introduces sterile heavy right-handed neutrinos to explain the very tiny left-handed neutrino masses that appear in the SM. Extra dimensions predict the existence of several new dimensions apart from the regular four space-time dimensions. These can help explain why gravity is so much weaker than the other fundamental forces of nature. According to these theories, gravitons (force carriers of gravity) can leak into these extra dimensions, while the other SM particles can not do so. The Vector-like Quark (VLQ) models predict the existence of massive non-chiral quarks. These quarks might get produced in the proton-proton (pp) collisions at 13 TeV at the LHC. The VLQs are the topic of this study and have been discussed in detail in the following section.

1.3 Vector-like quarks (VLQ)

The SM has only chiral fermions i.e. the left handed and right handed fermionic fields transform differently under the SU(2) gauge transformations. As a result, the Dirac mass term $m\bar{\psi}\psi$ is forbidden in this theory to preserve gauge invariance of the SU(2) Lagrangian. Thus SM charged fermions gain their masses from the BEH mechanism. In particular, the W boson only interacts

with left-handed fermions in the SM.

Several theoretical models, however, predict the existence of new heavy quarks which are non-chiral in nature. Such models include composite Higgs [17], Randall-Sundrum model [18], Grand Unified Theories (GUTs) [19] and little-Higgs models [20–22]. Unlike SM quarks, these quarks have vector couplings to the charged currents. These quarks are heavier partners of the top and bottom quarks, namely the T quark with a charge of $\frac{2}{3}$ e and the B quark with a charge of - $\frac{1}{3}$ e respectively. Some models predict that the T quarks can solve the hierarchy problem because they contribute to cancelling the divergent quantum loop corrections introduced by the top quark to the Higgs mass. These new quarks interact with the SM particles and can decay to a third generation quark and a heavy boson (for e.g. T$\rightarrow$ tZ,tH,bW and B $\rightarrow$ bZ,bH,tW). For such non-chiral fermions, the Dirac mass term m$\bar{\psi}\psi$ is not forbidden. The only way to determine the masses of these particles is to find them experimentally.

Previous searches for these VLQs have been conducted by both ATLAS [23–26] and CMS [27–30] at $\sqrt{s}$ = 7, 8 and 13 TeV. This thesis explores a particular decay topology of the T quark, which has been described in the next section.

1.4 Signal Topology

We will assume T $\rightarrow$ tZ with 100 % branching ratio. Previous searches at the LHC have inferred that for this channel, the mass of the T quarks is larger than 790 GeV [31]. We are going to look for decays of the T like the one shown in Fig. 1.2. One of the T quarks decays to a top quark and a Z boson, where $t \rightarrow W(q\bar{q}')b$ (hadronic top decay) and Z$\rightarrow q\bar{q}$ (hadronic Z decay). The other T decays to a hadronic top and a Z boson which decays to a dilepton pair (leptonic channel). This decay topology gives rise to a dilepton and a multi-jet final state. Since the T quark is massive, we expect its decay products to have a considerable boost and we will exploit this feature later to develop a search strategy for pair produced T quarks.

Branching ratios for various SM particles are summarized in Table 1.1. We will calculate the probability of our final state (Fig. 1.2) with the numbers given in Table 1.1. The probability of both the tops decaying hadronically is given by 0.6741*0.6741=0.4544. As one of the Z decays hadronically, we get 0.4544*0.6991= 0.3177. The other Z can decay to a e^+e^-pair or $\mu^+\mu^-$pair, which implies that the final probability is 0.3177*(0.03363+0.03366)= 0.0214. Finally, we multiply a factor of 2 to this number as there are two Z's. Thus, our final state has a probability of 4.28% .

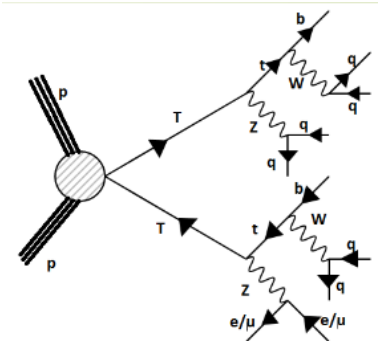


Figure 1.2: Pair produced T quarks decaying to a dilepton and multi-jet final state (quarks manifest themselves as jets in the detector).

Decay type	**BR**
$T \to tZ$	1
$Z \to q\bar{q}$	0.6991
$Z \to e^+e^-$	0.03363
$Z \to \mu^+\mu^-$	0.03366
$W \to q\bar{q}'$	0.6741
$t \to Wb$	1
Total probability of final state	0.0428

Table 1.1: Branching ratios for some particles used in this study [16] and the total probability of the final state as shown in Fig. 1.2.

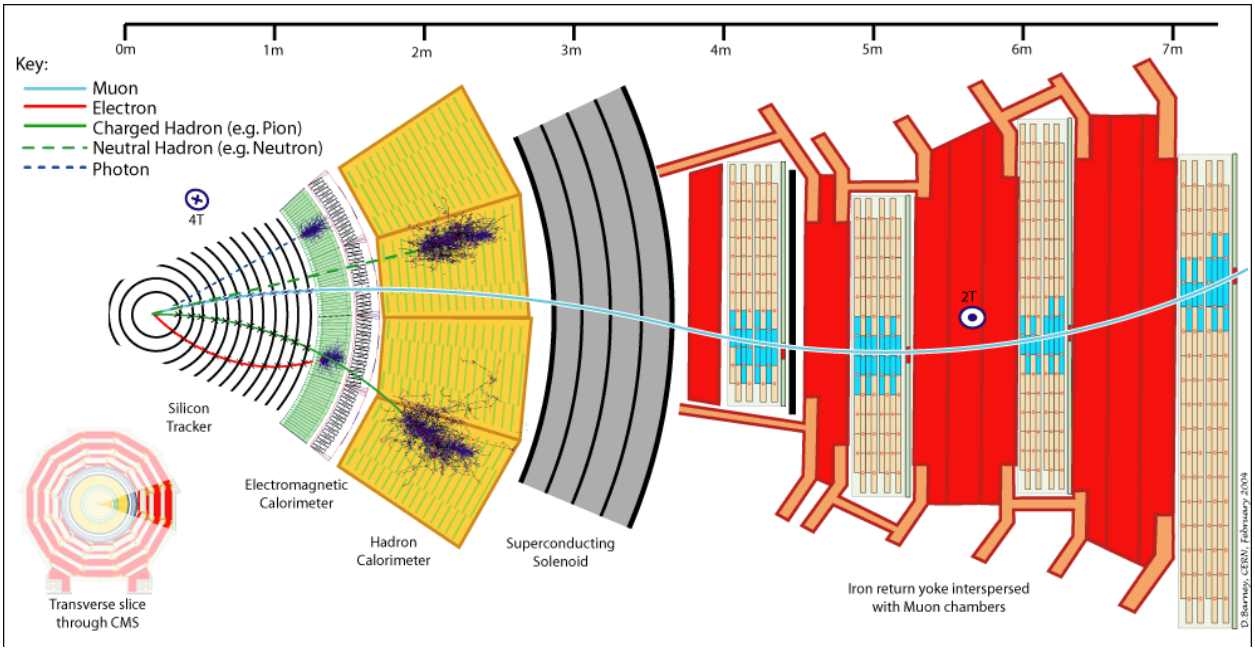


Figure 1.3: A transverse slice of the CMS detector [32].

1.5 The CMS detector

The Compact Muon Solenoid (CMS) is a particle detector built to study pp collisions at the LHC. The detector has a diameter of 15 m and a length of 21.6 m. The key feature of the detector is its large superconducting solenoid with an internal diameter of 6m that creates a magnetic field of 3.8 T. The magnetic field helps to bend the trajectories of charged particles in order to identify their momentum and charge. Fig 1.3 shows a transverse view of the detector.

The detector is built hermetically around the beam pipe (where pp collisions take place) and has several different modules specialized in identifying certain types of particles. Closest to the beam pipe are the silicon pixel and strip detectors, which help in identifying charged particle tracks. Surrounding this is the Electromagnetic Calorimeter (ECAL), which is made up of scintillating lead-tungstate crystals and helps to identify energy deposits from photons and electrons. The Hadron Calorimeter (HCAL) is a sampling calorimeter built from brass and a scintillator and helps to identify energy deposits from charged and neutral hadrons. The pixel and strip trackers, ECAL and the HCAL are enclosed within the solenoid (magnetic) volume. Outside the magnet are the muon chambers which are gas ionization chambers and are used to identify muons. These muon chambers are embedded in steel return yokes of the magnet. Muons can travel several metres without interacting. This is the reason why muon detectors are placed at the outer edge of the

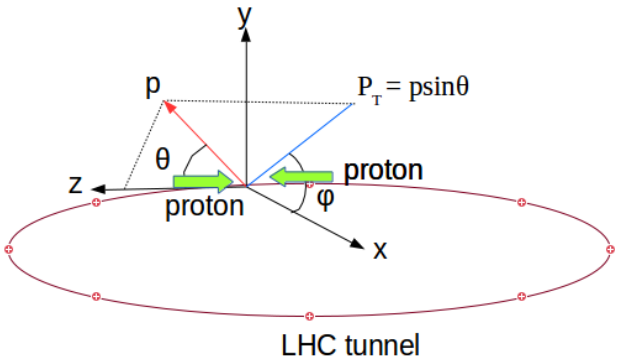


Figure 1.4: Right handed coordinate system used by the CMS.

hermetic detector. A detailed description of the CMS detector can be found in [33].

For specifying the directions of the outgoing particles, a right-handed coordinate system is used. The z-axis is along the anti-clockwise direction of the beam, the y-axis points upwards and perpendicular to the plane of the LHC ring and the x-axis points to the centre of the LHC ring. The x-y plane is also known as the transverse plane. The polar angle θ is measured from the positive z-axis and the azimuthal angle ϕ is measured in the plane transverse to the beam axis (Fig. 1.4). The pseudorapidity η is defined as -ln[tan($\frac{\theta}{2}$)]. This definition for η holds true only when the mass of the particle is negligible compared to its momentum. A non-zero total momentum after the collision in the transverse plane, also regarded as "missing energy", can serve as a useful indicator for neutral particles like neutrinos which otherwise escape unnoticed from the detector. Thus, the momentum of outgoing particles is measured in the transverse plane and is defined as p_T = psinθ, where p is the momentum in the 3D plane.

Chapter 2

Simulated samples, Collision datasets and Event reconstruction

2.1 Simulated samples and Collision datasets

Monte Carlo simulations are used to model pp collisions at the LHC. Various processes are generated based on their cross sections, kinematics and the dynamics of interaction. MC samples for both signal and background events were produced by the CMS generator team. The signal samples are generated by assuming equal branching ratios for $T \to tZ$, $T \to tH$ and $T \to Wb$. However, we select only $TT \to tZtZ$ events using the MC truth information. For signal samples, the leading order event generator MADGRAPH 5.1.3.30 [34] was used. The NNPDF2.3LO parton distribution functions (PDFs) [35] were used and samples were generated with up to two additional hard partons. The generated events were passed through a GEANT4 [45] simulation of the CMS detector to account for the detector response. Signal samples were generated by assuming T masses of 800, 900, 1000, 1100, 1200, 1300, 1400 and 1500 GeV. Cross-sections were calculated using next-to-next-to-leading order (NNLO) and next-to-next-to-leading log (NNLL) soft gluon resummation based on the Top++2.0 program [36] using 'MSTW2008nnlo68cl' PDF and using LHAPDF 5.9.0. Cross sections at 13 TeV for various T mass points are listed in Table 2.1.

The major SM process that contributes to our final state is Drell-Yan production of Z+jets along with smaller contributions from $t\bar{t}$ and diboson production. We will refer to these processes as

Sample	σ(pb)
TprimeTprime_M-800	0.196
TprimeTprime_M-900	0.0903
TprimeTprime_M-1000	0.044
TprimeTprime_M-1100	0.0224
TprimeTprime_M-1200	0.0118
TprimeTprime_M-1300	0.00639
TprimeTprime_M-1400	0.00354
TprimeTprime_M-1500	0.00200

Table 2.1: Production cross sections for pair produced Tquarks.

Sample	σ (pb)
$t\bar{t}$	831.76
DYJetsToLL_M-50_PT-100to250	83.12
DYJetsToLL_M-50_PT-250to400	3.047
DYJetsToLL_M-50_PT-400to650	0.3921
DYJetsToLL_M-50_PT-650toInf	0.03636
WW	118.7
WZ	46.74
ZZ	16.91

Table 2.2: Cross sections for background samples.

background. Statistics is usually low on the higher $\hat{p_T}$ side and hence, the Z+jets sample is generated in different $\hat{p_T}$ bins. These different samples are then stitched together with the proper cross sections to give the complete $\hat{p_T}$ profile of the underlying physics process. The Z+jets sample was produced using the aMC@NLO [37] generator. The cross-sections at NLO for Z+jets sample are listed in Table 2.2. We have also applied p_T dependent electroweak correction factors for the Z+jets sample later in the analysis [38].

The $t\bar{t}$ sample is generated by POWHEG generator [39] and the cross-section is considered at NNLO. The diboson samples were generated using PYTHIA 8.1 [40] generator. The cross-sections are considered at NNLO for WW [41] and ZZ [42] samples whereas for WZ [43] samples, the cross-section is considered at NLO. Production cross sections for the background samples are listed in Table 2.2.

All generated events were showered and hadronized using PYTHIA 8.1 and tune CUETP8M1 [44], except for the $t\bar{t}$ sample where tune CUETP8M2T4 was used. A GEANT4 [45] based simulation of

the CMS detector was used to simulate the response of the detector to particles traversing through it. Finally, reconstruction of the physics objects is done using the same software configuration which is used for the collision data. In a high luminosity machine such as the LHC, multiple proton-proton interactions happen per bunch crossing instead of just one. These extra events are known as pile-up events. To take this factor into account, on an average, at least 20 pile-up interactions are overlaid per hard scattering process in the Monte Carlo.

To make sure that the pile-up distribution in the data matches the corresponding distribution in MC, a correction is applied to the MC by matching the "true" number of interactions in the MC to the "true" number of interactions in the data. This procedure is known as pile-up re-weighting.

The collision datasets used were collected with the requirement of containing at least one muon or one electron, known as the 'Single Muon' and 'Single Electron' dataset respectively. These were taken in LHC Run-II in 2016 and have a total integrated luminosity of 35.9 fb^{-1} .

2.2 Event reconstruction

2.2.1 Trigger

At very high luminosities such as that at the LHC, it is impossible to register every event to the tape. A High Level Trigger (HLT) is designed so that one selects potentially interesting events (like events having a muon/electron or events with large missing energy (MET)). A fast reconstruction is done for physics objects at the HLT which is not as rigorous as the offline reconstruction. So an event that passes the online trigger requirement for some object reconstruction might fail to do so in the offline selection.

For this study, we have used separate triggers for the muon channel (Z$\rightarrow \mu^+\mu^-$) and the electron channel (Z$\rightarrow e^+e^-$). For the electron channel, we use a trigger which selects events having a single electron with p_T > 115 GeV (HLT_Ele115_CaloIdVT_GsfTrkIdT). For the muon channel, we use a trigger that selects events containing a single isolated muon with p_T > 24 GeV (HLT_IsoMu24 or HLT_IsoTkMu24). To account for the differences in the performance of the trigger in the data and the MC, we apply trigger scale factors for both the muon channel and the electron channel. A 2D histogram of the muon trigger scale factors (provided by the CMS Muon Object Group) as a function of the p_T and the η is plotted in Fig. 2.1.

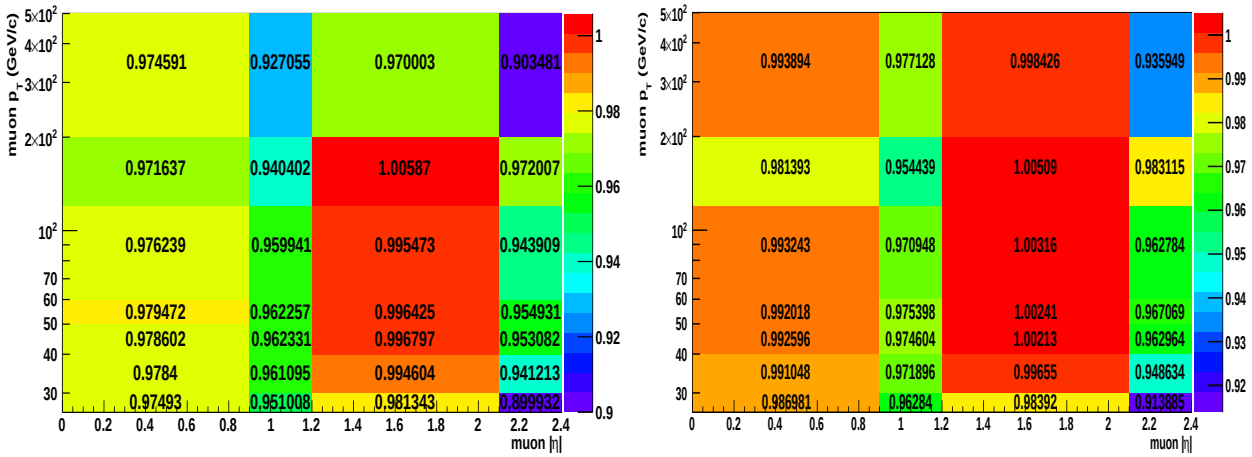


Figure 2.1: (Left) Muon trigger scale factors as a function of the p_T and the η for different periods within LHC Run-II in 2016.

The CMS uses a standard Particle Flow (PF) algorithm [46–48] to reconstruct objects. The algorithm combines information from the trackers, calorimeters and the muon systems to reconstruct candidates for electrons, muons, photons, charged hadrons and neutral hadrons.

2.2.2 Electron

The PF algorithm uses an iterative tracking method to reconstruct the best tracks from hits in the pixel detectors. These tracks are then matched to energy deposits in the ECAL cluster. Electron trajectories are then reconstructed by fitting a Gaussian Sum Filter (GSF) algorithm [49] to these tracks and care is taken to account for the radiative energy losses.

We use electron candidates with $|\eta| < 2.4$ and which pass the cut based electron identification and isolation as recommended by the CMS Electron Object group. These criteria are optimized for selecting electrons originating from W or Z boson decays. Relative isolation I_{rel} is defined as the sum p_T of charged, neutral hadrons and photons in a cone of Δ R = $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the lepton, divided by the lepton p_T. $\Delta\beta$ is defined as 1.5 times the sum p_T of charged hadrons from pile-up vertices. The factor 1.5 accounts for the fact that the ratio of the energy of the neutral hadrons to charged hadrons is 1:2 in QCD processes. Relative Isolation I_{rel} with $\Delta\beta$ pile-up corrections is required to be < 0.1 in a cone of Δ R < 0.3. This means that around the electron candidate if one considers a cone of size Δ R < 0.3, then no more than 10 % of the energy in

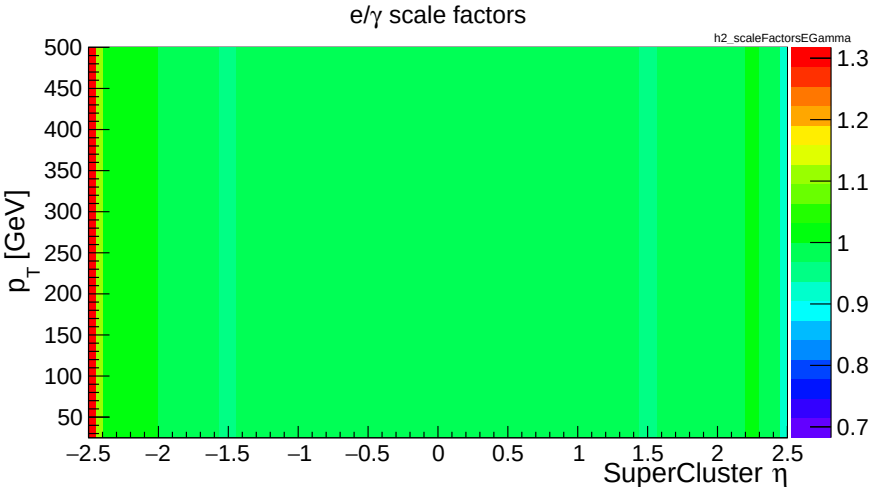


Figure 2.2: Electron identification scale factor as a function of the p_T and the η.

the cone should come from other charged/neutral hadrons, photons or pile-up. The identification requirements help in rejecting jets mis-identified as electrons or electrons from photon conversions. These requirements can have different performance efficiencies in the data and MC. To correct for these differences, we apply scale factors provided by the CMS Electron Object Group in the form of data/MC ratios to the MC that are dependent on both the p_T and the η of the electron. This ensures that these requirements perform exactly the same in the data and the MC. A 2D histogram of the scale factors as a function of the p_T and the η is plotted in Fig. 2.2

2.2.3 Muon

Muon candidates are reconstructed by the PF algorithm from the information in the tracker and the information from the muon systems. We use muon candidates with $|\eta| < 2.4$ and which pass the cut based muon identification as recommended by the CMS Muon Object group. Muon identification requirements suppresses cosmic muons and also rejects hadrons which accidentally reach the muon chambers and get mis-identified as muons. Relative isolation I_{rel} is defined as the sum p_T of charged, neutral hadrons and photons in a cone of Δ R = $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the lepton, divided by the lepton p_T. $\Delta\beta$ is defined as 0.5 times the sum p_T of the charged hadrons from pile-up vertices. The factor 0.5 comes from the ratio of the energy of the charged hadrons to neutral hadrons. Relative Isolation I_{rel} with $\Delta\beta$ pile-up corrections is required to be < 0.15 in a cone of Δ R < 0.4. This means that around the muon candidate if one considers a cone of size Δ R < 0.4, then no more than 15 % of the energy in the cone should come from other charged/neutral hadrons,

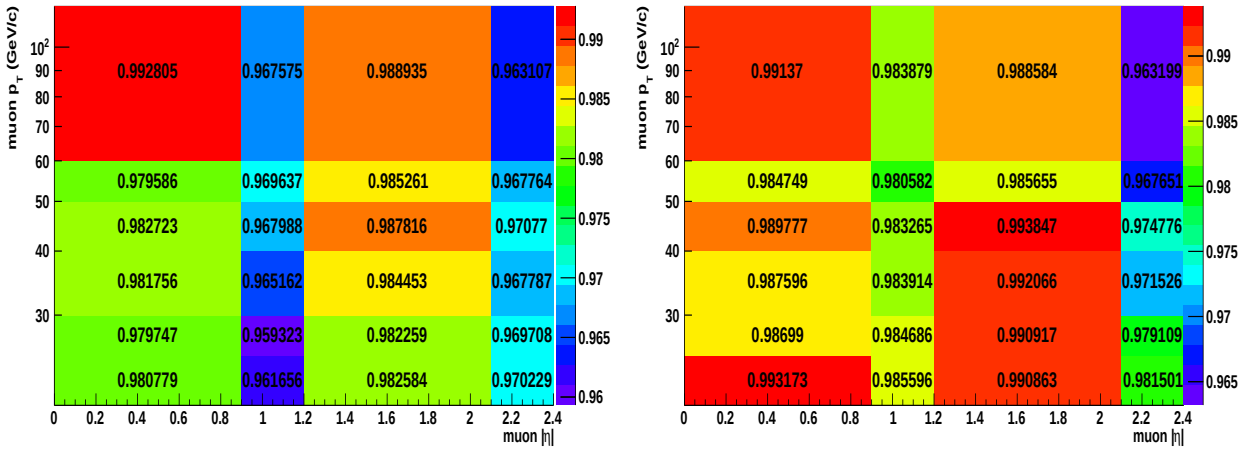


Figure 2.3: (Left) Muon identification scale factors as a function of the p_T and the η for different periods within LHC Run-II in 2016.

photons or pile-up.

