

Chapter 1

Introduction

1.0.1 The Standard Model

A physics model a description of a system using physics and mathematical concepts and language. Standard Model (SM) is a summarization of what is known in the subatomic world. For particle s. .

1.0.1.1 Formation of jets

1.0.1.2 W, Z and Higgs

1.0.2 The physics beyond Standard Model

Chapter 2

The CMS detector at the LHC

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. Inside the accelerator, two high-energy particle beams travel at close to the speed of light before they are made to collide. The beams travel in opposite directions in separate beam pipes —two tubes kept at ultrahigh vacuum. They are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets. The beams inside the LHC are made to collide at four locations around the accelerator ring, corresponding to the positions of four particle detectors ATLAS, CMS, ALICE and LHCb.

The focus of this chapter is to present a brief overview of the Compact Muon Solenoid (CMS) detector. The overall layout of CMS from different **aspects** are shown in Figure ?? and Figure ?. The dimensions of the CMS detectors are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 tons. In the following, we

CHAPTER 2. THE CMS DETECTOR AT THE LHC

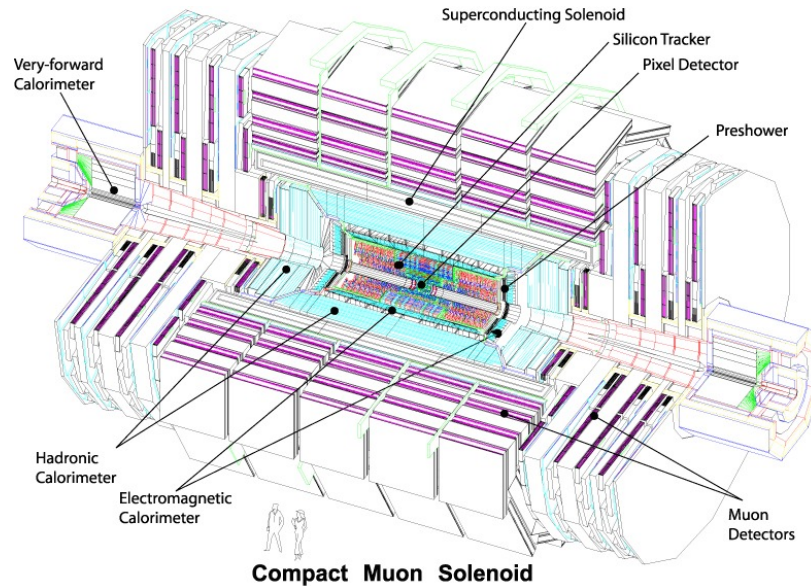


Figure 2.1: An exploded view of the CMS detector.

will first introduce the coordinates and then we will introduce the components of the CMS detector.

How to define the angles, $\eta\phi$ Guofan's thesis has a good explanation of the angles. GOOD! What kind of information do I want to show in the detector section. A detailed one or a brief one, how brief.

CHAPTER 2. THE CMS DETECTOR AT THE LHC

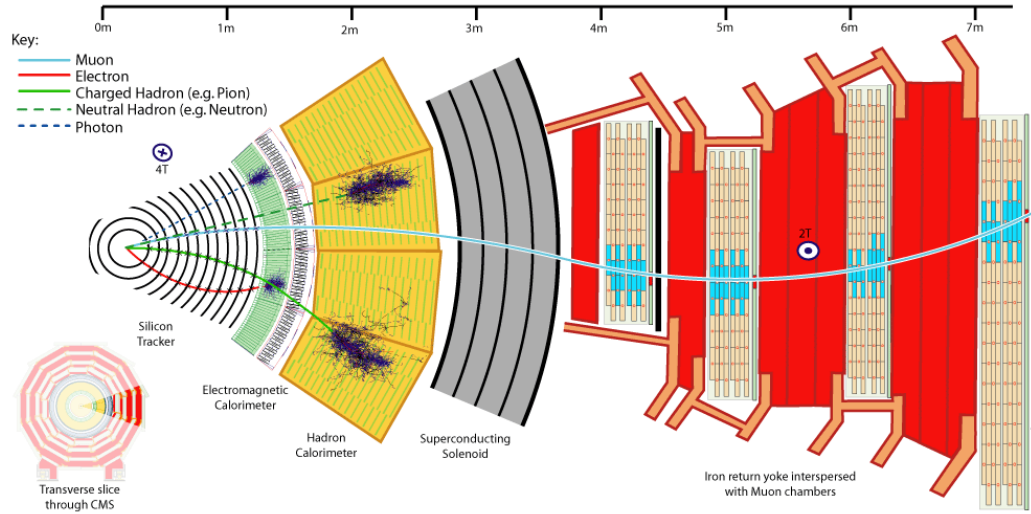


Figure 2.2: Transverse picture of the CMS detector. [Change the picture to pdf format](#)

2.1 The beam: luminosity and cross-section

2.2 The coordinates

2.3 The Magnet

In order to achieve good momentum resolution within a compact spectrometer without making stringent demands on muon-chamber resolution and alignment, a high magnetic field was chosen. The detailed parameters of this are shown in Table ??.

The bore of the magnet coil is also large enough to accommodate the inner tracker and the calorimetry inside.

CHAPTER 2. THE CMS DETECTOR AT THE LHC

Table 2.1: Parameters of the CMS superconducting solenoid.