Again, these requirements on the identification and isolation can have performance efficiencies which differ in the data and the MC. To account for these differences, scale factors that are dependent on both the p_T and the η of the muon are applied to the MC. These scale factors ensure that the identification and isolation requirements perform exactly the same in the data and the MC. A 2D histogram of the muon identification scale factors as a function of the p_T and the η is plotted in Fig. 2.3.

2.2.4 Jets

Quarks and gluons have color charge and can not exist freely in nature. They hadronize into a series of color-neutral hadrons which appear as a collimated shower of particles in the detector (known as a jet). Particles reconstructed by the PF algorithm are clustered into jets using the anti-kT algorithm [50] with a distance parameter of Δ R = 0.4 (AK4 jet) and Δ R = 0.8 (AK8 jet). Charged hadrons coming from pile-up vertices are removed from the cluster. We use AK4 jets with $p_T > 30$ GeV and $|\eta| < 2.4$. We use AK8 jets with $p_T > 200$ GeV and $|\eta| < 2.4$.

The detector has been calibrated at particular energies and the detector response is non-linear as a function of the energy and the η. The energy measured from the detector is not necessarily the true energy of the original particle. Jet energy corrections (JECs) provided by the CMS Jet-

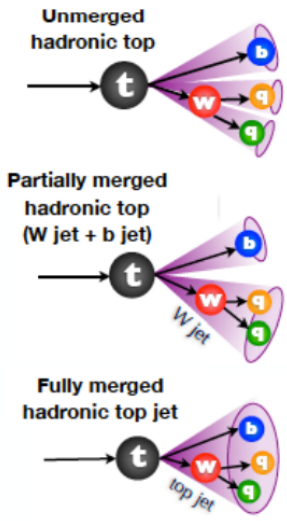


Figure 2.4: Top decay [51]

MET Group are applied to the jets to correct for the non-linear response and non-uniformity of the detector and also for pile-up corrections. These corrections are applied as a function of p_T and η and are obtained from collision datasets and GEANT4 based CMS MC simulations.

2.2.5 Boosted decays and substructure identifiers

For a top quark with p_T = 50 GeV decaying to a W boson and a b quark (where the W decays hadronically), we expect to see three isolated spots of energy in the calorimeter. For another top quark having a larger boost (p_T = 400 GeV), the three different jets originating from the top decay will merge into a single jet (also called as a fat jet) and we will see one large lump of energy in the detector (Fig. 2.4). In the latter situation, it becomes difficult to reconstruct individual jets and one needs special jet substructure identifiers.

Some of these substructure variables are:

1. **N-subjettiness [52]** : This is defined as

$$\tau_N = \frac{\sum_k p_{T,k} min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}\}}{d_0} \tag{2.1}$$

where k runs over all constituents of the jet, $\Delta R_{J,k}$ is the distance in $\eta - \phi$ plane between

candidate subjet J and particle k. The normalization factor d_0 is taken as

$$d_0 = \sum_k p_{T,k} R_0 \tag{2.2}$$

where R_0 is the jet radius. For $\tau_N \approx 0$, all the radiation is along the candidate subjet, and the jet has N (or fewer) subjets. For $\tau_N \gg 0$, a large fraction of the jet energy is scattered away from the candidate subjet, and the jet has at least N+1 subjets. One expects that for W or Z jets, τ_2 will be smaller as compared to the QCD background. However, QCD jets can also have small values of τ_2. Thus, to increase the discriminating power, $\frac{\tau_2}{\tau_1}$ (or τ_{21}) is used as an identifier for W jets. Lesser the value of τ_{21}, more likely it is that the jet is a W or Z jet. Similarly, although one expects τ_3 to be smaller for top jets, QCD jets can also end up having small values of τ_3. Thus, $\frac{\tau_3}{\tau_2}$ (or τ_{32}) is a better discriminator for top jets than τ_3 alone [52].

2. **Pruned mass** [53]: The jet clustering is re-run for each jet. When clusters i and j are being merged into a single cluster p, the conditions $min(p_{T_i}, p_{T_j})/p_{T_p} > 0.1$ and $\Delta R_{ij} < 0.5 * \frac{m_J}{p_{T_J}}$ are imposed, where m_J stands for the jet mass in the first clustering sequence. If the step i+j $\rightarrow p$ fails the above two conditions, the two clusters are not merged, and the softer cluster is thrown away. This ensures removal of soft particles and wide angled particles. It is seen that the jet mass, which is calculated as $m^2 = E^2 - p^2$, has an improved resolution with the above jet pruning technique.

3. **Soft Drop mass** [54, 55] : The last step of the jet clustering algorithm is undone to break a jet j in j_1 and j_2. j is called a soft drop jet if it passes the following condition :

$$\frac{min(p_{T_1}, p_{T_2})}{p_{T_1} + p_{T_2}} > z_{cut}(\frac{\Delta R_{12}}{R_0})^{\beta} \tag{2.3}$$

where z_{cut} (= 0.1) is the soft threshold, β (= 0) is an angular exponent, R_{12} is the distance (in $\eta - \phi$ plane) between j_1 and j_2 and R_0 is the jet radius.

If the above condition fails, j is defined to be the subjet with greater p_T and the above procedure is repeated. The iteration is repeated till j can no longer be declustered. At this stage j is either removed or kept as the final soft drop jet. This technique is also observed to improve jet mass resolution.

Based on the above sub-structure identifiers, we will now enlist the conditions for reconstructing

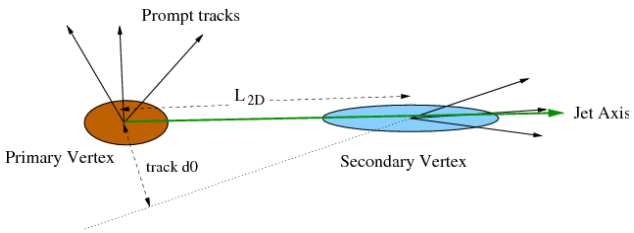


Figure 2.5: Displaced secondary vertex resulting from the decay of a long lived particle. Here, d_0 is the impact parameter and L_{2D} is the decay length of the particle in the transverse plane [56].

fat top jets or fat W or Z jets (Table 2.3). These requirements can perform differently in the data and the MC and to account for it, we apply top tagging and W or Z tagging scale factors to events with a top jet or W/Z jet in the MC.

Variable type	**Selection for Top**	**Selection for W or Z**
Soft Drop Mass	[105 - 220] GeV	-
Pruned Mass	-	[65 - 105] GeV
$\|\eta\|$	< 2.4	< 2.4
τ_{32}	< 0.67	-
τ_{21}	-	< 0.6
p_T	> 400 GeV	> 200 GeV

Table 2.3: Selection criteria for reconstructing boosted top or W or Z jets.

2.2.6 b-jet tagging

Bottom quarks hardonize to B hadrons. The B hadron has a longer lifetime and travels 400 μm before decaying. In a detector, this longer time of flight appears as a displaced secondary vertex (Fig. 2.5). This feature is used to identify jets originating from a b quark. Another useful variable for b-tagging is the mass of the secondary vertex.

The CMS event reconstruction uses a Combined Secondary Vertex Algorithm (CSV) to tag the b jets [57]. The distance between the primary and the secondary vertex (impact parameter) is usually higher for jets coming from B hadrons and this variable is taken into account to build a discriminator using the Multivariate Variable Analysis (MVA). We use the loose working point or CSVL (Value of discriminant > 0.5426). The loose point allows for a greater efficiency but at

the same time increases the possibility of light flavour jet contamination. The efficiency of the CSVL working point is 85%-90%. The MC might not model the data accurately and to account for the differences in the performance of the b-tagging algorithm in the data and the MC, residual corrections in the form of data/MC scale factors are applied to the MC.

Chapter 3

Analysis Strategy

In this chapter, we will describe a technique which can be used to reconstruct the invariant mass of the T quarks. A conventional 'cut and count' analysis looks at properties of AK4 jets, electrons or muons to identify subtle differences in the signal and background shapes for some kinematic distribution and tries to maximize the signal to background ratio. In this analysis, however, we will develop an alternative technique where we use properties of boosted AK8 jets. We will develop a χ^2 algorithm which chooses the best possible candidates for decay products of the T quark and from these candidates, we will reconstruct the invariant mass of the T quark. We want to mention here that for all of our offline analysis, we have used ROOT v5.34 [58], a data analysis framework provided by CERN.

3.1 Pre-selection of Events

According to Fig. 1.2, we expect to see a dilepton pair having an invariant mass within the Z mass window [75-105 GeV] in our signal events. So we select events with either a pair of oppositely charged electrons or muons with an invariant mass within the Z mass window. We have plotted the invariant mass of an oppositely charged lepton pair in Fig. 3.1. For a dimuon pair, we require that the leading muon should have a p_T > 45 GeV and the sub-leading muon should have a p_T > 25 GeV. For a dielectron pair, we require that the leading electron should have a p_T > 120 GeV and the sub-leading electron should have a p_T > 25 GeV.

We remove AK4 jets which fall within Δ R = 0.4 of a muon or an electron, to avoid the energy of

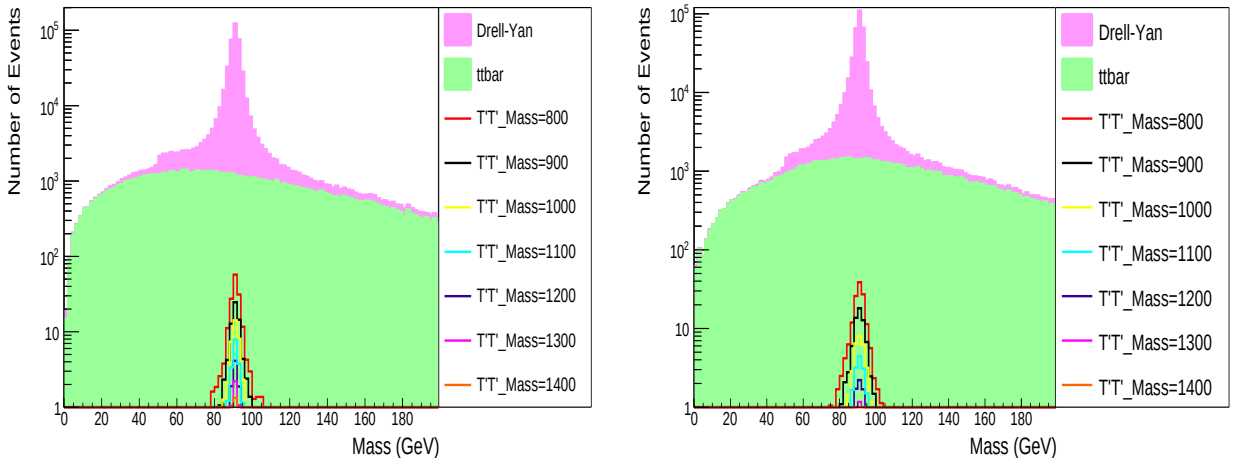


Figure 3.1: (Left) Invariant mass distribution for a dielectron pair ($\mathscr{L}_{int}$= 35.9 fb^{-1}). (Right) Invariant mass distribution for a dimuon pair ($\mathscr{L}_{int}$= 35.9 fb^{-1}).

the lepton from being counted twice (once as the lepton and once inside the jet). We also remove AK8 jets which fall within Δ R = 0.8 of a muon or an electron. With these selections in place, we decided to look at some kinematic distributions like dilepton p_T, H_T (sum p_T of all AK4 jets in an event), S_T (sum of H_T and leading dilepton p_T in an event) and the number of jets (Fig. 3.2).

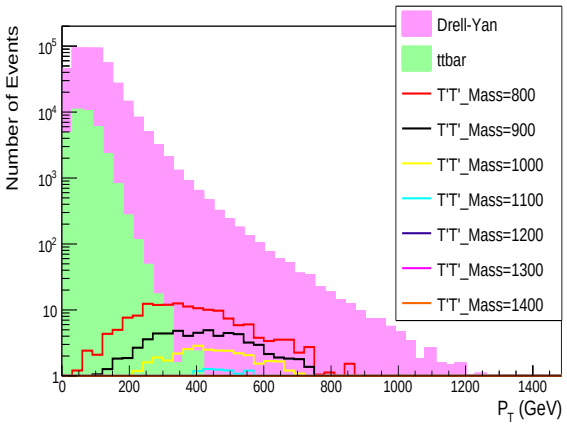
Based on Fig. 3.2, we can impose further conditions which will significantly reduce the background while having almost no effect on the signal numbers. The next few selections are:

- There should be at least 4 AK4 jets in the event
- Z(ll) p_T > 100 GeV
- H_T for AK4 jets > 500 GeV
- 2 loose b jets (CSV >0.5426)

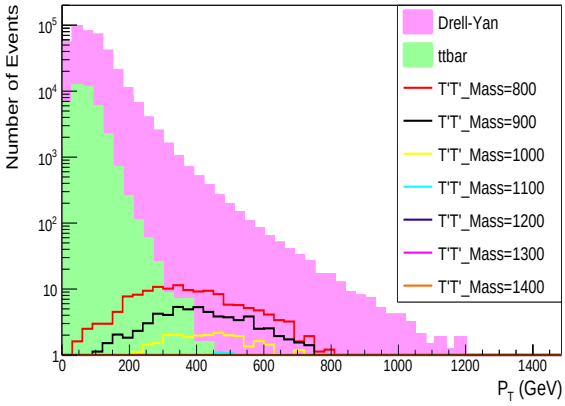
The pre-selections for this study have been summarised in Table 3.1.

3.1.1 Signal Optimization

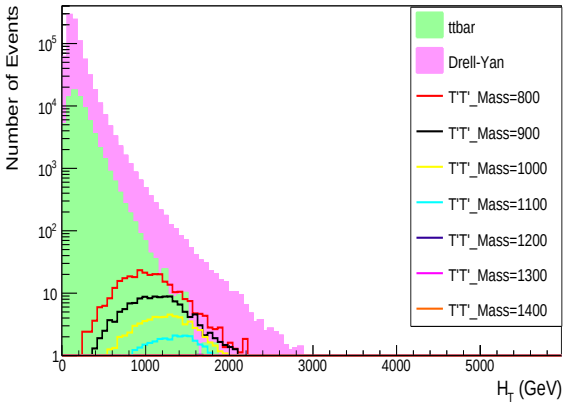
For this particular section, we study event yields based on a slightly different set of pre-selections as mentioned in Table 3.2. An event yields table shows the number of events surviving in the signal and background samples after each selection. Based on these numbers we calculate the



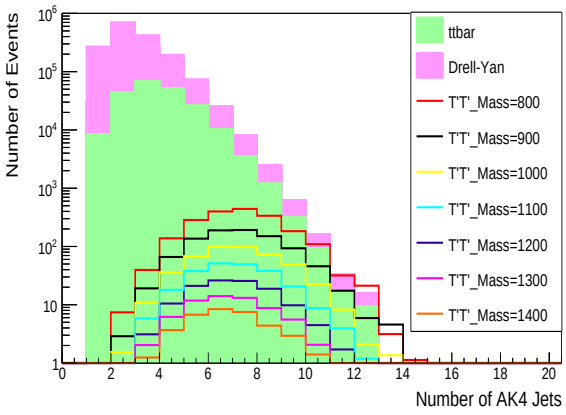
Dielectron p_T.



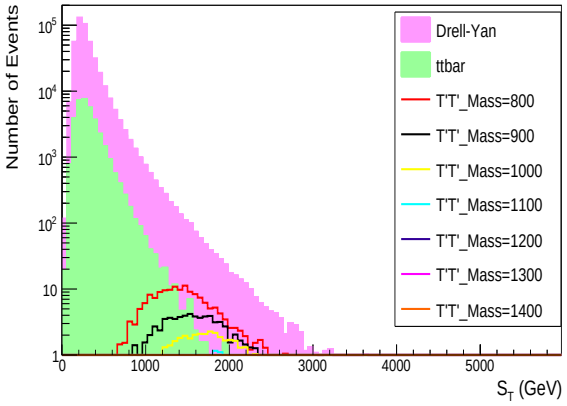
Dimuon p_T.



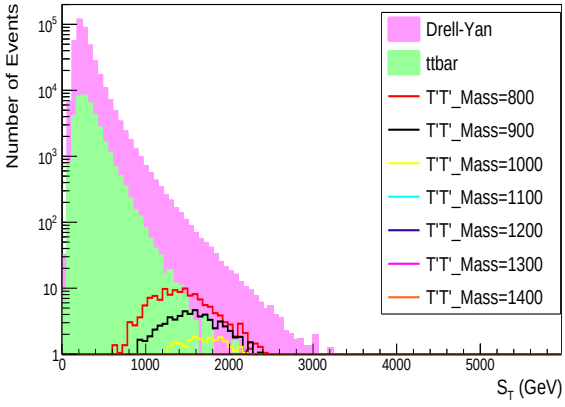
H_T distribution (e^+e^-+ $\mu^+\mu^-$channel).



Number of AK4 jets (e^+e^-+ $\mu^+\mu^-$channel).



S_T distribution (e^+e^-channel)



S_T distribution ($\mu^+\mu^-$channel)

Figure 3.2: Some kinematic distributions after the Z(ll) requirement with $\mathscr{L}_{int}$= 35.9 fb^{-1}. The top row shows the p_T distribution for the Z(ll) candidate in the Z (e^+e^-) channel and the Z ($\mu^+\mu^-$) channel separately. The middle row shows the H_T distribution and the AK4 jet multiplicity. The bottom row shows the S_T distribution for the Z (e^+e^-) channel and the Z ($\mu^+\mu^-$) channel separately.

Variable	Requirement
Leading muon p_T (if Z$\rightarrow \mu^+\mu^-$)	$\geq$45 GeV
Leading electron p_T (if Z$\rightarrow e^+e^-$)	$\geq$120 GeV
Sub-leading lepton (e/μ) p_T	$\geq$25 GeV
Number of e^+e^-or $\mu^+\mu^-$pair in the Z Mass window (75 - 105 GeV)	1
Number of AK4 jets	≥ 4
Z p_T	> 100 GeV
H_T	> 500 GeV
Number of loose b jets	$\geq$2

Table 3.1: Signal region pre-selection.

Variable	Requirement
Leading lepton p_T	$\geq$25 GeV
Sub-leading lepton p_T	$\geq$10 GeV
Number of e^+e^-or $\mu^+\mu^-$pair in the Z Mass window (60 - 120 GeV)	1
Number of AK4 jets	≥ 4
Z p_T	> 100 GeV
H_T (or S_T)	> 500 GeV
Number of loose b jets	$\geq$2

Table 3.2: Pre-selection requirements to study signal optimization. This set of pre-selections are valid only for this particular section. This table is comparable to Table 3.1, except that we use a loosened criteria on the lepton p_T and the dilepton invariant mass.

signal significance or $\frac{S}{\sqrt{S+B}}$, where S is the signal yield and B is the background yield. The event yields for background ($t\bar{t}$ and Z+jets) and T quark of mass 800 GeV is given in Table 3.3.

As can be seen from Table 3.3, the H_T selection gives better signal significance than the S_T selection. To understand this, we selected events with all our pre-selections as mentioned in Table 3.2 except for the H_T (or S_T) selection. Then for this particular pre-selection, we plotted the H_T (or S_T) distribution. Now from this H_T (or S_T) histogram, we made a cumulative distribution where for each value of H_T (or S_T), the number of bin entries where summed up from that bin till infinity (Fig. 3.3). For e.g., for H_T = 250 GeV, we sum up the bin entries starting from the bin corresponding to H_T = 250 GeV until the very last bin. We do not take into account the bin entries corresponding to H_T < 250 GeV. This exercise was done for both signal and background samples. Essentially, such a cumulative distribution tells us that given a particular value of H_T (or S_T), how many events would survive if we decided to select events above that value of H_T (or S_T).

Pre-selection with H_T			
Event selection	Background	M(T) = 800 GeV	$\frac{S}{\sqrt{S+B}}$
All Events	36906152.62	7036.40	1.16
One e^+e^-or $\mu^+\mu^-$pair in the Z Mass window (60 - 120 GeV)	840830.17	316.83	0.35
Number of AK4 jets $\geq$ 4	69260.60	299.45	1.14
Z p_T > 100 GeV	27764.98	290.92	1.74
H_T > 500 GeV	7549.48	279.74	3.16
At least 2 loose b jets	1838.16	222.22	4.90
Pre-selection with S_T			
Event selection	Background	M(T) = 800 GeV	$\frac{S}{\sqrt{S+B}}$
All Events	36906152.62	7036.40	1.16
One e^+e^-or $\mu^+\mu^-$pair in the Z Mass window (60 - 120 GeV)	840830.17	316.83	0.35
Number of AK4 jets $\geq$ 4	69260.60	299.45	1.14
Z p_T > 100 GeV	27764.98	290.92	1.74
S_T > 500 GeV	16756.16	290.60	2.23
At least 2 loose b jets	3615.38	228.97	3.69

Table 3.3: Event yields table for T quark of mass 800 GeV. Here, background refers to $t\bar{t}$ and Z+jets. The numbers have been normalized to $\mathscr{L}_{int}$= 35.9 fb^{-1}.

Afterwards, we plotted $\frac{S}{\sqrt{S+B}}$ using the cumulative signal and background distributions (Fig. 3.3). From Fig. 3.3, it can be seen that an H_T > 500 GeV requirement drastically reduces the background, without affecting the signal much. An S_T > 500 GeV requirement has almost no effect on the signal and background alike, thus giving lower significance values. Ideally, the optimum place to make a selection would be at the peak of the $\frac{S}{\sqrt{S+B}}$ curve. However, we need to retain sufficient signal events, and hence the optimum place will not always be the best place to select events. Usually values of H_T (or S_T) lower than the one corresponding to the peak are chosen for an analysis.

3.2 Mean and widths of reconstructed Top, W and Z

The main idea of the χ^2 algorithm (section 3.3) is to check how close the mass of a candidate object is to the true mass of the particle. Here, by candidate objects, we mean the decay products of the T quark which are the top quark and the Z boson. Hence, we need to first evaluate the width and mean masses of reconstructed standard model particles using the signal MC samples.

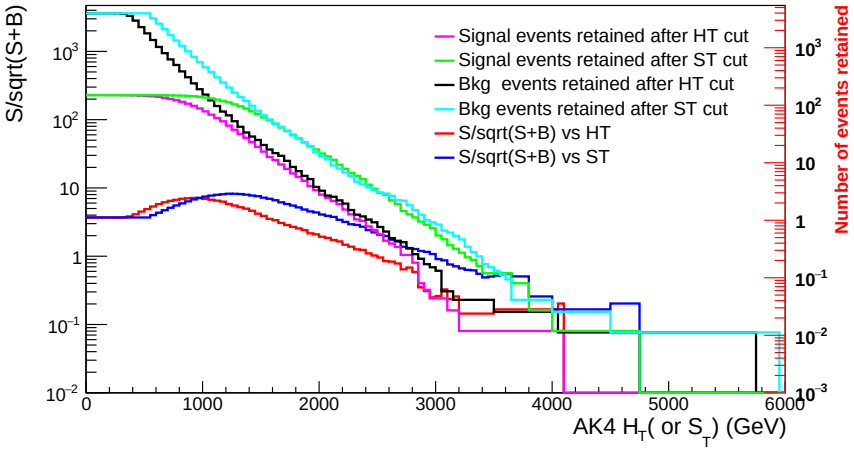


Figure 3.3: Cumulative plot for H_T (S_T) with $\mathscr{L}_{int}$= 35.9 fb^{-1}. Signal here refers to T quark of mass 800 GeV. All pre-selections (Table 3.2) have been applied except the H_T (or S_T) selection. For the cyan, black, green and magenta curves (i.e. number of events retained after imposing a selection on a particular value of H_T (or S_T)), refer to the axis on the right. For the red and blue curves ($\frac{S}{\sqrt{S+B}}$ corresponding to a particular value of H_T (or S_T)), refer to the axis on the left.

3.2.1 Top

For each top (t→W($q\bar{q}'$)b), we check if there are any AK8 jets with p_T > 300 GeV, $|\eta|$ < 2.4 and τ_{32} < 0.67 falling within Δ R < 0.4 from the top. On finding such a jet, we ensure that the W boson and b quark (daughters of the top) are matched to this AK8 jet within Δ R < 0.6. We also make sure that the daughters of the W are also matched to this AK8 jet. If all the above criteria are satisfied, we say that the AK8 jet is a top jet. The mass of the matched AK8 jet is plotted on the left in Fig. 3.4.

If the top is not very boosted, instead of reconstructing just one fat jet, we will end up getting a b jet along with a W tagged AK8 jet. Thus, we use an alternative scheme for cases where the top does not get matched to any AK8 jet. For the W boson (daughter of top), we loop over all AK8 jets tagged as V and look for a jet which falls within Δ R < 0.3 from the W. On finding such a jet, we also ensure that the daughters of the W($q\bar{q}$') fall inside this jet. If all the above criteria are satisfied, then the jet is declared as matched to the W boson.

After this, we look for loose b tagged jets that fall within ΔR < 0.4 from the b quark. The jet with the least ΔR from the b quark is matched to it. We now declare this b-jet and the AK8 jet matched to W as coming from the top. The invariant mass of the pair is plotted on the right in Fig. 3.4.

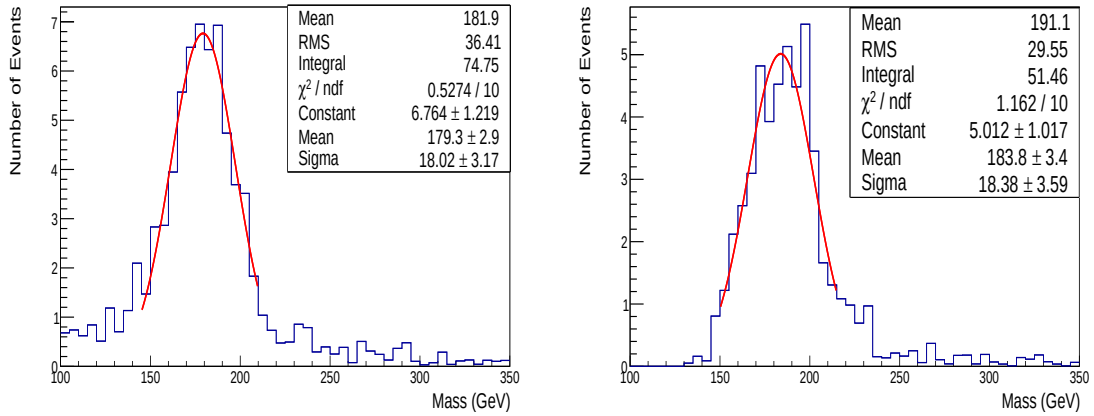


Figure 3.4: (Left) Mass of AK8 jets matched to a top. The histogram has been fitted with a Gaussian. (Right) Invariant Mass of AK8 W jet and b jet matched to a top. The histogram has been fitted with a Gaussian. We use the mean and sigma values given towards the end of the legend which were calculated from the fit. The mean and RMS values at the top correspond to the histogram.

For cases where the top p_T is small, we will reconstruct a tri-jet system corresponding to the three decay products of the top quark. We try to match the daughters of the hadronic top to three different AK4 jets (Δ R < 0.4). Once all the three jets are found, we check if the invariant mass of the tri-jet system falls within the top mass window (110 - 210 GeV). We also ensure that the Δ R between the top and the combined three jet system is less than 0.4. Apart from this, the two jets which get matched to the daughters of the W should also have an invariant mass within the W mass window (65 - 105 GeV) and fall within Δ R < 0.4 from the W. The jet that gets matched to the b quark should be loose b tagged. If all the above said criteria are satisfied, the tri-jet system is declared as a top candidate. The invariant mass of this tri-jet system is plotted in Fig. 3.5.

Please refer to Table 3.4 for a summary of the mean and width of reconstructed top quarks.

3.2.2 V jet originating from W or Z boson

To reconstruct hadronically decaying Z (qq) bosons, we look for jets with p_T >200 GeV/c, $|\eta|$ < 2.4 and τ_{21} < 0.6 and match it to Z bosons falling within ΔR < 0.4. Once we find the matched jet, we also ensure that the daughters of the Z ($q\bar{q}$) get matched to this AK8 jet. The mass of the matched jet is plotted on the left in Fig. 3.6.

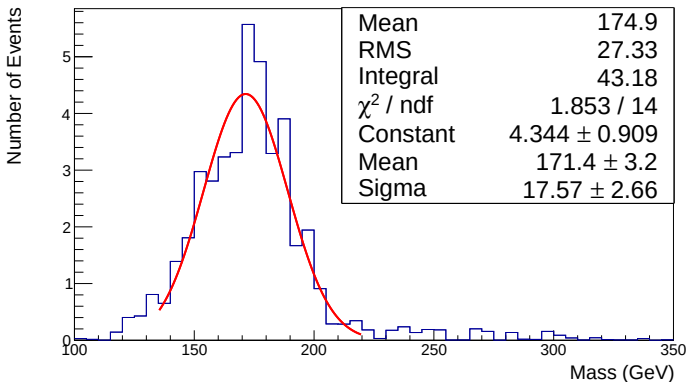


Figure 3.5: Invariant mass of three AK4 jets matched to the daughters of a top. The histogram has been fitted with a Gaussian. We use the mean and sigma values given towards the end of the legend which were calculated from the fit. The mean and RMS values at the top correspond to the histogram.

Particle	Mean (GeV)	Width (GeV)
Top (fully merged)	179.3	18.1
Top (partially resolved)	184	18.4
Top (fully resolved)	171.3	17.7
W ($q\bar{q}$')	80.9	9.5
Z (qq)	88.5	9.5

Table 3.4: Summary for mean and widths of reconstructed particles.

For each W ($q\bar{q}'$), we look for AK8 jets with $p_T > 200$ GeV/c, $|\eta| < 2.4$ and $\tau_{21} < 0.6$ falling within $\Delta R < 0.4$ of the W. Once we find the matched jet, we also ensure that the daughters of the W ($q\bar{q}'$) get matched to this AK8 jet. If all the above criteria are satisfied, we declare the jet as matched to the W. The mass of the matched jet is plotted on the right in Fig. 3.6.

Please refer to Table 3.4 for a summary of the mean and width of reconstructed top quarks and W or Z boson.

3.3 χ^2 minimization for T mass reconstruction

The key step to reconstruct the T mass is to correctly identify its daughters. Our signal topology has two top quarks, one hadronically decaying Z and one leptonically decaying Z. Our event selection

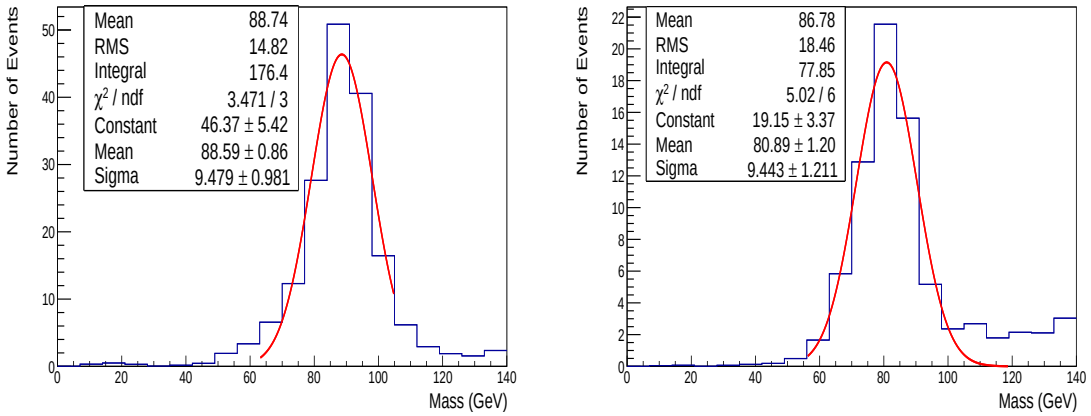


Figure 3.6: (Left) Mass of AK8 jets matched to Z. The histogram has been fitted with a Gaussian. (Right) Mass of AK8 jets matched to W. The histogram has been fitted with a Gaussian. We use the mean and sigma values given towards the end of the legend which were calculated from the fit. The mean and RMS values at the top correspond to the histogram.

has one dilepton ($Z \to \mu^+\mu^-$ or $Z \to e^+e^-$) candidate and 4 AK4 jets. But to reconstruct the other candidates, there are several possibilities at hand. For example, the top could get reconstructed as a fat jet, a partially resolved jet (W+b) or it might not get reconstructed at all from AK8 jets. Similarly, we might or might not be able to reconstruct V-jets from AK8 jets. It is thus a difficult task to identify the tops and the hadronically decaying Z bosons coming from the decay of the T quark.

The χ^2 minimization technique involves defining a χ^2 variable for a combination of jets in every event. The combination that minimizes this variable is considered as the correct reconstruction for the decay products of the T quark. The generic form of the χ^2 variable is defined in Eqn. 3.1.

$$\chi^2 = \frac{(M_t - m_{top_1})^2}{\sigma_t^2} + \frac{(M_t - m_{top_2})^2}{\sigma_t^2} + \frac{(M_Z - m_{Z(qq)})^2}{\sigma_Z^2} \tag{3.1}$$

where m_x is the mass of the reconstructed candidate object in the event, M_X is the mean mass of particle X (calculated in the previous section) and σ_X is the width of the particle X (calculated in the previous section). The values used for M_X and σ_X are given in Table 3.4. The names top_1 and top_2 refer to the two different reconstructed top-like candidates which are supposedly the decay products of the pair produced T quarks.

We have divided the algorithm into several categories such that each category has at least one boosted object. The categories are mentioned in Table 3.5.

Categories for χ^2 algorithm
At least 1 top jet and at least 2 V-jets
At least 1 top jet and 1 V-jet
1 top jet and no V-jet
No top jet and at least 2 V-jets
No top jet, 1 V-jet and at least 6 AK4 jets

Table 3.5: Categories used for the χ^2 algorithm

Now we will discuss each of the five categories for the χ^2 algorithm mentioned in Table 3.5.

3.3.1 At least 1 top jet and at least 2 V-jets

$$\chi^2 = \frac{(M_t - m_{top})^2}{\sigma_t^2} + \frac{(M_t - m_{V_1+b})^2}{\sigma_{V+b}^2} + \frac{(M_Z - m_{V_2})^2}{\sigma_Z^2} \tag{3.2}$$

Here, we consider that one top is reconstructed as a fully merged AK8 jet whereas the other top is reconstructed as partially resolved from a V-jet and an AK4 loose b-jet. The hadronic Z is also reconstructed from an AK8 jet. We consider all possible candidates for top, V_1, V_2 and the loose b-jet. The combination that gives the least value for Eqn 3.2 is considered as the best fit for reconstructed daughter objects.

3.3.2 At least 1 top jet and 1 V-jet

$$\chi^2 = \frac{(M_t - m_{top})^2}{\sigma_t^2} + \frac{(M_t - m_{V+AK4_3})^2}{\sigma_{V+b}^2} + \frac{(M_Z - m_{AK4_1+AK4_2})^2}{\sigma_Z^2} \tag{3.3}$$

Here again, we consider that one of the tops is completely boosted whereas the other top is reconstructed as partially resolved. The hadronic Z is reconstructed from two AK4 jets. We ensure that the jet $AK4_3$ passes the loose b-tagging criterion. We consider all possible combinations for the three AK4 jets arising in Eqn 3.3.

We also switch the positions of the V jet with $AK4_1$+$AK4_2$ with the added condition that the mass of $AK4_1$+$AK4_2$ should fall within the W mass window (65 - 105 GeV). This means that we also consider cases, where the hadronic Z is boosted and is reconstructed as the V jet and the other top is reconstructed as a tri-jet system ($AK4_1$+$AK4_2$+$AK4_3$). Eqn 3.3 gets modified as below:

$$\chi^2 = \frac{(M_t - m_{top})^2}{\sigma_t^2} + \frac{(M_t - m_{AK4_1+AK4_2+AK4_3})^2}{\sigma^2_{AK4+AK4+b}} + \frac{(M_Z - m_V)^2}{\sigma_Z^2} \tag{3.4}$$

The combination which then gives the least value for χ^2 is considered as the best fit for reconstructed objects.

3.3.3 1 top jet and no V-jet

$$\chi^2 = \frac{(M_t - m_{top})^2}{\sigma_t^2} + \frac{(M_t - m_{AK4_1+AK4_2+AK4_3})^2}{\sigma^2_{AK4+AK4+b}} + \frac{(M_Z - m_{AK4_3+AK4_4})^2}{\sigma_Z^2} \tag{3.5}$$

Here, we consider that one top is reconstructed as fully merged whereas the other top is reconstructed as completely resolved. The hadronic Z is also reconstructed as fully resolved. We make sure that the jet $AK4_3$ passes the loose b-tagging criterion and the mass of $AK4_1$+$AK4_2$ falls within the W mass window (65 - 105 GeV). To consider all possible cases, we take all possible combinations of the five AK4 jets in Eqn 3.5. The combination which then gives the least value for Eqn 3.5 is considered as the best fit for reconstructed objects.

3.3.4 No top jet and at least 2 V-jets

$$\chi^2 = \frac{(M_t - m_{V_1+AK4_3})^2}{\sigma^2_{V+b}} + \frac{(M_t - m_{V_2+AK4_4})^2}{\sigma^2_{V+b}} + \frac{(M_Z - m_{AK4_1+AK4_2})^2}{\sigma_Z^2} \tag{3.6}$$

Here, neither of the tops are boosted enough to be reconstructed as fat jets. We make sure that the AK4 jets used to reconstruct the partially resolved tops pass the loose b-tagging criterion. To consider all possible cases, we consider all possible candidates for V_1 and V_2 along with all possible combinations of the four AK4 jets in Eqn 3.6. We also switch the position of V_1 or V_2 with $AK4_1$+$AK4_2$ with the added condition that mass of $AK4_1$+$AK4_2$ should fall within W mass window (65 - 105 GeV). This means that we also consider cases, where one of the tops is reconstructed as completely resolved and the hadronic Z is reconstructed as a fat V jet. Eqn 3.6 gets modified as given below:

$$\chi^2 = \frac{(M_t - m_{AK4_1+AK4_2+AK4_3})^2}{\sigma^2_{AK4+AK4+b}} + \frac{(M_t - m_{V_2+AK4_4})^2}{\sigma^2_{V+b}} + \frac{(M_Z - m_{V_1})^2}{\sigma_Z^2} \tag{3.7}$$

Even for Eqn 3.7, we make sure that we consider all possible combination of AK4 jets and the two

V jets.

The combination which then gives the least value for χ^2 is considered as the best fit for reconstructed objects.

3.3.5 No top jet, 1 V-jet and at least 6 AK4 jets

$$\chi^2 = \frac{(M_t - m_{AK4_1+AK4_2+AK4_5})^2}{\sigma^2_{AK4+AK4+b}} + \frac{(M_t - m_{V+AK4_6})^2}{\sigma^2_{V+b}} + \frac{(M_Z - m_{AK4_3+AK4_4})^2}{\sigma^2_Z} \tag{3.8}$$

Here, one top is reconstructed as fully resolved and the other is reconstructed as partially resolved. For the fully resolved top, it is necessary that one of the three AK4 jets pass the loose b-tagging criterion and the mass of the other two AK4 jets combined should fall within the W mass window (65-105 GeV). To consider all possible cases, we consider all possible combinations of the 6 AK4 jets in Eqn 3.8. We also switch the position of V with $AK4_1$+$AK4_2$ or $AK4_3$+$AK4_4$ with the added condition that mass of $AK4_1$+$AK4_2$ or $AK4_3$+$AK4_4$ should fall within W mass window (65 - 105 GeV). This means that we also take into account, those cases where the hadronic Z is reconstructed as a fat V jet and both the tops are reconstructed as completely resolved tri-jet systems. Eqn 3.8 gets modified as given below:

$$\chi^2 = \frac{(M_t - m_{AK4_1+AK4_2+AK4_5})^2}{\sigma^2_{AK4+AK4+b}} + \frac{(M_t - m_{AK4_3+AK4_4+AK4_6})^2}{\sigma^2_{AK4+AK4+b}} + \frac{(M_Z - m_V)^2}{\sigma^2_Z} \tag{3.9}$$

The combination which then gives the least value for χ^2 is considered as the best fit for reconstructed objects.

For every case, once we find the most suitable candidate objects, there are two ways to reconstruct the T mass:

- A: Combine the four vectors of top_1 and Z(ll) to reconstruct the leptonically decaying T along with combining top_2 with Z(qq) to reconstruct the hadronically decaying T
- B: Combine the four vectors of top_2 and Z(ll) to reconstruct the leptonically decaying T along with combining top_1 with Z(qq) to reconstruct the hadronically decaying T

To resolve the ambiguity, we take the mass difference between the leptonic T and the hadronic T in both case A and case B from above. The case which has a lesser mass difference is considered

Scale factors applied
Trigger scale factor
Lepton identification and isolation scale factors
Z p_T dependent electroweak correction factors for the Z+jets sample
b-tagging efficiency scale factors
Top tagging scale factor
W or Z tagging scale factor

Table 3.6: Scale factors applied to the MC samples.

	Source	Uncertainty (in %)
Normalization	$t\bar{t}$ cross section	15
	Drell-Yan cross section	15
	Diboson cross section	20
	Integrated luminosity	2.6
	Lepton identification and isolation	3
	Trigger efficiency	1
Shape	b-tagging scale factor	shape
	Jet Energy Correction (JEC)	shape
	Jet Energy Resolution (JER)	shape

Table 3.7: Systematic Uncertainties for the background and signal yields. The uncertainty on the luminosity is motivated from [59] and the uncertainty on the cross-sections are motivated from [60–62].

as the correct reconstruction for the two T quarks in the event. For a given combination of jets, we expect the χ^2 to be lower for signal than for background.

We had discussed in Chapter 2 about the scale factors that should be applied to the MC to correct for the differences in efficiencies as measured in the data and those which are predicted in the MC simulation. We will enlist the scale factors that are applied again in Table 3.6.

We also consider systematic uncertainties that are associated with these scale factors and the theoretical uncertainties associated with the cross section calculations for various samples. The systematic uncertainties are mentioned in Table 3.7. Note that there are two different types of uncertainties, one that affects the overall normalization of any distribution and the other affects the shape of the distribution.

The event yields for all the five cases with the associated errors are shown in Table 3.8. The χ^2 variable for all the five cases is plotted in Fig. 3.7. Note that the signal tends to usually have lower

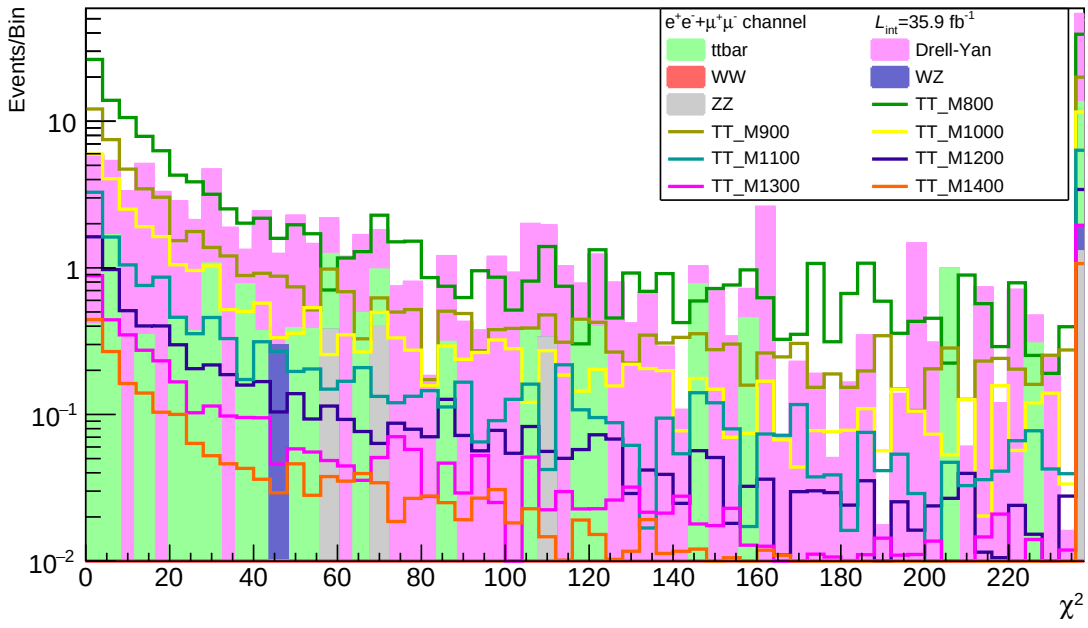


Figure 3.7: The χ^2 distribution.

values of χ^2 when compared to the background. The corresponding T mass is plotted on the left in Fig. 3.8. We can see that the distributions for M(T) = 800 GeV and M(T) = 900 GeV already stand out from the background. Also, the background distribution peaks at lower values of masses whereas the signal distribution peaks at higher values. We then impose a selection of $\chi^2 < 20$ on our events and the corresponding event yields are shown in Table 3.9 and the T mass is plotted on the right in Fig. 3.8. At this point, we can see even M(T) = 1000 GeV and M(T) = 1100 GeV standing out from the background.