Field	4 T
Inner Bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA
Store energy	2.7 GJ
Hoop stress	64 atm

2.4 The inner tracking system

Here we need a plot of the tracker layout and also the pixel tracker. If we could find a good description of the strip detector.

The tracking volume is give by a cylinder of length 5.8 m and diameter 2.6 m. In order to deal with high track multiplicities, CMS employs 10 layers of silicon microstrip detectors, which provide the required granularity and precision. In addition, 3 layers of silicon pixel detectors are placed closed to the interaction region to improve the measurement of impact parameter of charged-particle tracks, as well as the position of secondary vertices. The detailed layout of the tracker system is shown in Figure ??

Close to the interaction vertex, in the barrel region, are 3 layers hybrid pixel detectors at a radii of 4.4, 7.3 and 10.2 cm. Each layer is spilt into segments like tiny

CHAPTER 2. THE CMS DETECTOR AT THE LHC

kitchen tiles, each a little silicon sensor, $100\ \mu m$ by $150\ \mu m$, about two hairs widths. Knowing which pixels have been touched allows us to deduce the particle's trajectory. And because the detector is made of 2D tiles, rather than strips, and has a number of layers, we can create a three-dimensional picture. The spatial resolution is measured to be $\approx 10\ \mu m$ for the $r - \phi$ measurement and $\approx 20\ \mu m$ for the z measurement. After the pixels and on their way out of the tracker, particles pass through ten layers of silicon strip detectors, reaching out to a radius of 130 cm. The tracker silicon strip detector consists of four inner barrel (TIB) layers assembled in shells with two inner endcaps (TID), each composed of three small discs. The outer barrel (TOB) consists of six concentric layers. Finally two endcaps (TEC) close off the tracker. Each has silicon modules designed differently for its place within the detector.

2.5 Electromagnetic calorimeter (ECAL)

The ECAL is designed to calibrate the energy of electron and photons resulting from pp collisions, and its structure is shown in Figure ?? ECAL uses lead tungstate ($PbWO_4$) crystals with coverage in $|\eta|$ up to 3.0. This crystal is highly transparent and “scintillates” when electrons and photons pass through it, which produces light in proportion to the particles energy. Although lead tungstate has short radiation (0.89 cm) and Moliere (2.2 cm) lengths, are also fast and radiation hard, it produces relatively low light yield. So the silicon avalanche photodiodes (APDs) are used as