3.4 Data vs Monte Carlo comparisons

We need to ensure that the MC simulation closely reproduces various distributions as measured in the data and takes into account various effects like detector geometry, efficiency etc. For this, we compare the data with the MC samples in a background enriched region with negligible contribution from the signal. We will do this in two separate channels i.e. the electron channel ($Z \rightarrow e^+e^-$) and the muon channel ($Z \rightarrow \mu^+\mu^-$). We use the same systematic uncertainties as mentioned in Table 3.6 and the percentage of uncertainties considered are the same as given in Table 3.7. The

Background					
Category type	$t\bar{t}$	Z+jets	WW	WZ	ZZ
At least 1 top jet and at least 2 V-jets	0	0.39±0.6±0.4	0	0	0
At least 1 top jet and 1 V-jet	1.12±1.06±0.23	4.15±2.04±1.6	0	0	0
1 top and no V-jet	2.54±1.59±0.38	16.94±4.12±6.18	0	0	0.32±0.57±0.45
No top jet and at least 2 V-jets	3.31±1.81±0.9	16.67±4.08±7.41	0	0	0
No top jet, 1 V-jet and at least 6 AK4 jets	17.52±4.19±3.89	68.96±8.30±25.35	0	0.96±0.98±0.36	2.14±1.46±0.51
Signal					
Category type	M(T) = 800 GeV	M(T) = 900 GeV	M(T) = 1000 GeV	M(T) = 1100 GeV	M(T) = 1200 GeV
At least 1 top jet and at least 2 V-jets	4.15 ± 2.04±1.95	2.17±1.47±0.64	1.75±1.32±0.63	1.22±1.1±0.51	0.65±0.81±0.24
At least 1 top jet and 1 V-jet	27.47 ± 5.24±9.04	15.51±3.94±5.31	9.10±3.02±3.29	5.0±2.24±1.82	3.08±1.75±1.19
1 top and no V-jet	31.06 ± 5.57±9.84	15.67±3.96±5.15	8.68±2.95±2.77	4.31±2.08±1.43	2.26±1.50±0.77
No top jet and at least 2 V-jets	34.44 ± 5.87±10.84	15.73±3.96±5.16	9.04±3.01±2.96	4.31±2.08±1.48	2.18±1.46±0.80
No top jet, 1 V-jet and at least 6 AK4 jets	71.19±8.44±23.08	29.22±5.41±9.92	14.19±3.77±4.72	6.56±2.56±2.36	3.35±1.83±1.20

Table 3.8: Event yields from χ^2 algorithm for background and signal samples.The numbers are written as N±(stat)±(syst).

Background					
Category type	$t\bar{t}$	Z+jets	WW	WZ	ZZ
At least 1 top jet and at least 2 V-jets	0	0.11±0.34±0.37	0	0	0
At least 1 top jet and 1 V-jet	0.42±0.65±0.20	3.08±1.76±0.95	0	0	0
1 top and no V-jet	0.81±0.9±0.24	4.69±2.17±1.85	0	0	0
No top jet and at least 2 V-jets	0.77±0.87±0.21	6.61±2.57±3.57	0	0	0
No top jet, 1 V-jet and at least 6 AK4 jets	0.49±0.70±0.22	6.73±2.59±1.84	0	0	0
Signal					
Category type	M(T) = 800 GeV	M(T) = 900 GeV	M(T) = 1000 GeV	M(T) = 1100 GeV	M(T) = 1200 GeV
At least 1 top jet and at least 2 V-jets	3.31 ± 1.82±1.75	1.60±1.26±0.44	1.35±1.16±0.53	0.98±0.990±0.43	0.50±0.71±0.19
At least 1 top jet and 1 V-jet	22.02 ± 4.70±7.13	12.87±3.59±4.39	7.03±2.65±2.52	3.58±1.89±1.35	2.05±1.43±0.84
1 top and no V-jet	12.1 ± 3.48±4.35	6.17±2.48±2.05	2.96±1.72±0.97	1.38±1.17±0.51	0.63±0.79±0.23
No top jet and at least 2 V-jets	15.71 ± 3.96±5.62	6.48±2.55±2.54	3.28±1.81±1.34	1.19±1.09±0.48	0.55±0.74±0.3
No top jet, 1 V-jet and at least 6 AK4 jets	14.77±3.84±5.54	5.03±2.24±2.14	2.2±1.48±0.75	0.78±0.89±0.37	0.35±0.6±0.14

Table 3.9: Event yields for $\chi^2 < 20$ for background and signal samples. The numbers are written as N±(stat)±(syst).

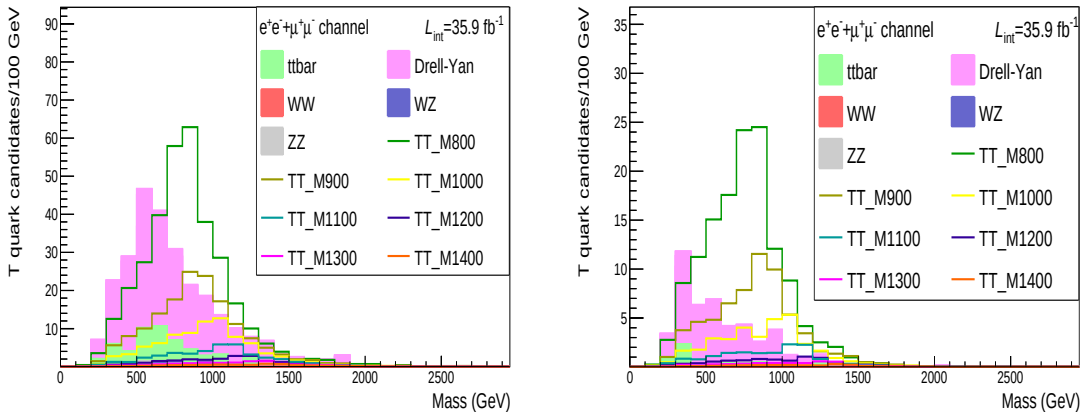


Figure 3.8: (Left) Mass of leptonic and hadronic T quarks reconstructed from the χ^2 algorithm. (Right) Mass of leptonic and hadronic T quarks reconstructed after imposing $\chi^2 < 20$.

Variable	Requirement
Leading muon p_T (if Z $\rightarrow \mu^+\mu^-$)	$\geq$45 GeV
Leading electron p_T (if Z $\rightarrow e^+e^-$)	$\geq$120 GeV
Sub-leading lepton p_T (electron or muon)	$\geq$25 GeV
Number e^+e^-or $\mu^+\mu^-$pair in the Z Mass window (75 - 105 GeV)	1
Number of AK4 jets	≥ 3
Z p_T	> 100 GeV
H_T	> 500 GeV

Table 3.10: Background enriched control region pre-selection.

pre-selections that we will use to make these comparisons are listed in Table 3.10.

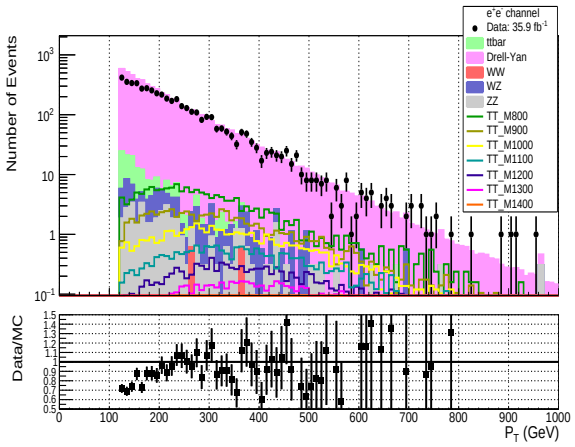
3.4.1 Electron Channel

We will show how applying scale factors to the MC affects the agreement between the data and the MC. We will first show a raw comparison between the two, without applying any scale factors to the MC. The comparisons are shown on the left in Fig. 3.9. Clearly, the agreement is rather off throughout the entire p_T range. We then apply trigger scale factors, electron identification and isolation scale factors to all the MC samples and the Z p_T dependent electroweak correction factors for the Drell-Yan sample. We again compare the MC and the data and the plots are shown on the right in Fig. 3.9. Clearly, the agreement improves quite a bit, especially in the low p_T range.

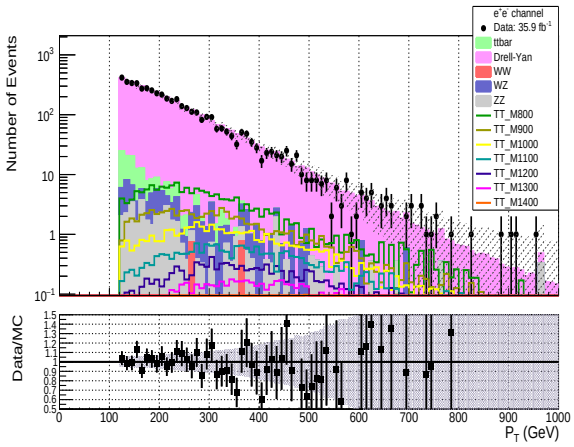
We will now show several other comparisons before and after applying the scale factors (Fig. 3.10 and Fig. 3.11).

Lastly, we will show the effect of applying b-tagging scale factors on the b-jet multiplicity. The comparison is shown in Fig. 3.12. Again, one can see that this improves the Data/MC ratio. After making these comparisons, we can safely say that the MC reproduces very closely, the effects which are seen in the data, in the electron channel.

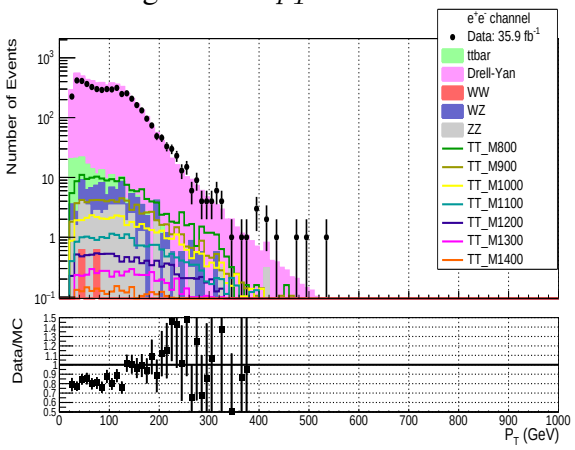
If one wishes to see the effect of applying just the trigger and lepton identification and isolation scale factors to the MC or the effect of the Z p_T dependent electroweak correction factors to the Drell-Yan sample, please refer to Appendix A.



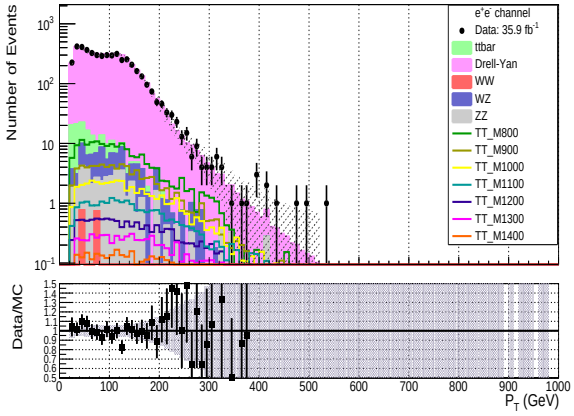
Leading electron p_T distribution before.



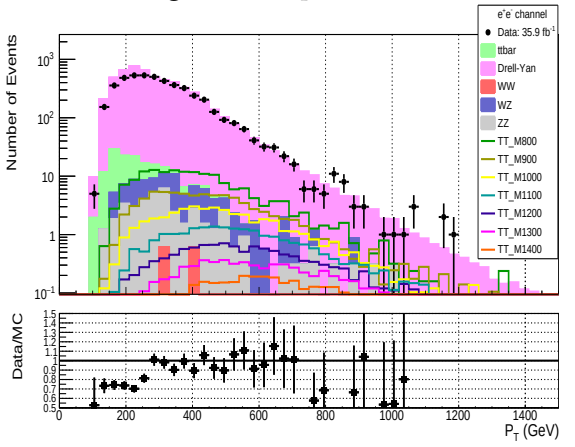
Leading electron p_T distribution after.



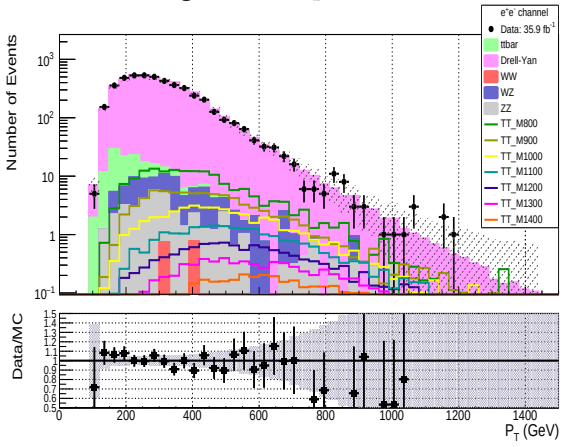
Sub-leading electron p_T distribution before.



Sub-leading electron p_T distribution after.

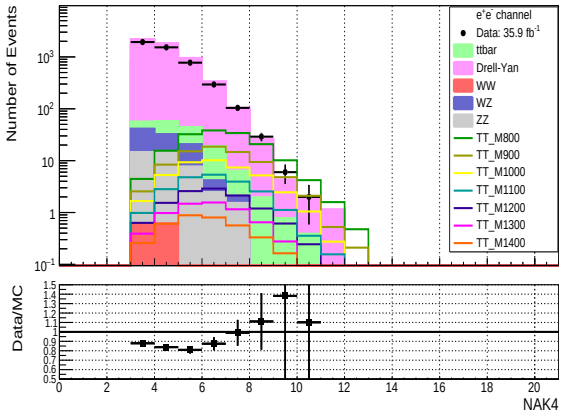


Dielectron p_T distribution before.

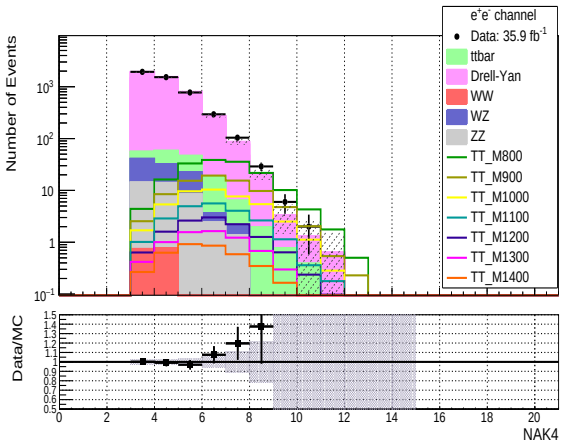


Dielectron p_T distribution after.

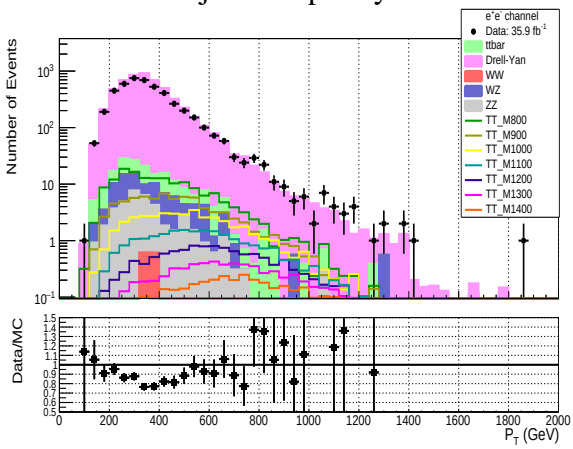
Figure 3.9: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the electron identification and isolation scale factors, the trigger scale factors and the Z p_T dependent electroweak correction factors for Drell-Yan. The error band shows both statistical uncertainty as well as the systematic uncertainty on the three scale factors applied along with the theoretical uncertainty on the cross section for the background samples and the uncertainty on the integrated luminosity.



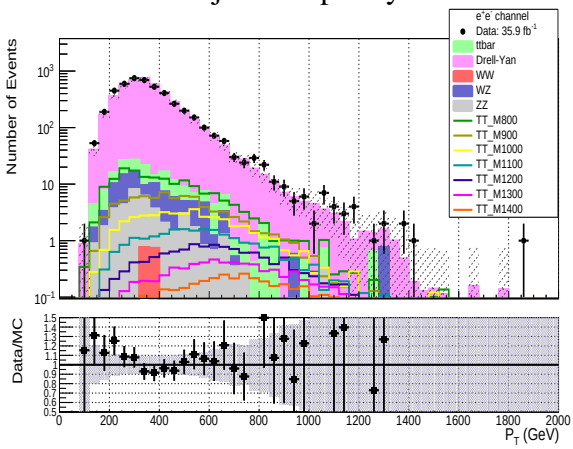
AK4 jet multiplicity before.



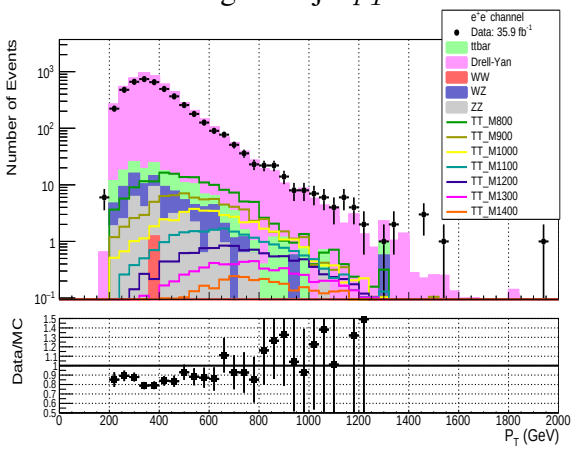
AK4 jet multiplicity after.



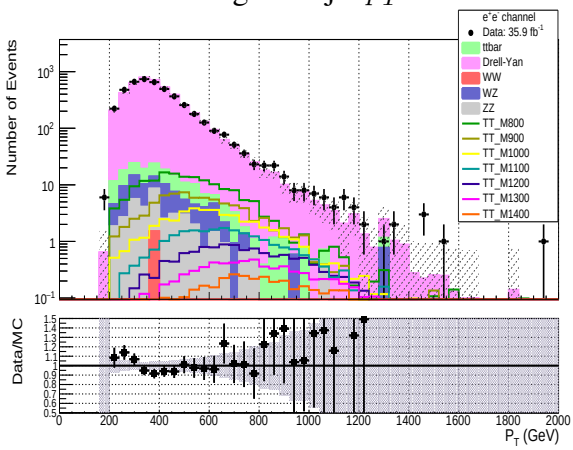
Leading AK4 jet p_T before.



Leading AK4 jet p_T after.

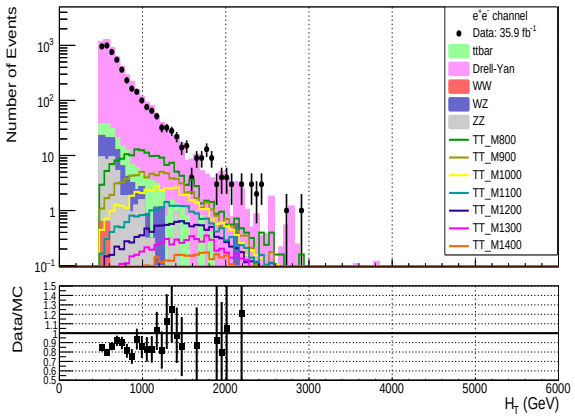


Leading AK8 jet p_T before.

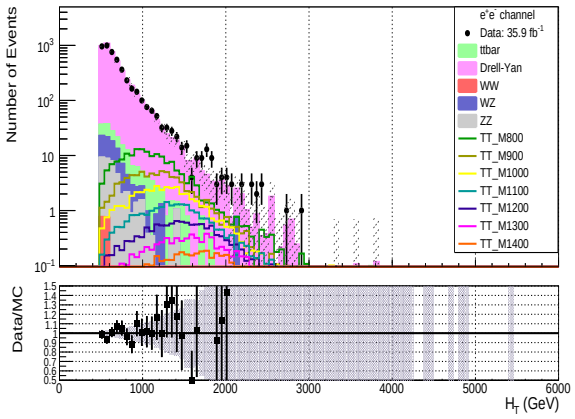


Leading AK8 jet p_T after.

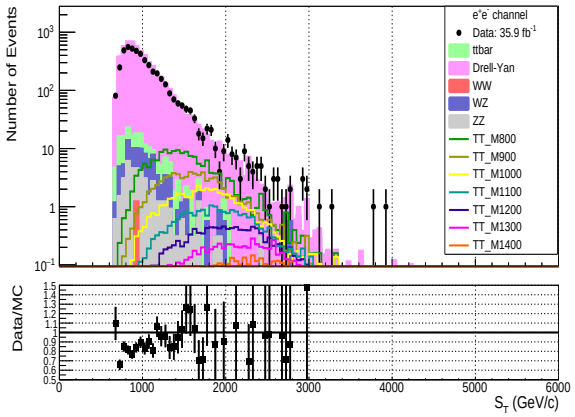
Figure 3.10: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the electron identification and isolation scale factors, the trigger scale factors and the Z p_T dependent electroweak correction factors for Drell-Yan. The error band shows both statistical uncertainty as well as the systematic uncertainty on the three scale factors applied along with the theoretical uncertainty on the cross section for the background samples and the uncertainty on the integrated luminosity.



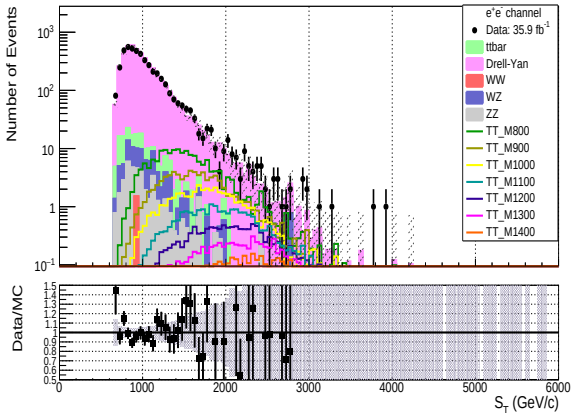
H_T distribution before.



H_T distribution after.

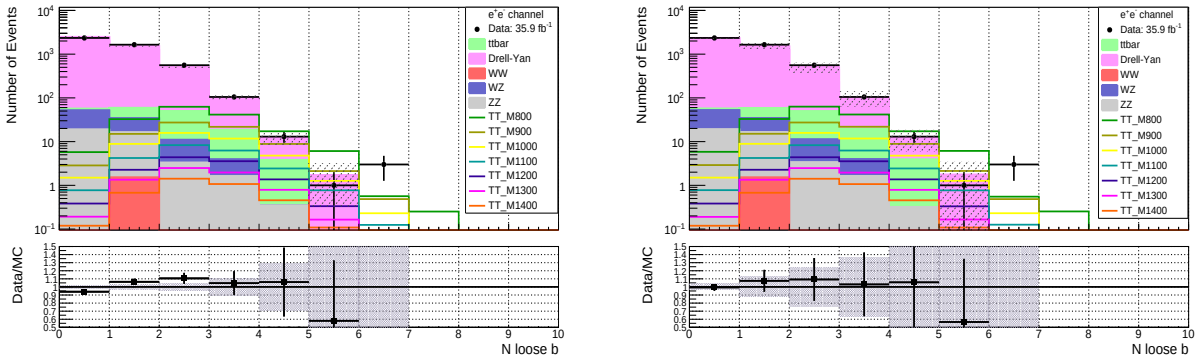


S_T distribution before.



S_T distribution after.

Figure 3.11: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the electron identification and isolation scale factors, the trigger scale factors and the Z p_T dependent electroweak correction factors for Drell-Yan. The error band shows both statistical uncertainty as well as the systematic uncertainty on the three scale factors applied along with the theoretical uncertainty on the cross section for the background samples and the uncertainty on the integrated luminosity.



Loose b-jet multiplicity before. Loose b-jet multiplicity after.

Figure 3.12: (Left) The plot shows the distribution with just the electron identification and isolation scale factors, trigger scale factors and the Z p_T dependent electroweak correction factors to the Drell-Yan sample. (Right) The same plots after applying the b-tagging scale factors. The error band for the plot on the right includes statistical uncertainty, theoretical uncertainty on the cross-sections of the background samples, the uncertainty on the integrated luminosity and the systematic uncertainties for all the scale factors applied, including the b-tagging shape uncertainty and the JEC and JER shape uncertainties.

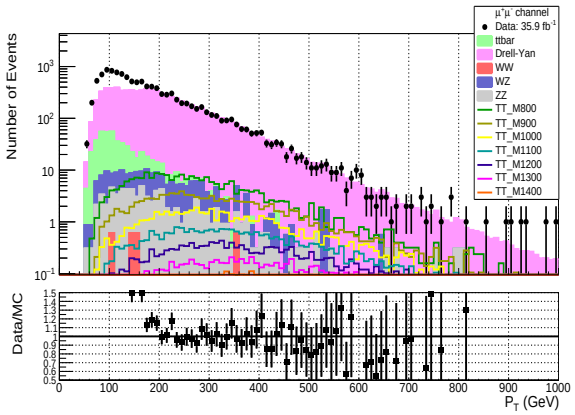
3.4.2 Muon Channel

We will now carry out a similar exercise in the Muon channel and show how different scale factors affect the agreement between the data and the MC. We will begin by making a raw comparison between the data and the MC (i.e. without any scale factors applied to the MC). The comparisons are shown on the left in Fig. 3.13. Again, the agreement is rather off throughout the entire p_T range, especially on the lower side. We then apply trigger scale factors, muon identification and isolation scale factors and the Z p_T dependent electroweak correction factors for the Drell-Yan sample. We again compare the MC and the data and the plots are shown on the right in Fig. 3.13. It can be seen that the agreement improves quite a bit.

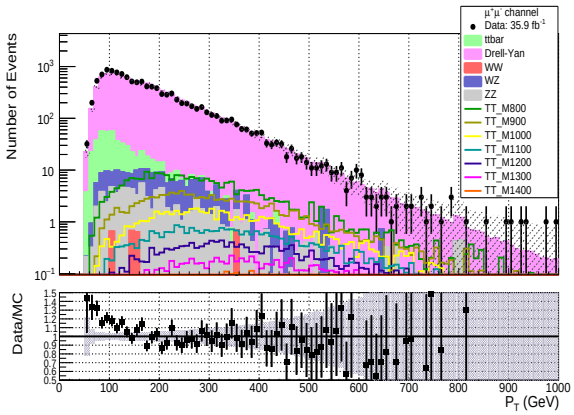
We will now show several other comparisons before and after applying the scale factors for the MC (Fig. 3.14 and Fig. 3.15).

Lastly, we will show the effect of applying b-tagging scale factors on the b-jet multiplicity. The comparison is shown in Fig. 3.16. Again, one can see that this improves the Data/MC ratio.

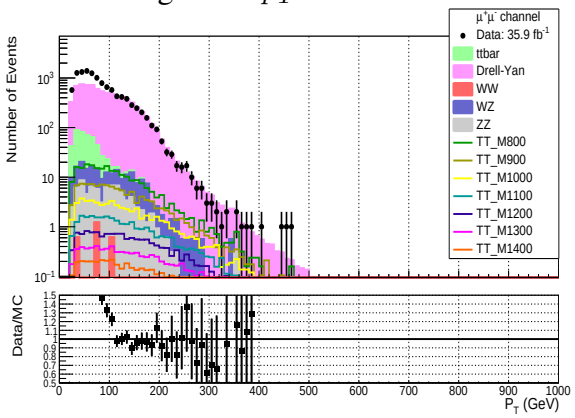
After making these comparisons, we can safely say that the MC reproduces very closely, the effects which are seen in the data, in the muon channel as well. If one wishes to see the effect of applying just the trigger and lepton identification and isolation scale factors to the MC or the effect of the Z



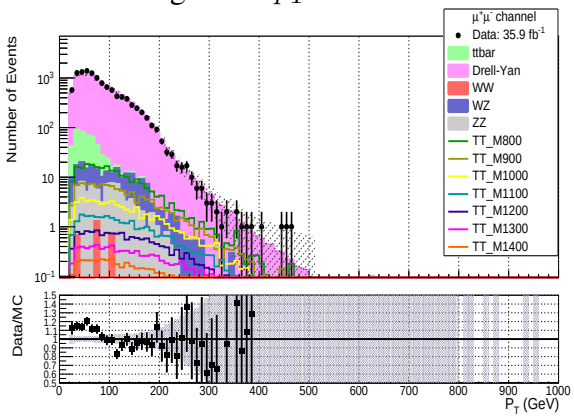
Leading muon p_T distribution before.



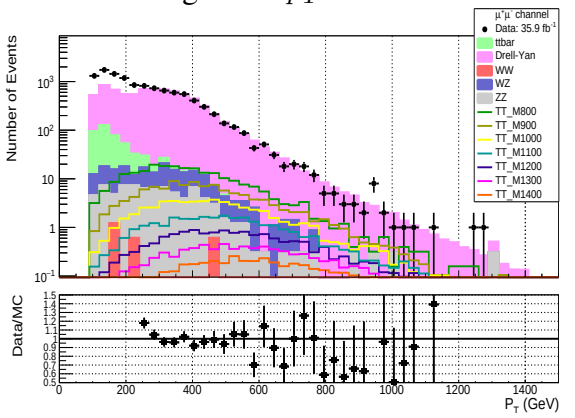
Leading muon p_T distribution after.



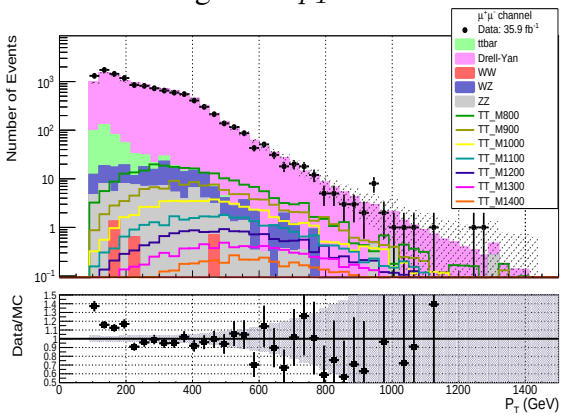
Sub-leading muon p_T distribution before.



Sub-leading muon p_T distribution after.

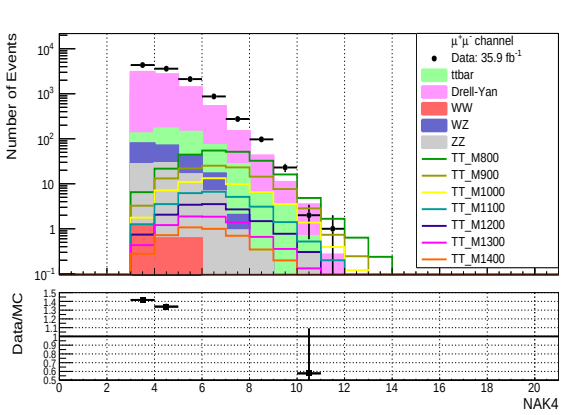


Dimuon p_T distribution before.

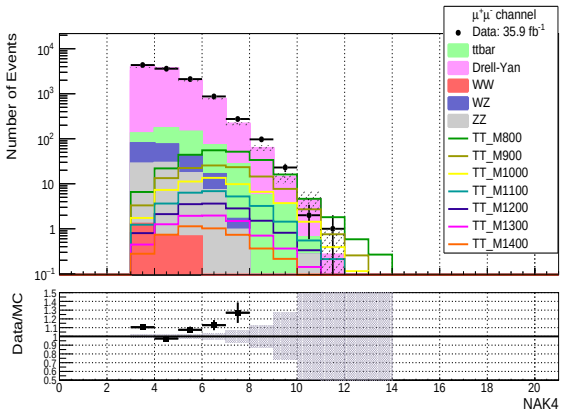


Dimuon p_T distribution after.

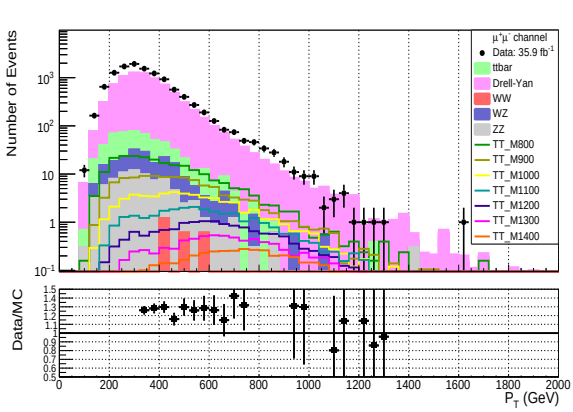
Figure 3.13: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the muon identification and isolation scale factors, the trigger scale factors and the Z p_T dependent electroweak correction factors for Drell-Yan. The error band shows both statistical uncertainty as well as the systematic uncertainty on the three scale factors applied along with the theoretical uncertainty on the cross section for the background samples and the uncertainty on the integrated luminosity.



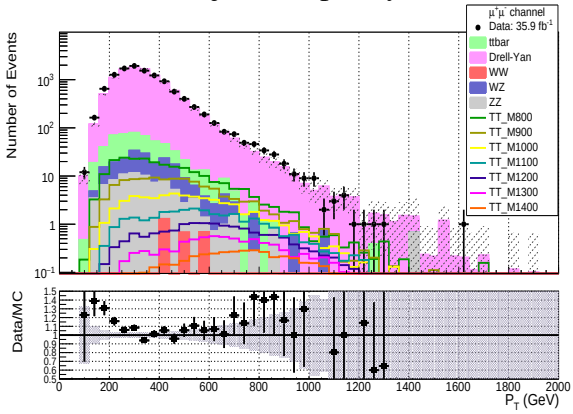
AK4 jet multiplicity before.



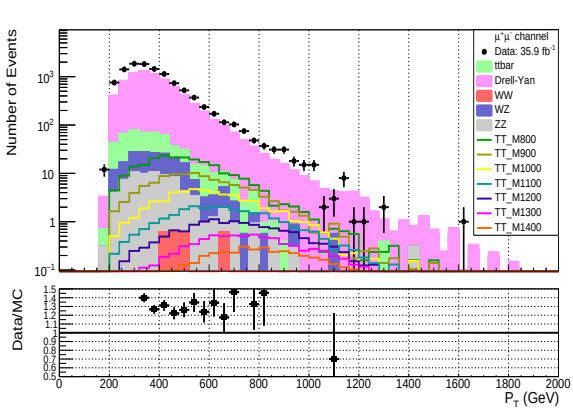
AK4 jet multiplicity after.



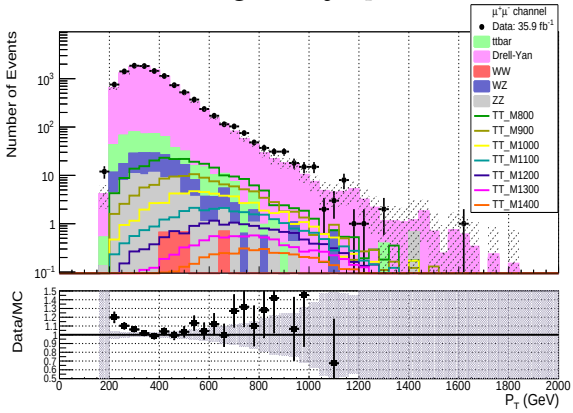
Leading AK4 jet p_T before.



Leading AK4 jet p_T after.

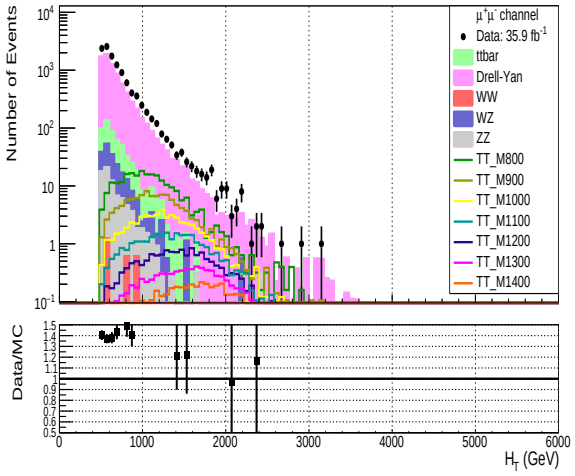


Leading AK8 jet p_T before.

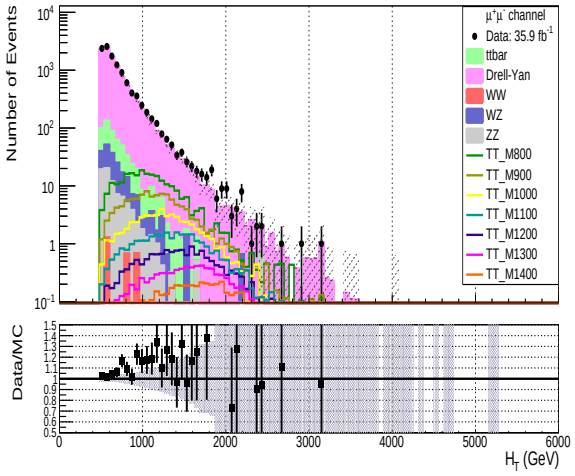


Leading AK8 jet p_T after.

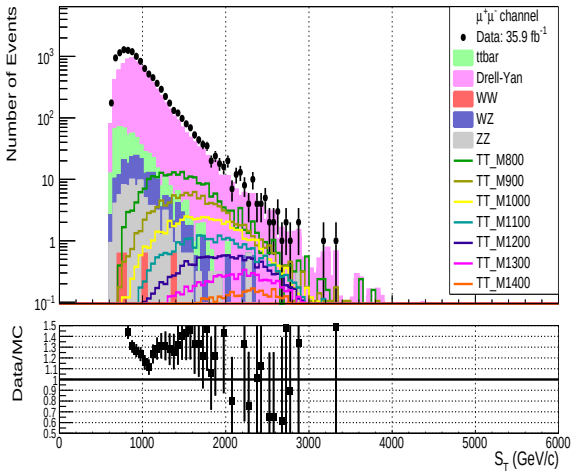
Figure 3.14: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the muon identification and isolation scale factors, the trigger scale factors and the Z p_T dependent electroweak correction factors for Drell-Yan. The error band shows both statistical uncertainty as well as the systematic uncertainty on the three scale factors applied along with the theoretical uncertainty on the cross section for the background samples and the uncertainty on the integrated luminosity.



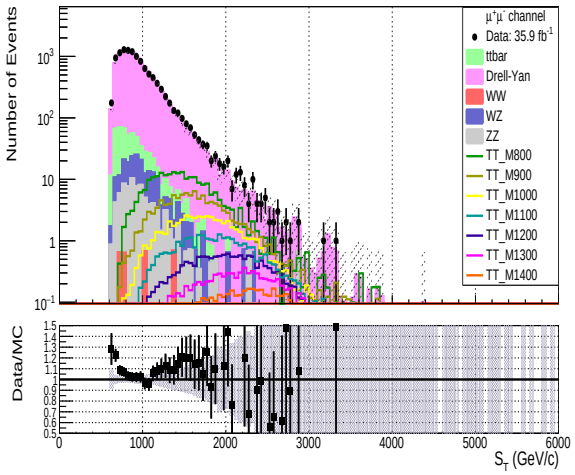
H_T distribution before.



H_T distribution after.

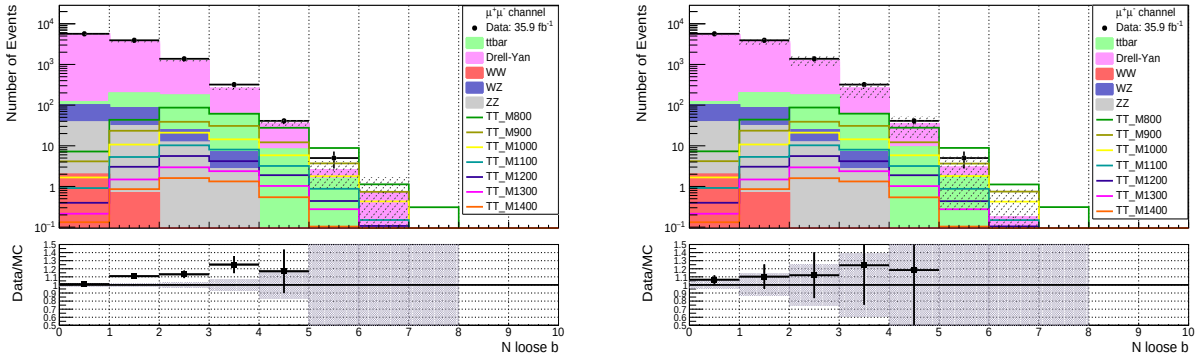


S_T distribution before.



S_T distribution after.

Figure 3.15: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the muon identification and isolation scale factors, the trigger scale factors and the Z p_T dependent electroweak correction factors for Drell-Yan. The error band shows both statistical uncertainty as well as the systematic uncertainty on the three scale factors applied along with the theoretical uncertainty on the cross section for the background samples and the uncertainty on the integrated luminosity.



Loose b-jet multiplicity before. Loose b-jet multiplicity after.

Figure 3.16: (Left) The plot shows the distribution with just the muon identification and isolation scale factors, trigger scale factors and the Z p_T dependent electroweak correction factors to the Drell-Yan sample. (Right) The same plots after applying the b-tagging scale factors. The error band for the plot on the right includes statistical uncertainty, theoretical uncertainty on the cross-sections of the background samples, the uncertainty on the integrated luminosity and the systematic uncertainties for all the scale factors applied, including the b-tagging shape uncertainty and the JEC and JER shape uncertainties.

p_T dependent electroweak correction factors to the Drell-Yan sample, please refer to Appendix A.

Chapter 4

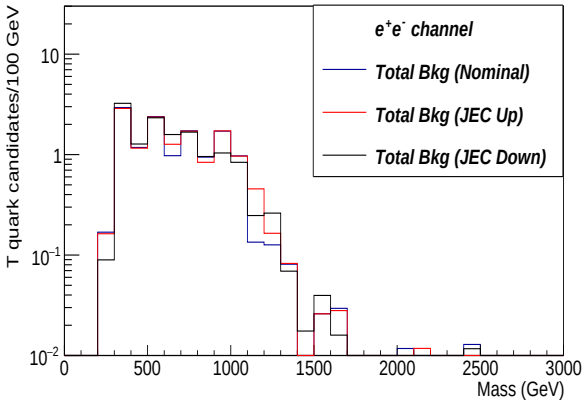
Results

In this chapter, we will consider systematic uncertainties that can affect the normalization and the shape of the T mass distributions. We test the χ^2 algorithm on the data with an integrated luminosity of 35.9 fb^{-1} and evaluate 95 % CL expected and observed upper limits on TT production cross section at each of the mass points considered in Table 2.1.

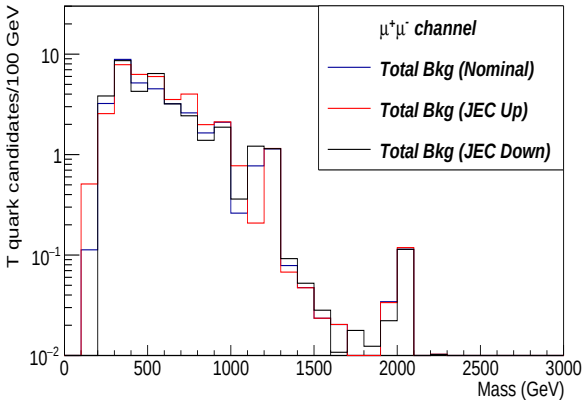
4.1 Systematic Uncertainties

We have previously mentioned some of the systematic uncertainties that we consider in this analysis (Table 3.7). Some of these uncertainties only affect the normalization of the distribution and we take a flat percentage to account for such uncertainties. There are some uncertainties, however, that affect the shape of the distribution. They affect different bins differently for a given distribution. We consider three types of shape uncertainties: the b-tagging uncertainty, the JEC and the JER uncertainties. The JEC and JER uncertainties are calculated by varying these parameters by $\pm 1\sigma$ and re-creating kinematic distributions for the MC. The b-tagging shape uncertainty is calculated by scaling up and down, the b-tagging scale factors for heavy b,c and light flavour quarks. The effect of these uncertainties on the shape of the background mass distribution as evaluated by the χ^2 algorithm after a requirement of $\chi^2 < 20$ is shown in Fig. 4.1. It is clear that these uncertainties cause only small changes in the overall mass distribution.