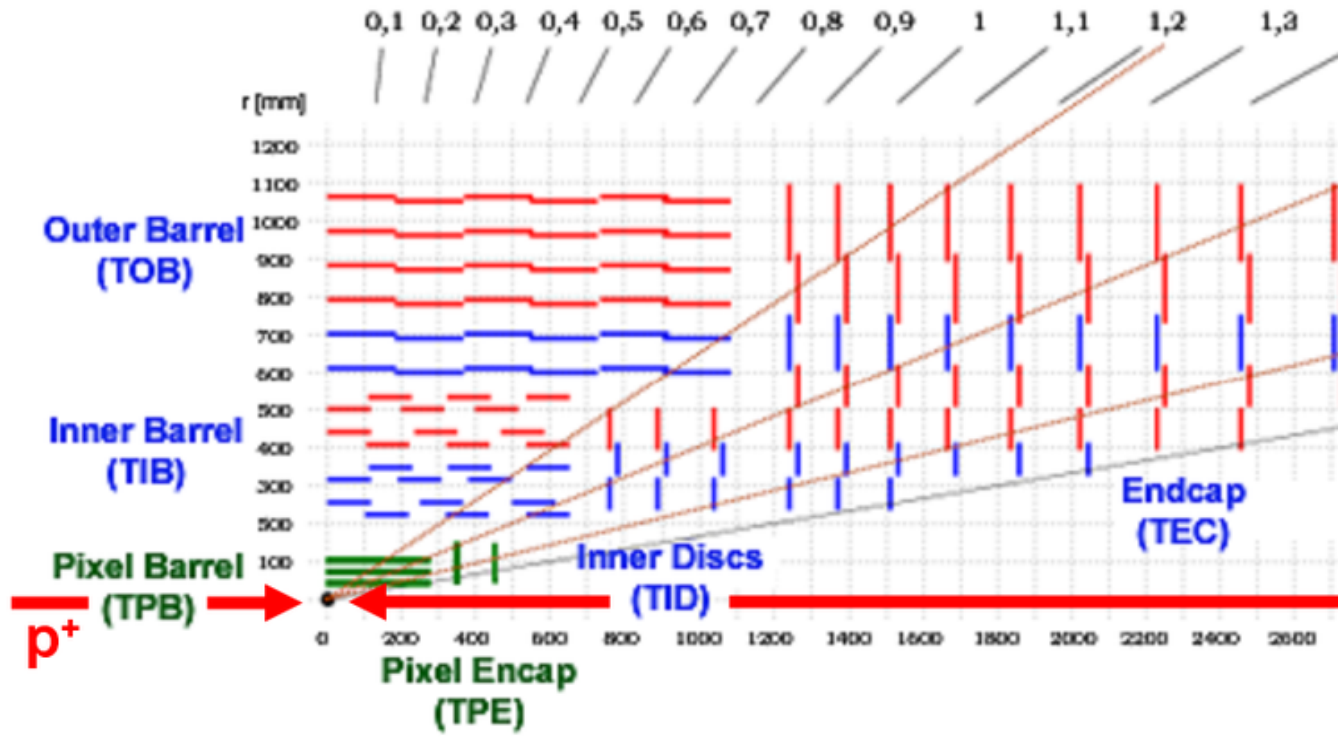


Figure 2.3: The tracker layout of CMS

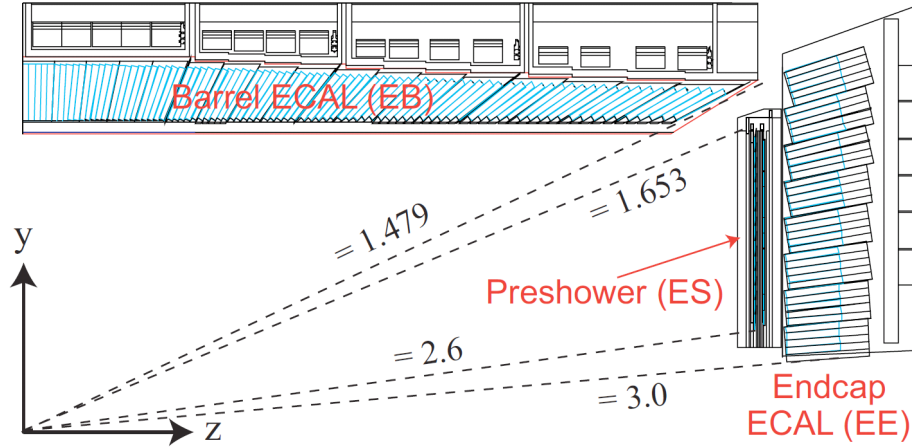


Figure 2.4: Geometric view of one quarter of the ECAL.

photodetectors in the ECAL barrel (EB) and vacuum phototriodes (VPTs) in the ECAL endcap (EE).

These photodetectors that have been especially designed to work within the high magnetic field, are also glued onto the back of each of the crystals to detect the scintillation light and convert it to an electrical signal that is amplified and sent for analysis.

A preshower system is installed in the front of EE. The preshower is made of two planes of lead followed by silicon sensors, similar to those used in the tracker. The reason for the preshower system is that short-lived particles called neutral pions, also produced in collisions, can inadvertently mimic high-energy photons when they decay into two closely-spaced lower energy photons that the ECAL picks up together. And for Higgs discovery, the high energy photons from Higgs decay is the important signature of $H \rightarrow \gamma\gamma$ channel.

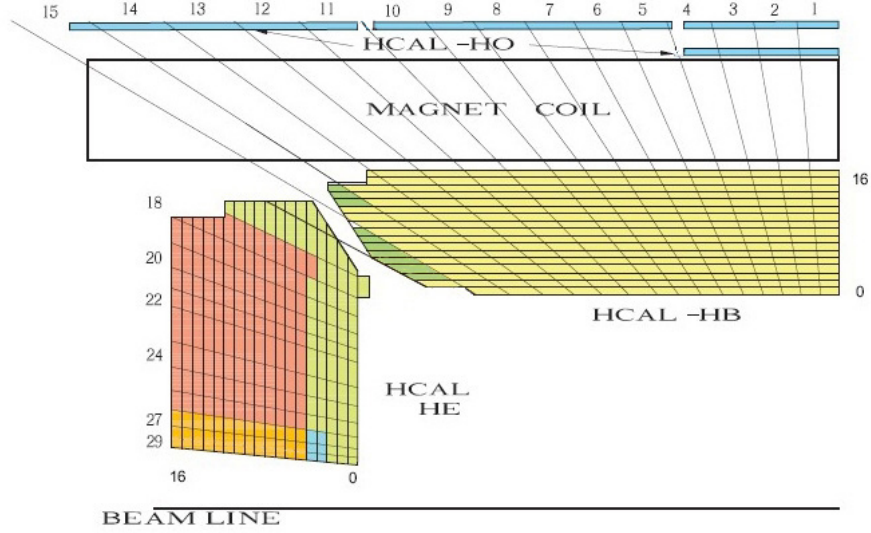


Figure 2.5: Geometric view of one quarter of the HCAL. find a plot has HF

2.6 Hadronic calorimeter (HCAL)

HCAL is designed to measure the energy of particles other than electron and photons, for example proton, neutron, pion and kaon. It is also provide good containment and hermitically for the transverse missing energy E_T^{miss} measurement, resulting from neutrinos or possible new particles.

As shown in Figure ??, HCAL is composed by four parts, HCAL barrel (HB), HCAL endocap (HE), HCAL outer (HO) and HCAL forward(HF). HF is not presented in the plot. However, HF sits on the outside of the magnet and covers $3 < |\eta| < 5.0$.

HCAL is a brass/scintillator sampling hadron calorimeter.

2.7 Muon system

CMS is called the "compact muon solenoid". Muon detecting at CMS is very important.

Centrally produced muons are measured 3 times: in the inner tracker, after the coil, and in the return flux. Measurement of the momentum of muons using only the muon system is essentially determined by the muon bending angle at the exit of the 4 T coil, taking the interaction point of pp collision as the origin of the muon. For low-momentum muons, the best momentum resolution is given by the resolution obtained in the silicon tracker. For high-momentum muons, combining the inner tracker and muon detector measurements will highly improve the muon momentum resolution. At CMS, in $0 < |\eta| < 2.0$, for μ with p_T 200 400 GeV, $\delta p/p$ is measured to be $\leq 3\%$.