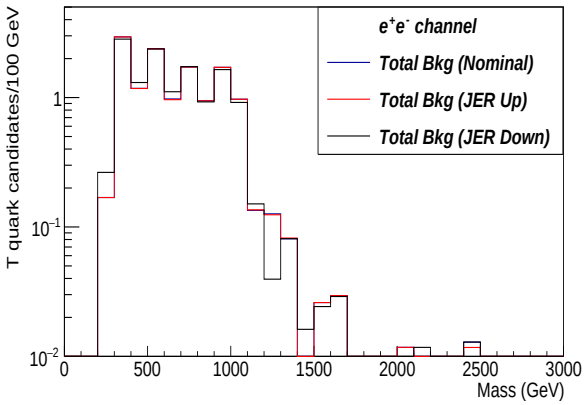
Apart from the already mentioned uncertainties, we also take into account a 3% uncertainty due to pileup re-weighting, a 1 % uncertainty on the PDF distributions [35], a 2.6 % uncertainty on



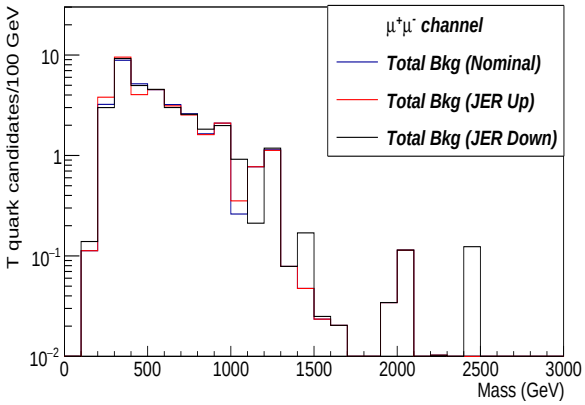
JEC systematics for electron channel.



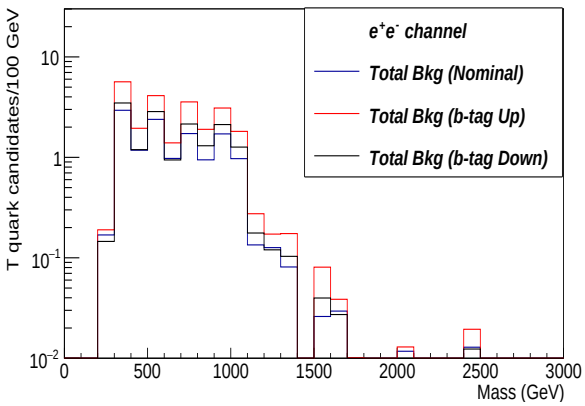
JEC systematics for muon channel.



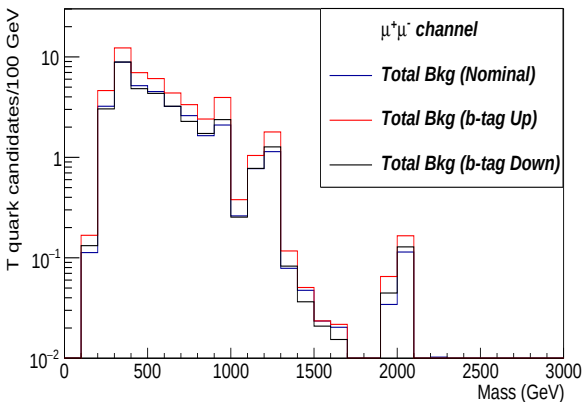
JER systematics for electron channel.



JER systematics for muon channel.



b-tagging systematics for electron channel.



b-tagging systematics for muon channel.

Figure 4.1: JEC, JER and b-tagging shape systematics for the background. The mass distribution is evaluated by the χ^2 algorithm after a requirement of $\chi^2 < 20$.

	Source	Uncertainty (in %)
Normalization	$t\bar{t}$ cross section	15
	Drell-Yan cross section	15
	Diboson cross section	20
	Integrated luminosity	2.6
	Lepton identification and isolation	3
	Trigger efficiency	1
	Pileup re-weighting	3
	PDF uncertainty	1
	W or Z tagging	4
	top tagging	7
Shape	b-tagging scale factor	shape
	Jet Energy Correction (JEC)	shape
	Jet Energy Resolution (JER)	shape

Table 4.1: Systematic Uncertainties for the background and signal yields.

Category type	Data N±(stat)	$t\bar{t}$ N±(stat)±(syst)	Z+jets N±(stat)±(syst)	M(T) = 800 GeV N±(stat)±(syst)	M(T) = 900 GeV N±(stat)±(syst)
At least 1 top jet and at least 2 V-jets	0	0	0.11±0.34±0.37	3.31 ± 1.82±1.75	1.60±1.26±0.44
At least 1 top jet and 1 V-jet	0	0.42±0.65±0.20	3.08±1.76±0.95	22.02 ± 4.70±7.13	12.87±3.59±4.39
1 top and no V-jet	6.0± 2.45	0.81±0.9±0.24	4.69±2.17±1.85	12.1 ± 3.48±4.35	6.17±2.48±2.05
No top jet and at least 2 V-jets	3.0 ± 1.73	0.77±0.87±0.21	6.61±2.57±3.57	15.71 ± 3.96±5.62	6.48±2.55±2.54
No top jet, 1 V-jet and at least 6 AK4 jets	11.0 ± 3.31	0.49±0.70±0.22	6.73±2.59±1.84	14.77±3.84±5.54	5.03±2.24±2.14

Table 4.2: Event yields in different categories for $\chi^2 < 20$ in the data and the MC. The event yields for the diboson samples were not included here because they were zero in all categories (Table 3.9).

the luminosity [59], a selection efficiency of 1.04 $\pm$ 0.04 on the W or Z tagging and a selection efficiency of 0.93 $\pm$ 0.07 on the top tagging. The uncertainty on the cross-sections are motivated from [60–62]. Table 4.1 lists all the systematic uncertainties used.

4.2 Results of the χ^2 algorithm on the data

We have unblinded the data by finding the results of the χ^2 algorithm on the data (Fig 4.2). The mass distribution for $\chi^2 < 20$ for both the data and the MC were separated into the electron and muon channels and the plots along with the systematic uncertainties are shown in Fig. 4.3. The mass distribution for the combined electron and muon channel is shown in Fig. 4.4. We have also listed the event yields for $\chi^2 < 20$ in the data and the MC for different categories of the χ^2 algorithm in Table 4.2.

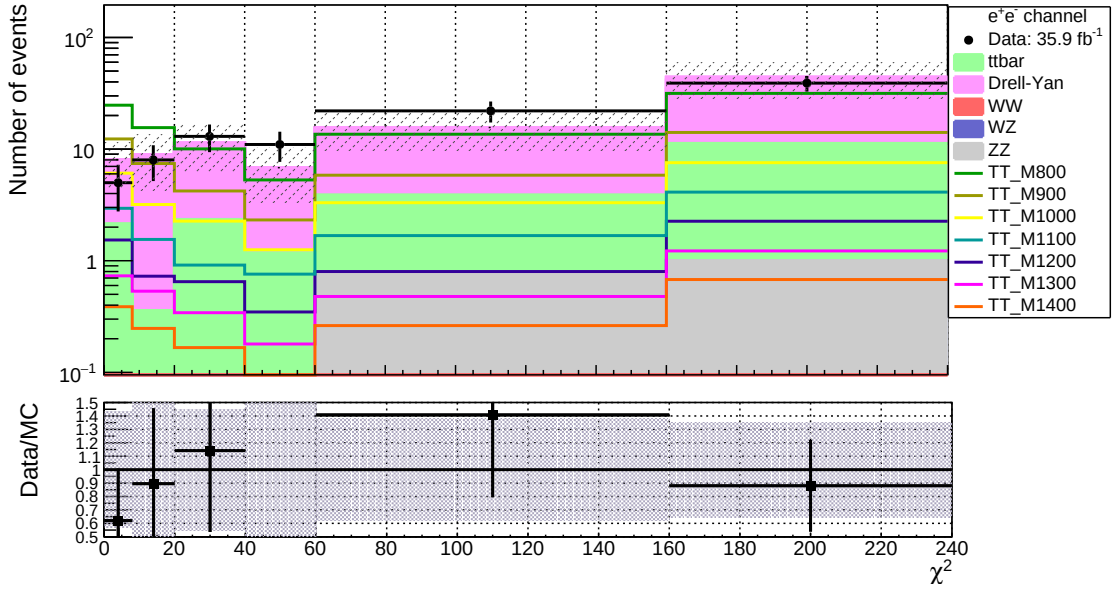
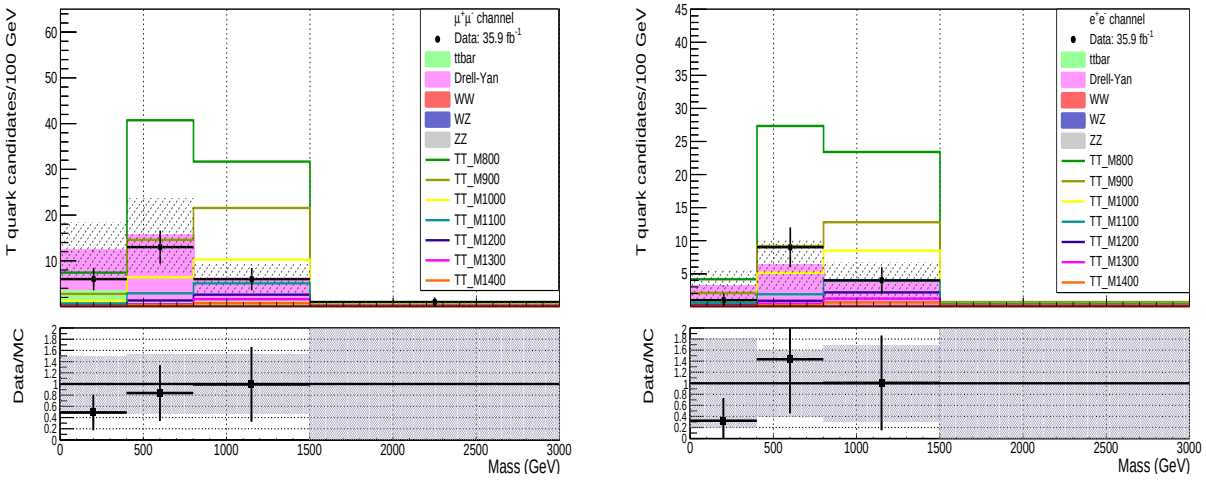


Figure 4.2: χ^2 distribution for the combined e^+e^-+ $\mu^+\mu^-$channel.



Muon channel Electron channel

Figure 4.3: T mass distribution for $\chi^2 < 20$.

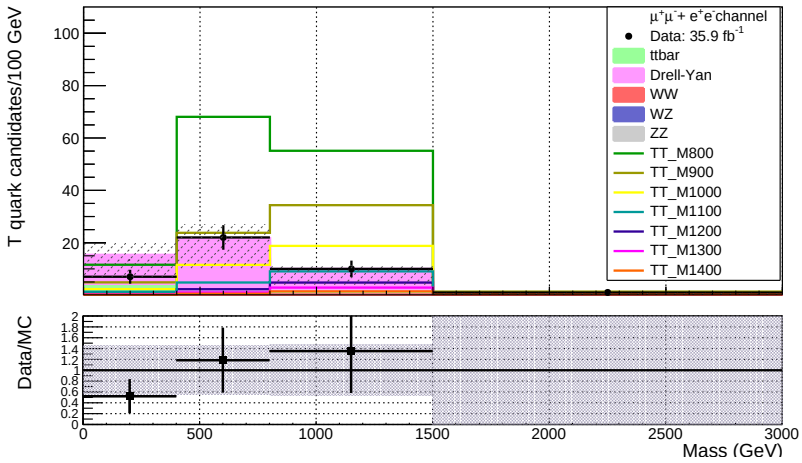


Figure 4.4: T mass distribution for $\chi^2 < 20$ in the combined e^+e^-+ $\mu^+\mu^-$channel.

4.3 Expected and Observed Limits on the T mass

We used the CMS Higgs Combined Limit tool to calculate expected and observed 95% CL upper limits on TT production cross-section using the Asymptotic CLs method [63]. We take the nuisance parameters as mentioned in Table 4.1 and calculate a likelihood fit for the T mass reconstructed with the χ^2 algorithm with the requirement of $\chi^2 < 20$. The actual mass distributions which were used to calculate the limits can be found in Appendix B (Fig. B.1). The expected and observed 95% CL upper limits on the TT production cross-section as a function of the T mass for an integrated luminosity of 35.9 fb^{-1} are plotted in Fig. 4.5. The expected 95% CL lower limit on the T quark mass is 1170 GeV for the decay TT→tZtZ with 100% branching ratio. We did not observe any excess in the data and the observed lower limit is 1095 GeV on the T mass for the decay TT→tZtZ with 100% branching ratio.

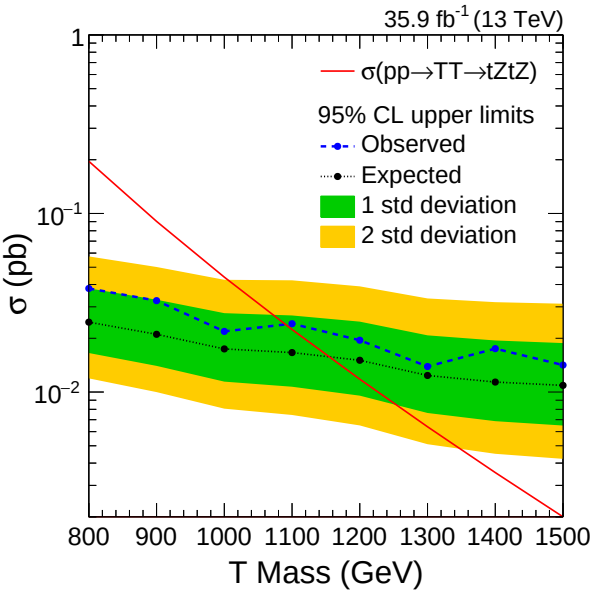


Figure 4.5: The observed and expected limits for TT→tZtZ with 100% branching ratio.

Chapter 5

Summary and Conclusions

In this study, we searched for pair produced vector-like T quarks decaying to a dilepton and multi-jet final state at $\sqrt{s}$ = 13 TeV. We assumed T$\rightarrow$tZ with 100% branching ratio. Instead of doing a traditional 'cut and count' analysis, we decided to exploit the properties of a boosted topology to design a novel χ^2 algorithm. This algorithm effectively picks out the best possible candidates for the decay products of the T quarks in order to do a mass reconstruction for the T quark. We divided the χ^2 algorithm into several categories, each of which contained at least one boosted top jet or W/Z jet. The background events were found to have higher values of χ^2 and therefore, a selection on the lower side of the χ^2 variable helped to distinctly separate the signal shape from the background. The small drawback of this technique is that one misses out events where there are no boosted objects. For such events, computing a χ^2 variable can be computationally very expensive, as one has to first reconstruct top jets or W/Z jets from several possible combinations of AK4 jets. However, as the mass of the prospective new particle increases, it is likely that more and more events will start falling in the boosted categories and the fraction of events which have no boosted objects will reduce significantly. So for sufficiently boosted topologies, this technique is quite effective. Another comment that we would like to make here, is that as the probable mass of the T quark increases, it is more and more unlikely that the T quarks will be pair produced in pp collisions because the LHC is not going to see any energy upgrade in the future and all the data taking will continue at $\sqrt{s}$ = 13 TeV. So for future studies which look exclusively at single production of vector-like quarks at very high mass ranges, the χ^2 variable should be easy to evaluate even in categories with no boosted objects because of the reduced amount of combinatorics.

We made sure that our MC simulations closely reproduced the distributions as measured in data.

We then went ahead and obtained the results for the χ^2 algorithm in the data (integrated luminosity of 35.9 fb^{-1}). We took into account several systematic uncertainties and calculated 95% CL expected and observed upper limits on the TT production cross-section using an Asymptotic CLs method. We did not observe any excess in the data and could exclude T masses below 1095 GeV with 95% CL.

By the end of 2018, LHC is supposed to have collected data with an integrated luminosity of 100 fb^{-1}. With more data, let's hope to discover some exciting new physics!

Bibliography

[1] Abdus Salam, J. C. Ward, Phys. Lett. 13, 168-171 (1964).

[2] S. L. Glashow, Nucl. Phys. 22 (1961) 579-588.

[3] Steven Weinberg, Phys. Rev. Lett. 19 (1967) 1264-1266.

[4] M. Kobayashi and T. Maskawa, Progress of Theoretical Physics, Vol. 49 (1973).

[5] G. Aad et al. (ATLAS Collaboration, CMS Collaboration), Phys. Rev. Lett. 114, 191803, arXiv:1503.07589.

[6] F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964), 321-323.

[7] P. W. Higgs, Phys. Rev. Lett. 13 (1964), 508-509.

[8] Wikipedia: Standard Model, https://en.wikipedia.org/wiki/Standard_Model

[9] C. L. Bennett et al., The Astrophysical Journal, Supplement 208 (2013), 20., arXiv:1212.5225.

[10] G. Hinshaw et al., The Astrophysical Journal, Supplement 208 (2013), 19. arXiv:1212.5226.

[11] P. Ade et al. (Planck Collaboration), A&A 594, A13 (2016), arXiv:1502.01589.

[12] S.M. Bilenky, Phys. Scripta T121 (2005) 17-22, arXiv:hep-ph/0410090

[13] S. P. Martin, Adv. Ser. Direct. High Energy Phys.21 (2010), 1-153, arXiv:hep-ph/9709356.

[14] S.F. King, Rept.Prog.Phys. 67 (2004) 107-158, arXiv:hep-ph/0310204

[15] M. Shifman, Int.J.Mod.Phys.A25:199-225, 2010, arXiv:0907.3074.

[16] Particle Data Group collaboration, Chin. Phys. C, Vol. 40, 100001 (2016)

[17] D. Marzocca, M. Serone and J. Shu, JHEP 08 (2012) 013.

[18] L. Randall and R. Sundrum, Phys. Rev. Lett. 83 (1999) 3370-3373, arXiv:hep-ph/9905221.

[19] J. Kang, P. Langacker, and B. Nelson, Phys. Rev. D 77 (2008) 035003, arXiv:0708.2701v2.

[20] N. Arkani-Hamed, A. Cohen, E. Katz, and A. Nelson, JHEP 0207 (2002) 034, arXiv:hep-ph/0206021v2.

[21] M. Schmaltz, Nucl. Phys. Proc. Suppl. 117 (2003) 40, arXiv:hep-ph/0210415

[22] M. Schmaltz and D. Tucker-Smith, Ann. Rev. Nucl. Part. Sci 55 (2005) 229-270, arXiv:hep-ph/0502182v1.

[23] ATLAS Collaboration, JHEP 08 (2015) 105, arXiv:1505.04306

[24] ATLAS Collaboration, JHEP 11 (2014) 104, arXiv:1409.5500

[25] ATLAS Collaboration, Phys. Rev. D 91 (2015) 112011, arXiv:1503.05425

[26] ATLAS Collaboration, Phys. Lett. B 718 (2013) 1284, arXiv:1210.5468

[27] CMS Collaboration, Phys. Rev. Lett. 107 (2011) 271802, arXiv:1109.4985

[28] CMS Collaboration, Phys. Lett. B 729 (2014) 149, arXiv:1311.7667

[29] CMS Collaboration, Phys. Rev. D 93 (2016) 112009, arXiv:1507.07129

[30] CMS Collaboration, CMS-PAS-B2G-16-001, arXiv:1701.07409

[31] CMS Collaboration, Phys. Rev. D 93, 012003 (2016), arXiv:1509.04177

[32] R. Davis (CMS Collaboration), Aug 2016, CMS-OUTREACH-2016-027, https://cds.cern.ch/record/2205172

[33] CMS Collaboration, JINST 3 (2008) S08004.

[34] F. Maltoni and T. Stelzer, JHEP 02 (2003) 027, arXiv:hep-ph/0208156.

[35] NNPDF Collaboration, JHEP 04 (2015) 040, arXiv:1410.8849.

[36] M. B. Czakon, P. Fiedler, and A. Mitov, Phys.Rev.Lett. 110 (2013), no. 25, 252004, arXiv:1303.6254.

[37] J. Alwall et. al., JHEP07 (2014) 079, arXiv:1405.0301v2.

[38] S. Kallweit et. al., J. High Energ. Phys. (2016), arXiv:1511.08692.

[39] J. Baglio, Phys. Rev. D 93, 054010 (2016), arXiv:1512.05787.

[40] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. 178 (2008) 852-867, arXiv:0710.3820.

[41] T. Gehrmann et. al., Phys. Rev. Lett. 113 (2014) 212001, arXiv:1408.5243v1.

[42] F. C. et. al., PLB 735 (2014) 311, arXiv:1405.2219v2.

[43] R. K. E. J. M. Campbell and C. Williams, JHEP 018 (2011) 1107, arXiv:1105.0020.

[44] Khachatryan, V., Sirunyan, A.M., Tumasyan, A. et al., Eur. Phys. J. C (2016) 76: 155.

[45] J. Allison et al., IEEE Transactions on Nuclear Science 53 (Feb, 2006) 270-278.

[46] CMS Collaboration, CMS Physics Analysis Summary, CMS-PAS-PFT-09-001, CERN, 2009.

[47] CMS Collaboration, CMS Physics Analysis Summary, CMS-PAS-PFT-10-002, CERN, 2010.

[48] CMS Collaboration, CMS Physics Analysis Summary, CMS-PAS-PFT-10-003, CERN, 2010.

[49] CMS Collaboration, CMS Note, CMS-NOTE-2005-001, CERN, 2005

[50] M. Cacciari, G. P. Salam, and G. Soyez, JHEP 04 (2008) 063, arXiv:0802.1189.

[51] S. Khalil, HEP Seminar, SLAC (USA), https://indico.cern.ch/event/622651/

[52] J. Thaler and K. Van Tilburg, JHEP 1103:015 (2011), arXiv:1011.2268

[53] S. D. Ellis, C. K. Vermilion, J. R. Walsh, Phys. Rev. D 81:094023 (2010), arXiv:0912.0033

[54] A. J. Larkoski, S. Marzani, G. Soyez, J. Thaler, JHEP 1405 (2014) 146, arXiv:1402.2657

[55] M. Dasgupta et. al., JHEP09 (2013) 029, arXiv:1307.0007

[56] J.Nielsen, SCIPP 11/06, arXiv:1106.2516.

[57] CMS Collaboration, CMS Physics Analysis Summary, CMS-PAS-BTV-15-001.

[58] ROOT Data Analysis Framework, https://root.cern.ch/documentation

[59] CMS Collaboration, CMS Physics Analysis Summary, CMS-PAS-LUM-17-001.

[60] CMS Collaboration, Phys. Rev. Lett. 116, 052002 (2016), arXiv:1510.05302

[61] CMS Collaboration, Phys. Lett. B 766 (2017) 268, arXiv:1607.06943

[62] CMS Collaboration, Phys. Lett. B 763 (2016) 280, arXiv:1607.08834

[63] The ATLAS and CMS Collaborations and The LHC Higgs Combination Group, CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11

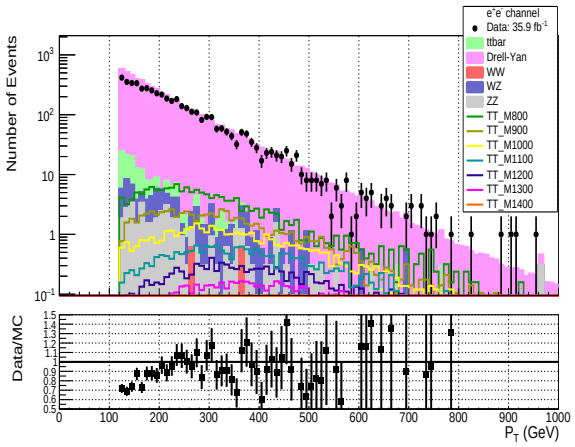
Appendix A

Data vs MC Comparisons

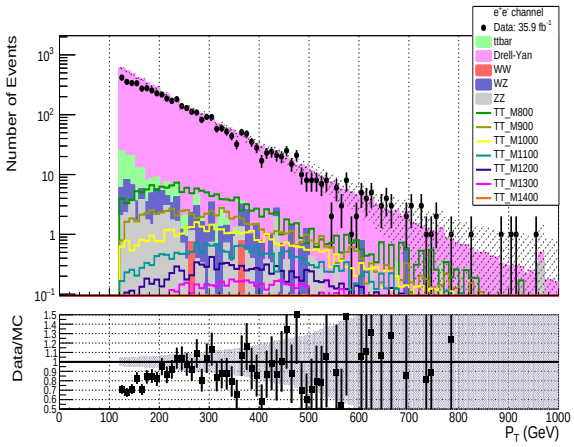
A.1 Electron Channel

In section 3.4.1, we didn't look at the effect of just the trigger and the lepton identification and isolation scale factors without the affect of the Z p_T dependent electroweak correction factors for the Drell-Yan sample. We will show the affect of applying the trigger scale factor and the lepton identification and isolation scale factors to the MC here and see how the comparison looks between the data and the MC. We use the same pre-selections as mentioned in Table 3.10. The plots are shown in Fig. A.1. There is not much overall difference to the distributions if one looks at the Data/MC ratios. We use the same systematic uncertainties as mentioned in Table 3.6 and the percentage of uncertainties considered are the same as given in Table 3.7.

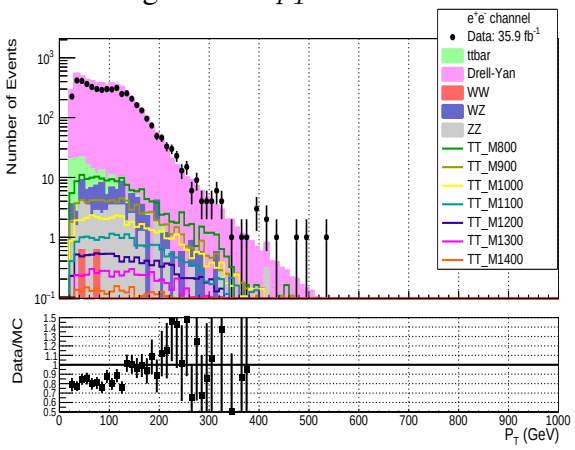
The Z p_T dependent electroweak correction factors to the Drell-Yan sample really make a difference to the MC. As Z+jets is the dominant background, the Data/MC ratio is largely affected by the contribution from the Z+jets sample. We will now show several comparisons just before and after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample (Fig. A.2, Fig. A.3 and Fig. A.4).



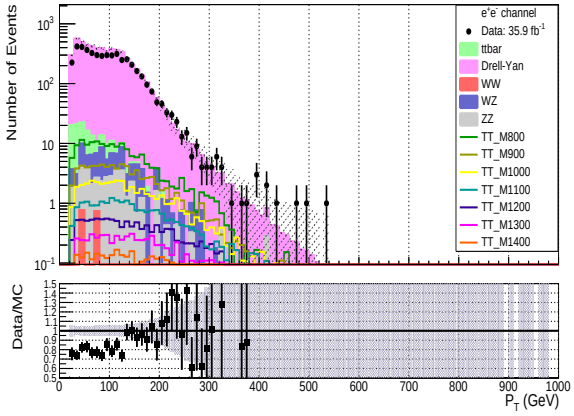
Leading electron p_T distribution before.



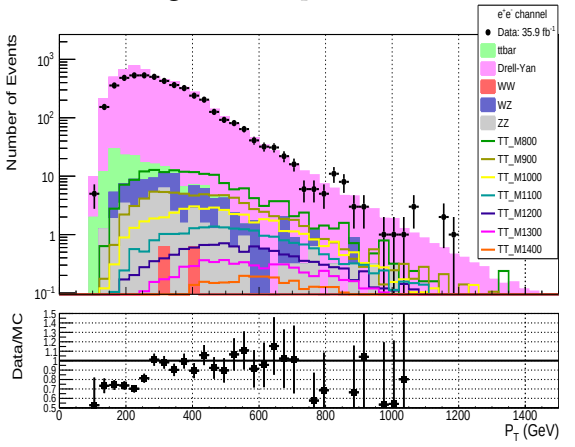
Leading electron p_T distribution after.



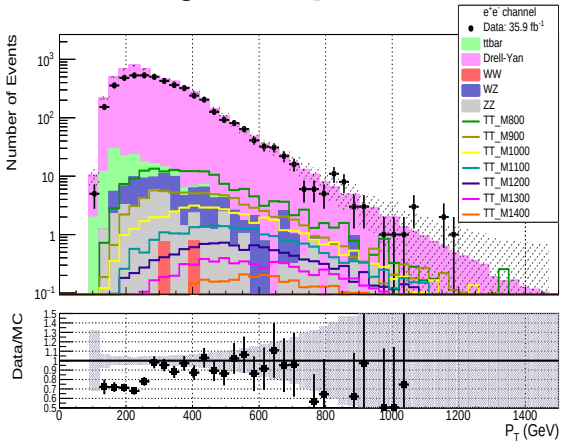
Sub-leading electron p_T distribution before.



Sub-leading electron p_T distribution after.

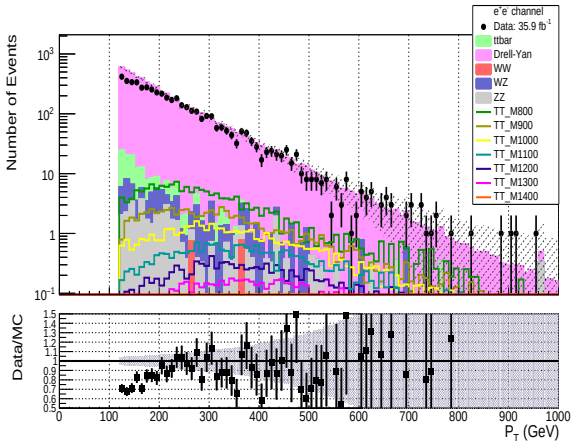


Dielectron p_T distribution before.

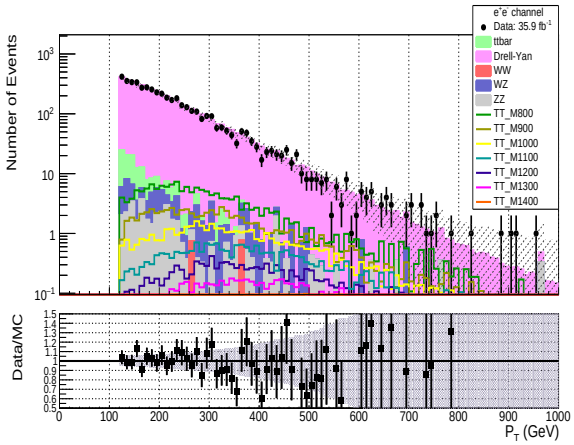


Dielectron p_T distribution after.

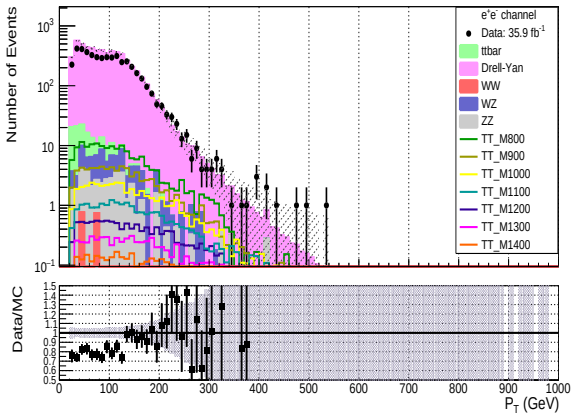
Figure A.1: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the electron identification and isolation scale factors and the trigger scale factors. The error band shows both statistical uncertainty as well as the systematic uncertainty on the two scale factors applied, the uncertainty on the integrated luminosity and the uncertainty on the theoretical cross section of the backround samples.



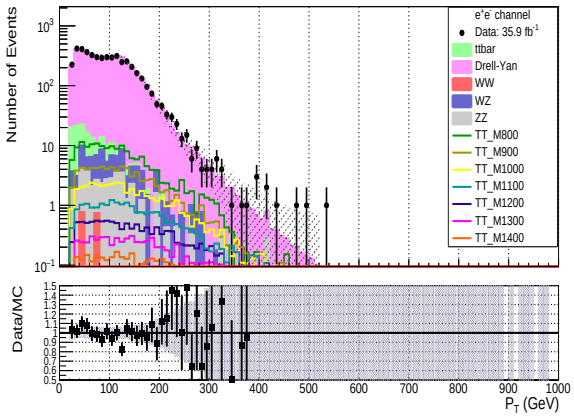
Leading electron p_T distribution before.



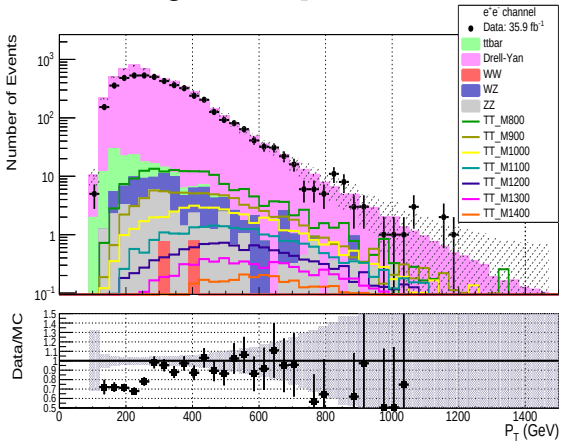
Leading electron p_T distribution after.



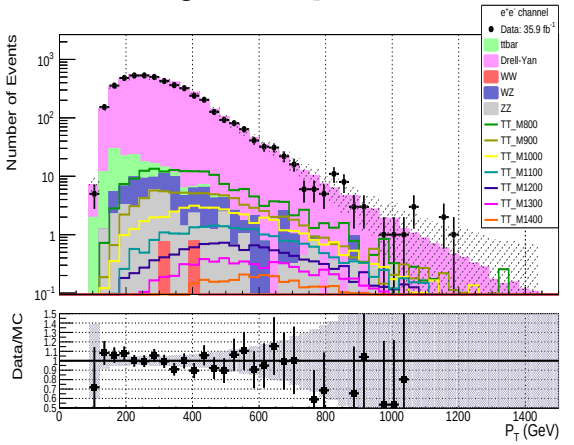
Sub-leading electron p_T distribution before.



Sub-leading electron p_T distribution after.

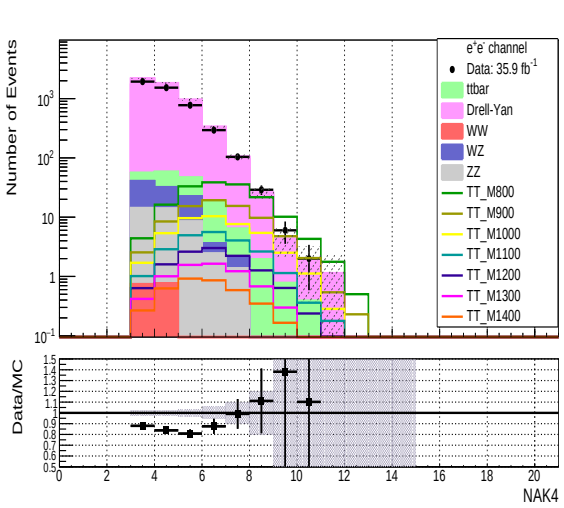


Dielectron p_T distribution before.

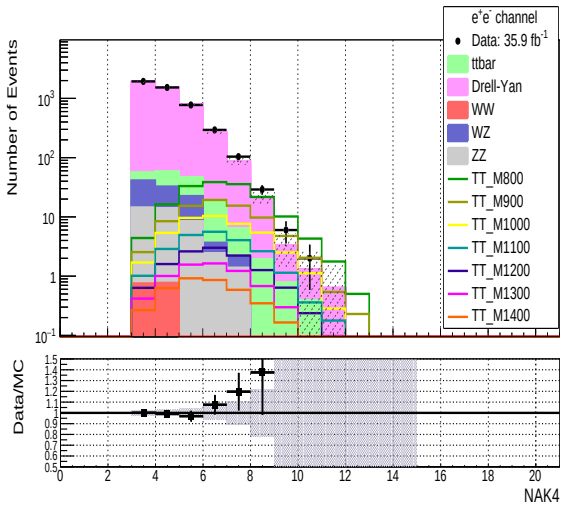


Dielectron p_T distribution after.

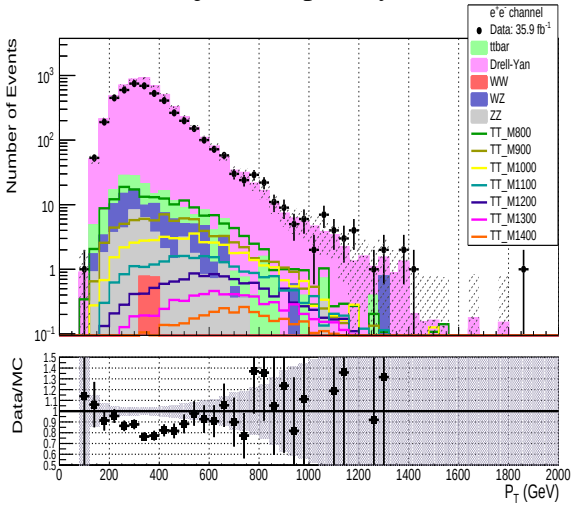
Figure A.2: (Left) The plots show the distribution with just the electron identification and trigger scale factors applied. (Right) The same plots after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample.



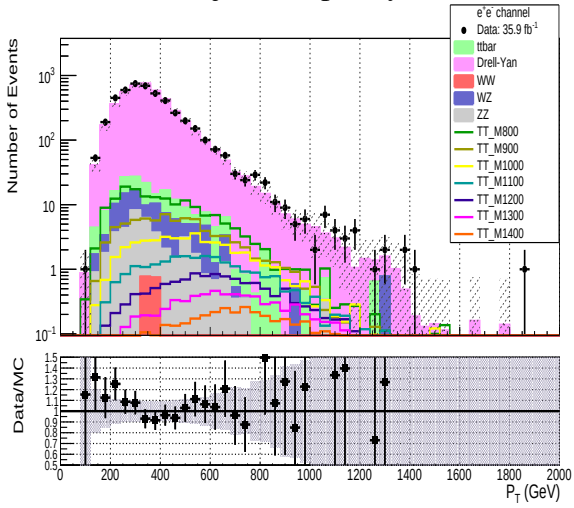
AK4 jet multiplitcity before.



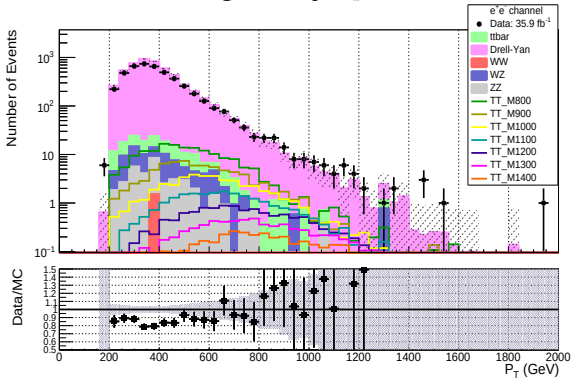
AK4 jet multiplicity after.



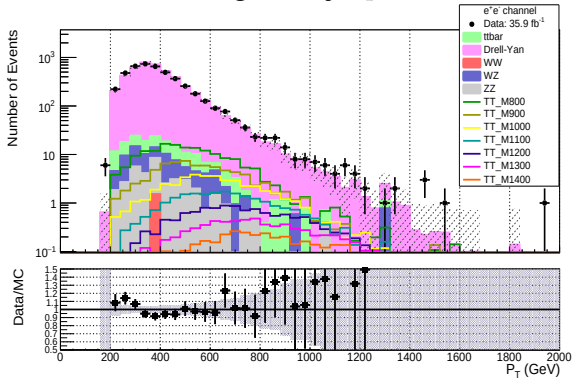
Leading AK4 jet p_T before.



Leading AK4 jet p_T after.

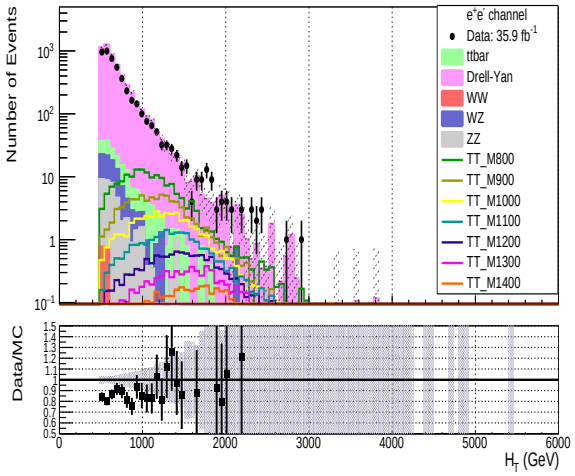


Leading AK8 jet p_T before.

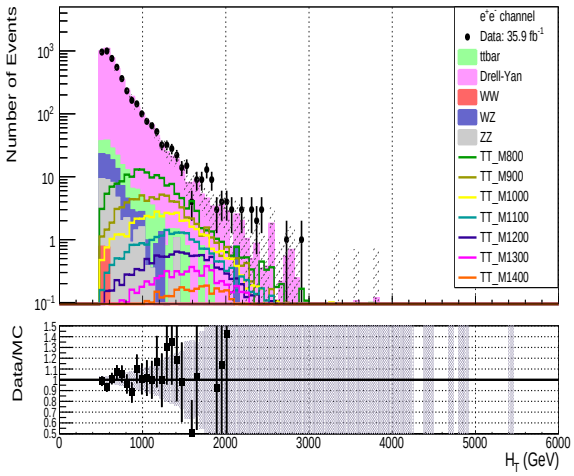


Leading AK8 jet p_T after.

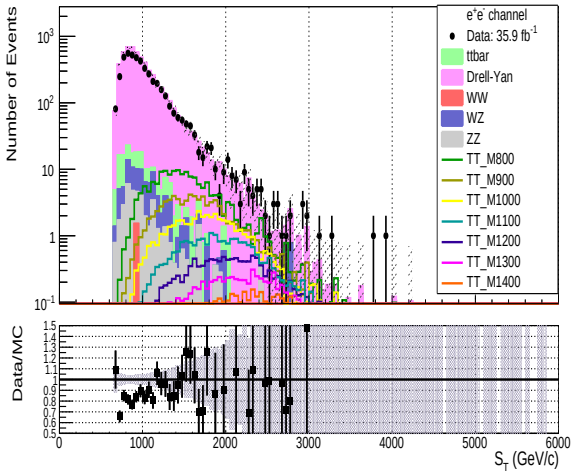
Figure A.3: (Left) The plots show the distribution with just the electron identification and trigger scale factors applied. (Right) The same plots after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample.



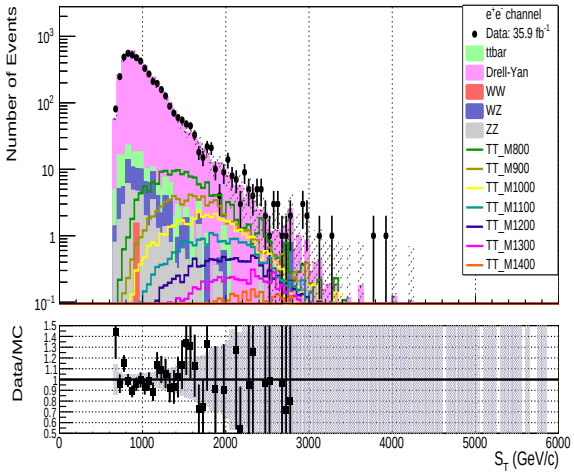
H_T distribution before.



H_T distribution after.



S_T distribution before.



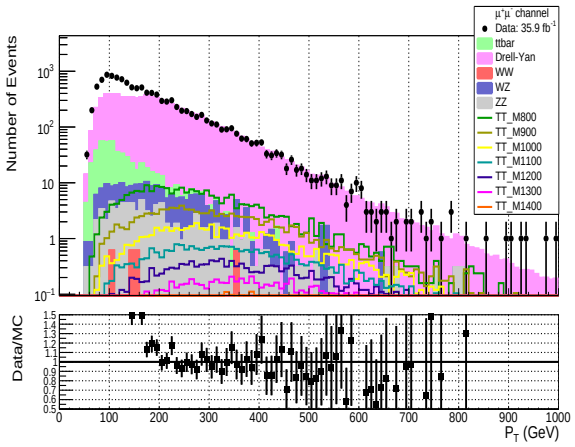
S_T distribution after.

Figure A.4: (Left) The plots show the distribution with just the electron identification and trigger scale factors applied. (Right) The same plots after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample.

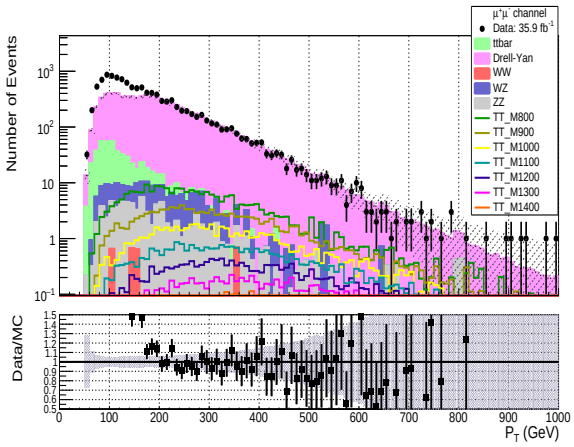
A.2 Muon Channel

In section 3.4.2, we didn't look at the effect of just the trigger and the lepton identification and isolation scale factors without the affect of the Z p_T dependent electroweak correction factors for the Drell-Yan sample. We will show the affect of applying the trigger scale factor and the lepton identification and isolation scale factors to the MC here and see how the comparison looks between the data and the MC. We use the same pre-selections as mentioned in Table 3.10. The comparison plots are shown in Fig. A.5. These scale factors however don't make a great overall difference to the distributions if one looks at the Data/MC ratios. We use the same systematic uncertainties as mentioned in Table 3.6 and the percentage of uncertainties considered are the same as given in Table 3.7.