The layout of the one quarter of the CMS muon system for the initial low luminosity running is shown in Figure ?? and the transverse view of the muon stations (MS) are shown in Figure ?. In the Muon Barrel (MB) region, 4 stations of detectors are arranged in cylinders interleaved with the iron yoke.

Three types of gaseous detectors are used to identify and measure muons. [cite some paper](#). Drift tube (DT) is used in the barrel region ($|\eta| < 1.2$). In this region, the residual magnetic field in the chambers is low and also muon rate is low, so drift tube works well. The maximum drift length is 2.0 cm and the single point resolution is $\approx 200 \mu m$. In the endcap region, cathode strip chambers (CSC) are used because of the high residual magnetic field and also high muon rate. In addition to this, resistive

plate chambers (RPC) are used in both the barrel and the endcap regions. RPC could provide a fast response with good time resolution but with a coarser position resolutions than DTs and CSCs. So RPC could identify the correct bunch crossing.

2.8 Trigger and data acquisition

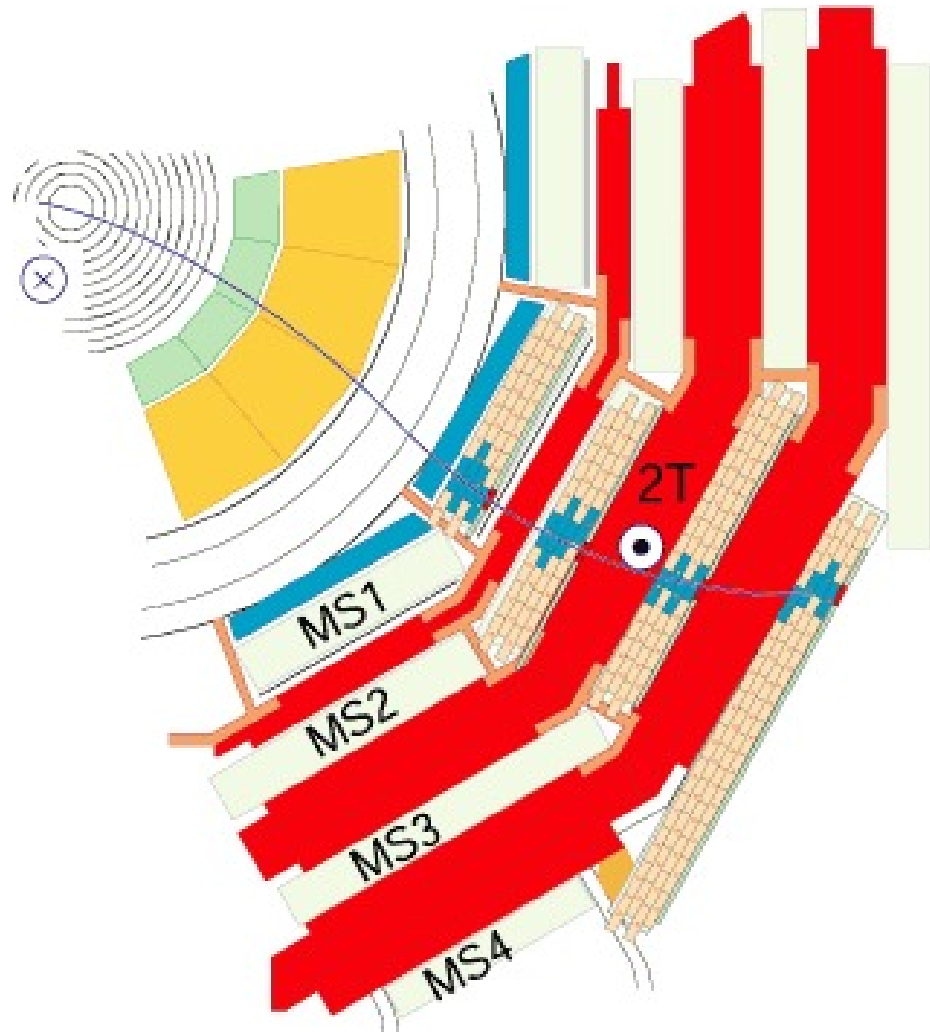


Figure 2.6: The Muon stations in the transverse view

CHAPTER 2. THE CMS DETECTOR AT THE LHC

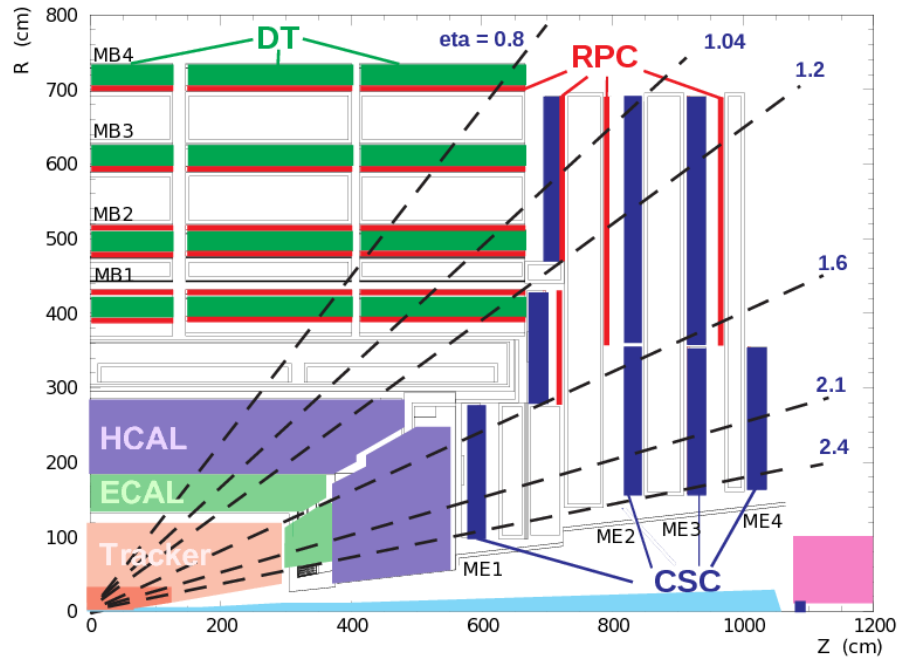


Figure 2.7: Layout of one quarter of the CMS muon system for initial low luminosity running. The RPC system is limited to $|\eta| < 1.6$ in the endcap, and for the CSC system only the inner ring of the ME4 chambers have been deployed. [reference the book or the website](#)

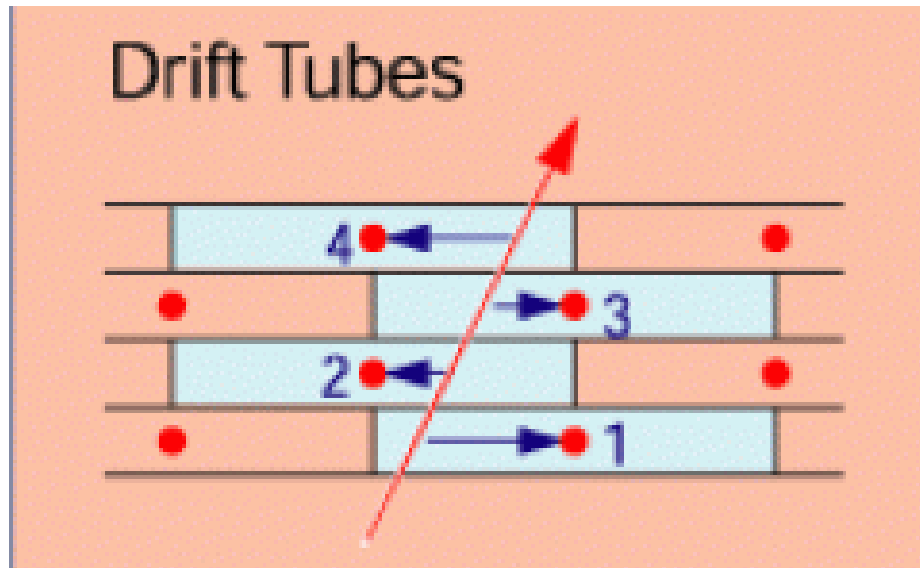


Figure 2.8: Layout of the drift tube.

Chapter 3

EXO-12-024

3.1 Introduction

As we mentioned in Chapter ??, SM is not perfect. And several scenarios of physics beyond the standard model predict the existence of resonances with masses above 1 TeV which decay into a quark jet and a W or Z boson, or into a pair of W or Z bosons. A vector boson emerging from such a decay is usually sufficiently boosted that the hadronization products from its daughter quarks merge into a single massive jet [?]. We present a search for these special dijet topologies, performed in the pp collision data collected by the CMS experiment in 2012 at a center-of-mass energy of 8 TeV.