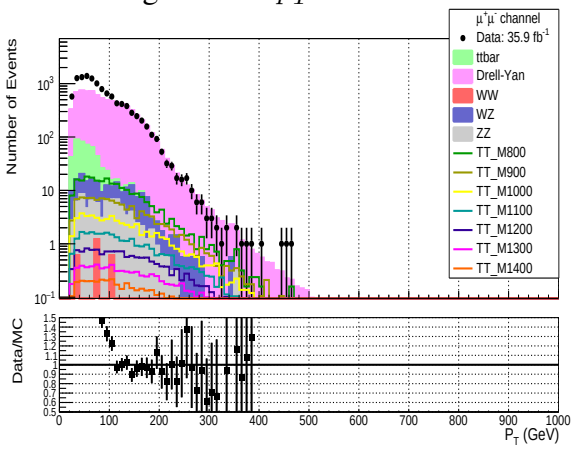
The Z p_T dependent electroweak correction factors to the Drell-Yan sample really make a difference to the MC. As Z+jets is the dominant background, the Data/MC ratio is largely affected by the contribution from the Z+jets sample. We will now show several comparisons just before and after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample (Fig. A.6, Fig. A.7 and Fig. A.8).



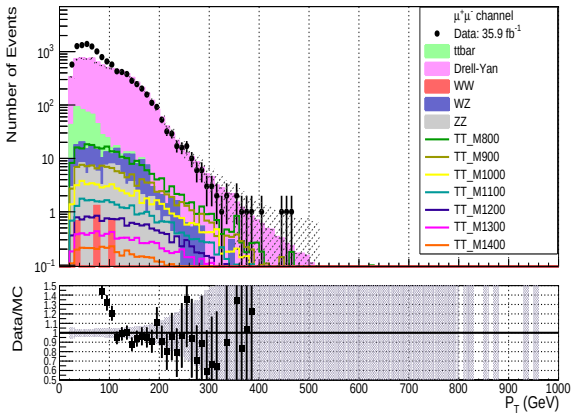
Leading electron p_T distribution before.



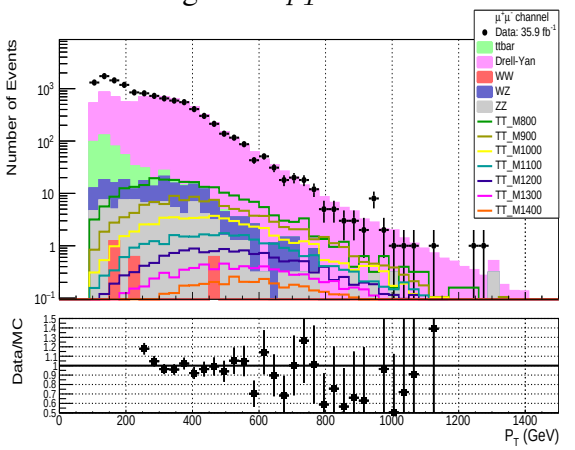
Leading muon p_T distribution after.



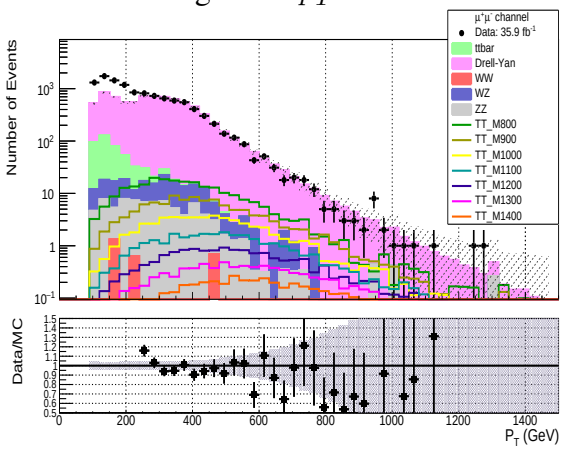
Sub-leading muon p_T distribution before.



Sub-leading muon p_T distribution after.

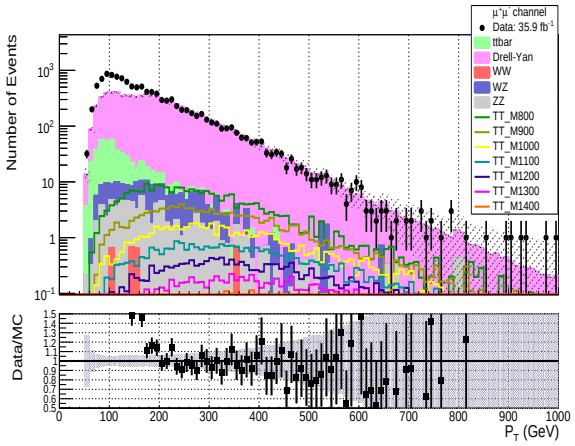


Dimuon p_T distribution before.

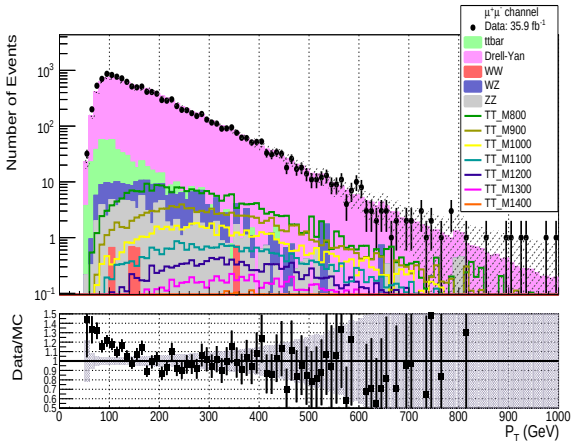


Dimuon p_T distribution after.

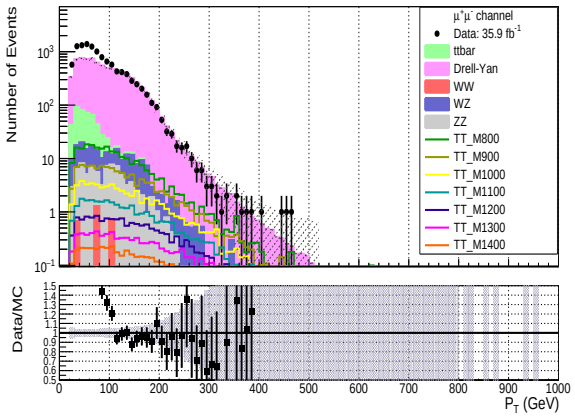
Figure A.5: (Left) Comparison between the Data and the MC before applying any scale factors. The error bars only reflect the statistical uncertainty. (Right) The Data and MC comparisons after applying the muon identification and isolation scale factors and the trigger scale factors. The error band shows both statistical uncertainty as well as the systematic uncertainty on the two scale factors applied, the uncertainty on the integrated luminosity and the uncertainty on the theoretical cross section of the backround samples.



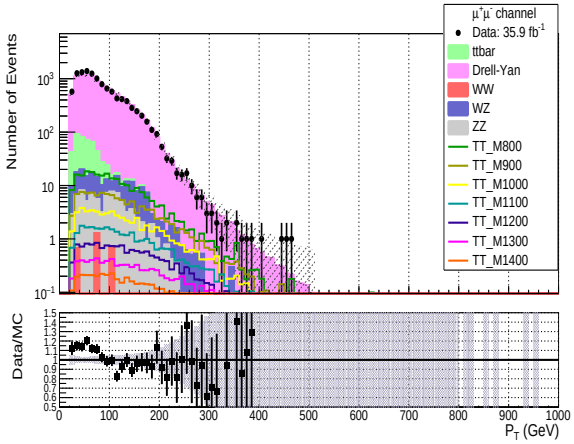
Leading electron p_T distribution before.



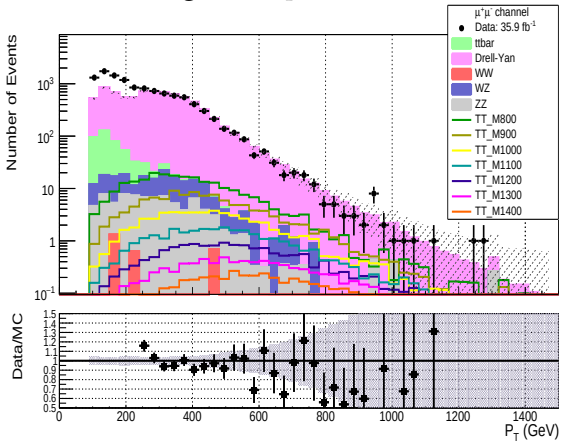
Leading muon p_T distribution after.



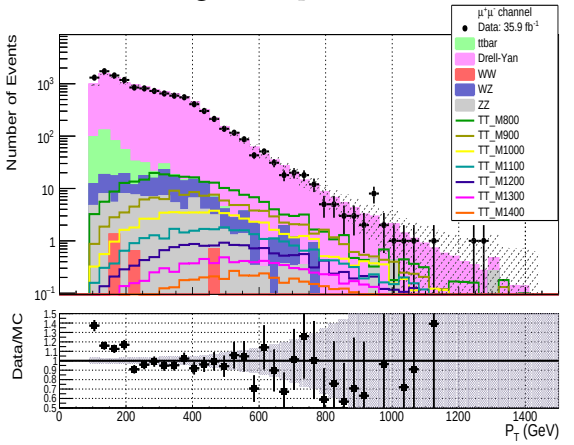
Sub-leading muon p_T distribution before.



Sub-leading muon p_T distribution after.

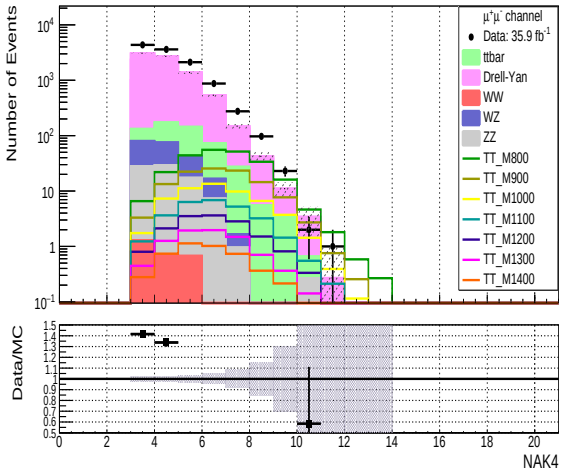


Dimuon p_T distribution before.

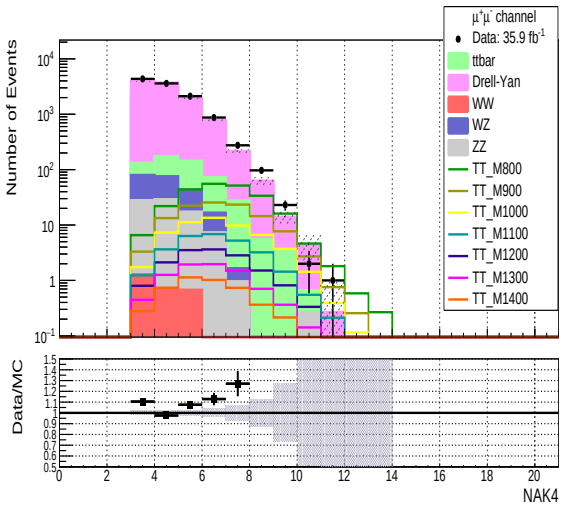


Dimuon p_T distribution after.

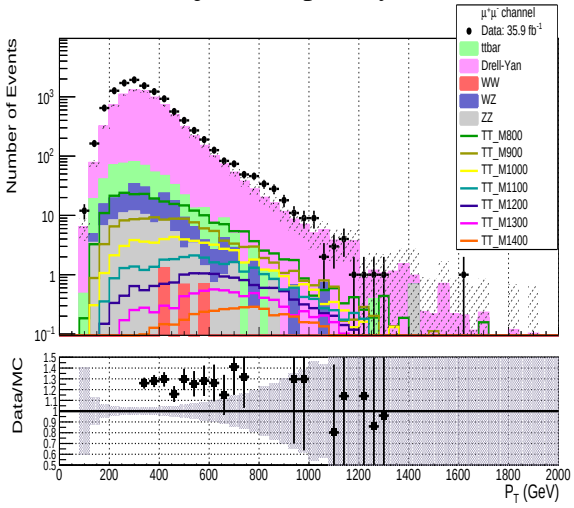
Figure A.6: (Left) The plots show the distribution with just the muon identification and trigger scale factors applied. (Right) The same plots after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample.



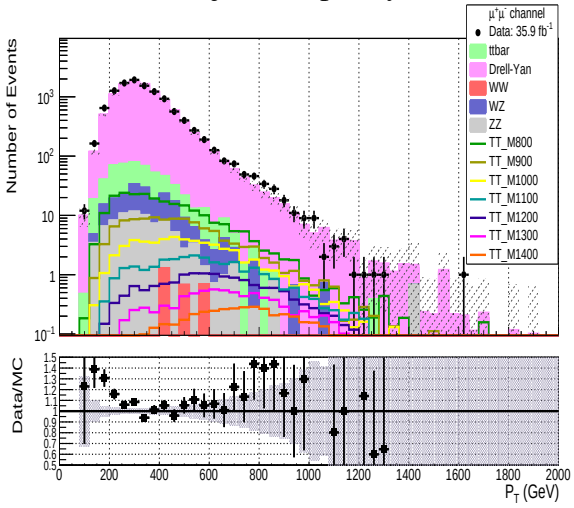
AK4 jet multiplitcity before.



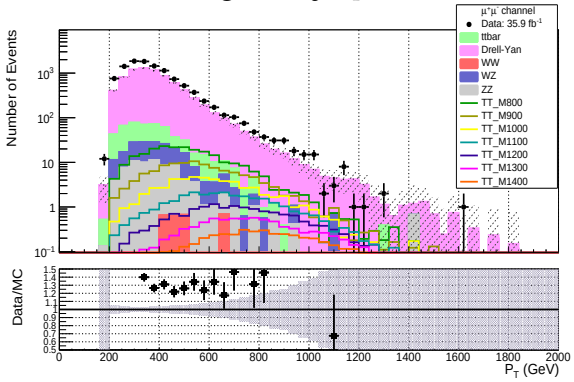
AK4 jet multiplicity after.



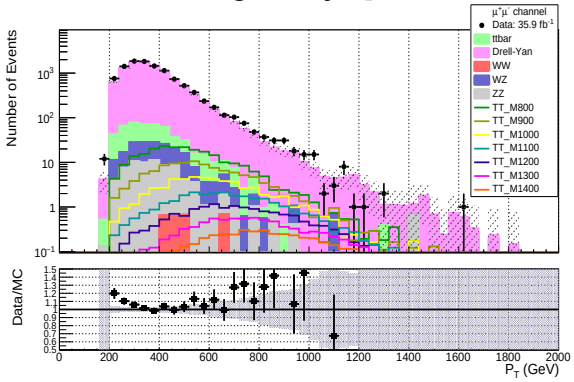
Leading AK4 jet p_T before.



Leading AK4 jet p_T after.

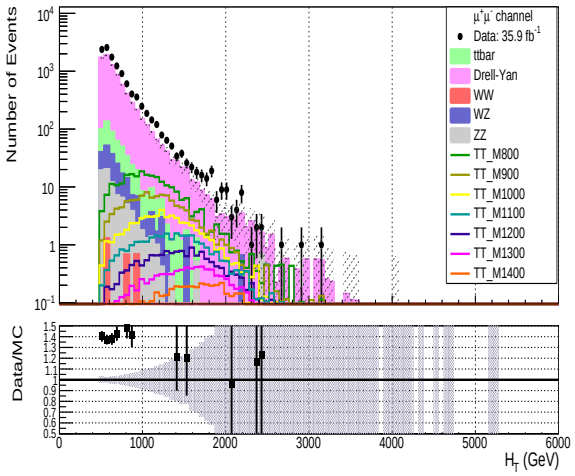


Leading AK8 jet p_T before.

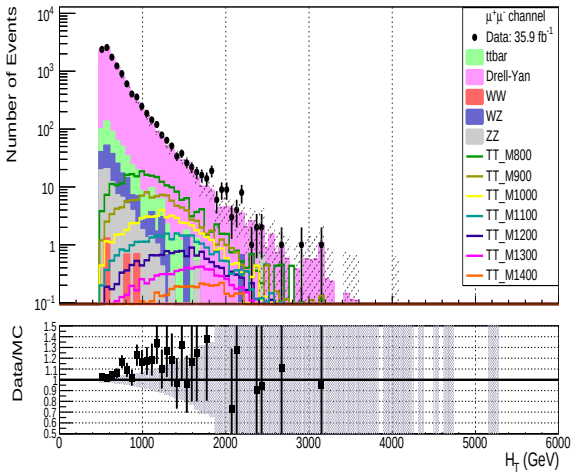


Leading AK8 jet p_T after.

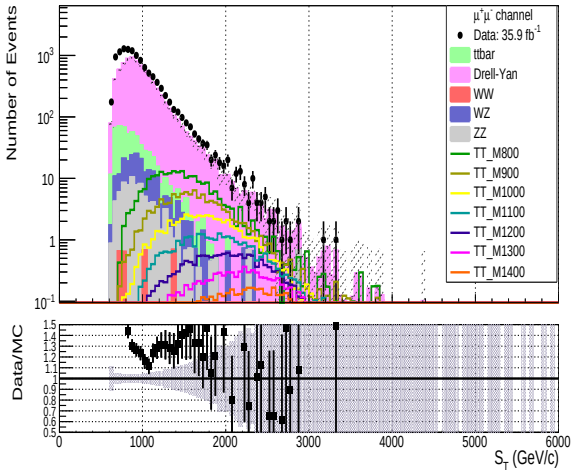
Figure A.7: (Left) The plots show the distribution with just the muon identification and trigger scale factors applied. (Right) The same plots after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample.



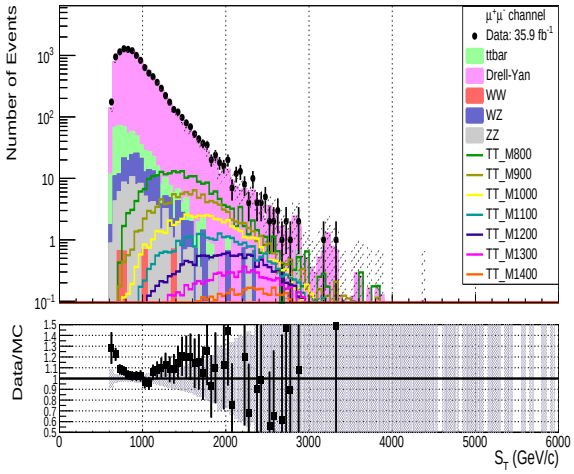
H_T distribution before.



H_T distribution after.



S_T distribution before.



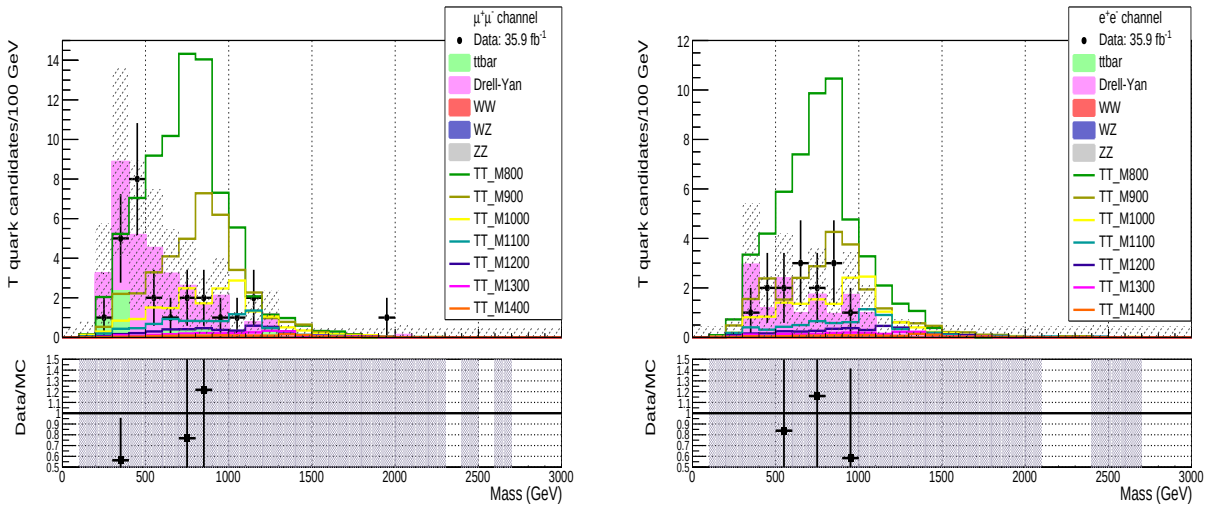
S_T distribution after.

Figure A.8: (Left) The plots show the distribution with just the muon identification and trigger scale factors applied. (Right) The same plots after applying the Z p_T dependent electroweak correction factors to the Drell-Yan sample.

Appendix B

Mass distributions for calculating expected and observed limits on the T mass

The mass distributions (obtained from the χ^2 algorithm for $\chi^2 < 20$) which were used to calculate the expected and observed limits in Sec. 4.3 are given in Fig. B.1. These distributions were used because of their finer binning (as compared to Fig. 4.3) which relays better information about the shape of the distribution and also helps in placing better limits on the T mass. The mass distribution for the combined electron and muon channel is shown in Fig. B.2.



Muon channel Electron channel

Figure B.1: T mass distribution for $\chi^2 < 20$.

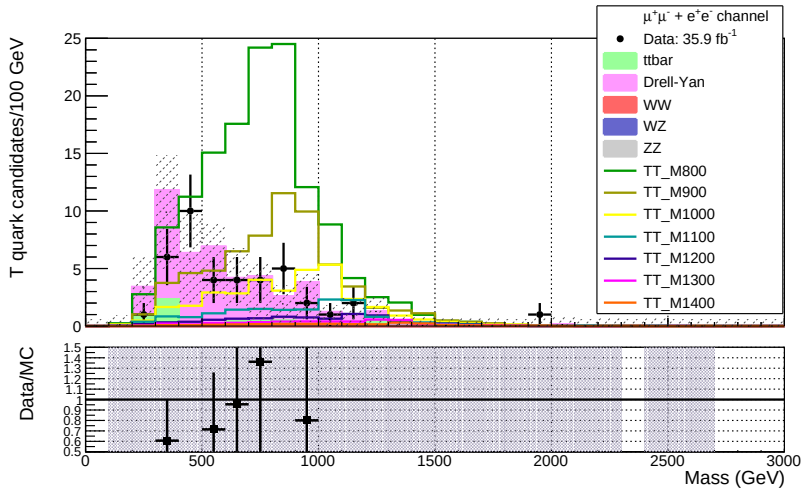


Figure B.2: T mass distribution for $\chi^2 < 20$ in the combined e^+e^-+ $\mu^+\mu^-$channel.