The signal is characterized by a peak in the dijet invariant mass, emerging from the falling standard model background comprised mainly of QCD events with a dijet

topology. The analysis presented here enhances the sensitivity to processes with jets from W/Z bosons by the application of techniques that can identify W/Z -jets and suppress quark and gluon jets (“ W/Z -tagging”). This search is an update of a previous study [?] performed on the 7 TeV CMS data. Besides a larger dataset and an increased signal cross section due to a higher center-of-mass energy, this update also benefits from an improved W/Z -tagger based on “N-subjettiness” variables [?].

We consider four benchmark scenarios that would produce singly- or doubly-tagged events: an excited quark q^* [?] decaying into a quark and a W or Z boson; a Randall–Sundrum (RS) graviton G_{RS} decaying to WW or ZZ [?]; a Bulk graviton G_{Bulk} decaying to WW or ZZ [?, ?, ?]; and a heavy partner of the SM W boson W' which decays to WZ [?]. The most stringent limits on the q^* model have been set in dijet resonance searches at the LHC by considering inclusive all-hadronic final states [?, ?, ?]. The most stringent lower limit (at 95% CL) on the q^* mass to date is 3.47TeV [?]. Specific searches for the qW and qZ final states have previously been reported at the Tevatron [?, ?], which exclude resonances decaying to qW or qZ with masses up to 0.54 TeV, and at the LHC [?, ?], which extends the mass exclusion of $qW(qZ)$ to 2.38TeV(2.15TeV). The WW and ZZ final states have also been explored experimentally [?, ?, ?, ?], setting lower limits on the G_{RS} and G_{Bulk} mass as a function of the coupling parameter $k/\overline{M}_{\text{Pl}}$, where k determines the curvature of the warped space and \overline{M}_{Pl} is the reduced Planck mass ($\overline{M}_{\text{Pl}} \equiv M_{\text{Pl}}/\sqrt{8\pi}$). The G_{Bulk} and G_{RS} models differ in the fact that the G_{Bulk} favors the decay into vector bosons rather

than photons or fermions and favors the production of longitudinal polarized W or Z bosons. For the W' , the most stringent limits are reported in searches with leptonic final states [?, ?], and the current lower limit on the W' mass is 2.9TeV. The limit varies by 0.1TeV, depending on the chirality of the W' couplings. Specific searches in the WZ final state have also been reported [?, ?, ?] setting a lower limit of 1.1TeV.

This analysis is the reload of the previous analysis EXO-11-095 [?]. This analysis proceeds via the following steps:

1. The search is performed in the dijet sample, using the same preselection as the standard search for resonances decaying to dijets [?].
2. We identify events with substructure: in each jet which is a candidate to originate from merging of W or Z daughter jets
 - we require a pruned jet mass cut, and
 - an n-subjettiness cut
3. After the full event selection, a potential signal would be characterized as a peak in the dijet invariant mass, on top of a falling background distribution.
4. We model the background distribution with a smoothly falling analytical function.
5. Finally, in absense of a signal, we set the limits on the production cross section of models with qW, qZ, WW, WZ and ZZ final states using the CL_s approach.

3.2 Dijet analysis with jet substructure tagging

3.2.1 Jet reconstruction

Based on CMSSW 5.3.x software package, events are reconstructed using the particle-flow reconstruction algorithm [?], which attempts to reconstruct all stable particles in an event by combining information from all subdetectors. The algorithm categorizes all particles into five types: muons, electrons, photons, charged and neutral hadrons. The resulting particle flow candidates are passed to each jet clustering algorithm, in this case the Cambridge-Aachen (CA) [?, ?] jet clustering algorithm, as implemented in FastJet version 3.0.1 [?, ?], to create "particle flow jets". The CA clustering sequence is only determined by the distance between clusters and is not weighted by their momentum, as is done for the k_T and anti- k_T algorithms. A distance parameter of size $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.8$ is used for the CA algorithm.

Charged hadrons identified as pileup are removed from the inputs to the jet clustering algorithms. The remaining neutral component of pileup is removed by applying a residual area-based correction as described in Ref. [?, ?]. The mean p_T per unit area is computed with the k_T algorithm with the "active area" method, with a distance parameter of 0.6, and the jet energy is corrected by the amount of pileup expected in the jet area. The amount of energy expected from the underlying event is added

back into the jet. The pileup-subtracted jet four momenta are finally corrected for nonlinearities in η and p_T with simulated data, with a residual η -dependent correction added to correct for the difference in simulated and true responses [?, ?].

The jet energy corrections for the CA $R = 0.8$ jets are derived from studies using the anti- k_T $R = 0.7$ jet algorithm. Simulation studies confirm that these anti- k_T -derived jet corrections are adequate for the CA $R = 0.8$ jet algorithm for the jet momenta considered here [?].

3.2.2 Event selection

Events are selected using the following cuts:

- The event must have a well reconstructed primary vertex as computed by a deterministic annealing filter (DAF) ($|z_{\text{Primary Vertex}}| < 24 \text{ cm}$, $N_{\text{DOF}} > 6$).
- The following recommended noise event filters are used:
 - CSC tight beam halo filter
 - HBHE noise filter with isolated noise rejection
 - HCAL laser event filter (HBHE) and HCAL laser event filter 2012
 - ECAL dead cell trigger primitive (TP) filter
 - The beam scraping filter
 - Bad EE supercrystal filter

CHAPTER 3. EXO-12-024

- The tracking failure filter
 - Good primary vertex filter
 - Tracking coherent noise filter
 - Tracking TOBTEC fakes filter
- The events are required to have at least two ungroomed CA8 jets with
 - $p_T > 30 \text{ GeV}$, $|\eta| < 2.5$
 - to have muon energy fraction < 0.8
 - pass tight particle flow jet ID. The tight PF jet ID is listed below:
 - * Neutral Hadron (EM) Fraction $< 0.90(< 0.90)$, for all jet η
 - * Number of Constituents > 1 , for all jet η
 - * Charged Hadron (EM) Fraction $> 0(< 0.99)$, for jet $|\eta| < 2.4$
 - * Charged Multiplicity > 0 , for jet $|\eta| < 2.4$
 - Beam background events are removed using the following requirements:
 - In events with at least 10 tracks, a minimum of 25% of these tracks must be high purity tracks.
 - We also require $E_T^{miss} / \sum E_T < 0.5$ to further suppress the noise producing large fake E_T^{miss} .
 - The events must pass $|\Delta\eta| < 1.3$, $m_{jj} > 890 \text{ GeV}$

CHAPTER 3. EXO-12-024

This sample of dijet events is then tested for presence of hadronically decaying W or Z bosons.

3.3 Data and Monte Carlo samples

The data sample of proton-proton collisions at $\sqrt{s} = 8$ TeV was collected in 2012 and corresponds to an integrated luminosity of 19.7 fb^{-1} . The datasets and also the certifications used are summarized in Table ???. The dijet sample is dominated by light flavored and gluon jets, which we denote as the "QCD background". The QCD background is obtained from data by fitting an analytic parameterization of the dijet invariant mass distribution.

Signal events have been simulated using JHUGEN [?, ?], PYTHIA 6.426 [?] and HERWIG++ 2.5.0 [?] event generators and processed through a simulation of the CMS detector, based on GEANT4 [?]. PYTHIA 6 is used with CTEQ61L [?] and HERWIG++ with MRST2001 [?] parton distribution functions. Tune Z2* (a modification of tune Z1 [?]) is used with PYTHIA 6, while the tune version 23 is used with HERWIG++. The process $q^* \rightarrow W/Z + \text{jet}$ is generated using PYTHIA 6. RS graviton production is studied with $k/\overline{M}_{\text{Pl}} = 0.1$, which determines a resonance width of about 1% of the resonance mass which is about a factor five smaller than the experimental resolution for dijets. While HERWIG++ contains a more detailed description of the angular distributions for G_{RS} than PYTHIA 6 for this process [?] and is therefore used to model the G_{RS} resonance shape, the PYTHIA 6 cross section is used to maintain consistency with reference models used in related analyses [?]. Bulk graviton production is studied with $k/\overline{M}_{\text{Pl}} = 0.2$ and is generated with JHUGEN interfaced with PYTHIA 6 for the showering. Bulk graviton cross sections are cal-

culated using CalcHEP. The process $W' \rightarrow WZ$ is generated using PYTHIA 6 with Standard Model $V - A$ couplings and without applying k-factors.

To validate our RS graviton resonance Monte Carlo, we compare Pythia6, Herwig++ and a generator including full angular correlations developed by the JHU group (which we denote “JHU generator”). Figure ?? shows the comparisons of invariant mass and $\Delta\eta$ of two Z bosons at generator level, in which Herwig++ and Pythia6 are compared with the JHU generator which describes the angular distributions exactly. Pythia6 does not implement the angular correlations, and from Figure ?? one can indeed conclude that in its description of this effect it is inferior to Herwig++.

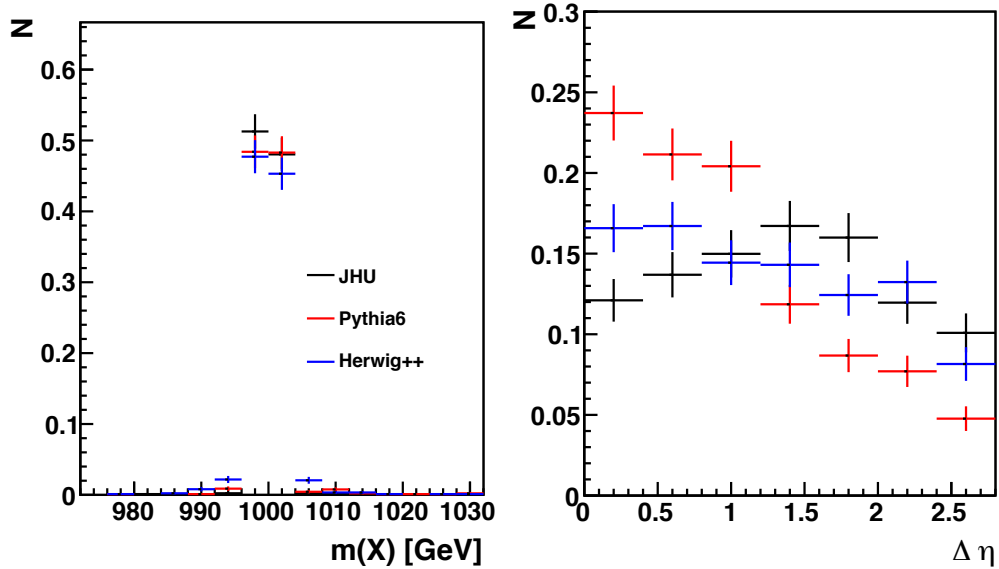


Figure 3.1: Invariant mass and $\Delta\eta$ of two Z bosons at generator level for Pythia6 and Herwig++ models of a 1 TeVRS graviton resonance with $k/M_{PL} = 0.02$ with the JHU generator which includes all angular correlations.

CHAPTER 3. EXO-12-024

All Monte Carlo events are fully simulated and reconstructed via the Geant4-based CMS simulation and reconstruction software.

Tables ??, ??, ?? and ?? summarize the new physics simulation datasets used in this analysis.

Dataset
/Jet/Run2012A-22Jan2013-v1/AOD
/JetHT/Run2012B-22Jan2013-v1/AOD
/JetHT/Run2012C-22Jan2013-v1/AOD
/JetHT/Run2012D-22Jan2013-v1/AOD

Table 3.1: Summary of 8 TeV collision data used in this analysis. The certification file used for these data is `Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`.

Table ?? describes a single-tagged process: $q^* \rightarrow W/Z + jet$ with a large cross section. We generated the MC using Pythia6 with Tune Z2*. The configuration is in the appendix of this note. The parameters `RTCM(43)`, `RTCM(44)`, `RTCM(45)` are set to 1 and the scale `RTCM(41)` is set to the resonance mass `PMAS(343,1)=PMAS(344,1)`. Only decays in to qW or qZ are allowed. We generate the process $q^* \rightarrow W/Z + jet$ using Pythia6 with Tune Z2*.

Table ?? shows a double-tagged process: $G_{RS} \rightarrow WW/ZZ$. This is produced using Herwig++ with Tune23 and as a cross check also in Pythia6 with Tune Z2*. In Pythia6, the parameter `PARP(50)` corresponding to $5.4k/\bar{M}_{Pl}$ which impacts the

Process	Generator	Events	X-sec[pb]
$qW(m=750GeV)$	Pythia6	30000	1.133E+02
$qW(m=1000GeV)$	Pythia6	30000	2.647E+01
$qW(m=1500GeV)$	Pythia6	30000	2.540E+00
$qW(m=2000GeV)$	Pythia6	30000	3.510E-01
$qW(m=3000GeV)$	Pythia6	30000	1.008E-02
$qZ(m=750GeV)$	Pythia6	30000	4.071E+01
$qZ(m=1000GeV)$	Pythia6	30000	9.405E+00
$qZ(m=1500GeV)$	Pythia6	30000	8.937E-01
$qZ(m=2000GeV)$	Pythia6	30000	1.231E-01
$qZ(m=3000GeV)$	Pythia6	30000	3.465E-03

Table 3.2: Summary of the simulated Monte Carlo samples used in this analysis for process $q^* \rightarrow Z/W + jet$

width and cross section of the resonance. In Herwig++, the cross section and width

are given by the ratio of `RS/Model:Lambda_pi` and the resonance mass `/Herwig/Particles/Graviton`

The process $G_{RS} \rightarrow WW/ZZ$ is generated using Herwig++ with Tune23 and its cross

section is taken from Pythia6 with Tune Z2*. We study RS graviton production with

$k/\bar{M}_{Pl} = 0.1$, defining a resonance width smaller than the experimental resolution

for dijets. Table ?? describes another double-tagged process: $W' \rightarrow WZ$. This is

produced using Pythia6 with Tune Z2*. The decay of the W' is restricted to WZ with

`MDME(331,1)=1`. The process $W' \rightarrow WZ$ is generated using Pythia6 with Tune Z2*.

CHAPTER 3. EXO-12-024

All Monte Carlo events are fully simulated and reconstructed via the CMS simulation and reconstruction software.

Process	Generator	Events	Pythia6 x-sec [pb]
WW(m=750GeV)	Herwig++/Pythia6 Z2*	30000	2.220E+00
WW(m=1000GeV)	Herwig++/Pythia6 Z2*	30000	4.254E-01
WW(m=1500GeV)	Herwig++/Pythia6 Z2*	30000	3.298E-02
WW(m=2000GeV)	Herwig++/Pythia6 Z2*	30000	4.083E-03
WW(m=2500GeV)	Herwig++/Pythia6 Z2*	30000	6.191E-03
WW(m=3000GeV)	Herwig++/Pythia6 Z2*	30000	1.010E-04
ZZ(m=750GeV)	Herwig++/Pythia6 Z2*	30000	1.120E+00
ZZ(m=1000GeV)	Herwig++/Pythia6 Z2*	30000	2.137E-01
ZZ(m=1500GeV)	Herwig++/Pythia6 Z2*	30000	1.662E-02
ZZ(m=2000GeV)	Herwig++/Pythia6 Z2*	30000	2.027E-03
ZZ(m=2500GeV)	Herwig++/Pythia6 Z2*	30000	3.077E-04
ZZ(m=3000GeV)	Herwig++/Pythia6 Z2*	30000	5.099E-05

Table 3.3: Summary of the simulated Monte Carlo samples used in this analysis for process $G_{RS} \rightarrow WW, ZZ$.

Process	Generator	Events	X-sec[pb]
$WZ(m=750\text{GeV})$	Pythia6	30000	5.391E-01
$WZ(m=1000\text{GeV})$	Pythia6	30000	1.444E-01
$WZ(m=1500\text{GeV})$	Pythia6	30000	1.804E-02
$WZ(m=2000\text{GeV})$	Pythia6	30000	3.129E-03
$WZ(m=2500\text{GeV})$	Pythia6	30000	6.781E-04
$WZ(m=3000\text{GeV})$	Pythia6	30000	1.894E-04

Table 3.4: Summary of the simulated Monte Carlo samples used in this analysis for process $W' \rightarrow WZ$.

Process	Generator	Events	X-sec[pb]
$WW(m=1000\text{GeV})$	JHU Z2*	50000	0.001774
$WW(m=1500\text{GeV})$	JHU Z2*	50000	9.207E-05
$WW(m=2000\text{GeV})$	JHU Z2*	50000	8.004E-06
$WW(m=2500\text{GeV})$	JHU Z2*	50000	8.851E-07
$WW(m=3000\text{GeV})$	JHU Z2*	50000	-
$ZZ(m=1000\text{GeV})$	JHU Z2*	50000	0.0009044
$ZZ(m=1500\text{GeV})$	JHU Z2*	50000	4.622E-05
$ZZ(m=2000\text{GeV})$	JHU Z2*	50000	4.029E-06
$ZZ(m=2500\text{GeV})$	JHU Z2*	50000	4.460E-07
$ZZ(m=3000\text{GeV})$	JHU Z2*	50000	-

Table 3.5: Summary of the simulated Monte Carlo samples used in this analysis for process $G_{Bulk} \rightarrow WW, ZZ$.

3.4 Trigger

Events are selected if one of the following triggers has fired: HLT_HT750, HLT_PFHT650, HLT_PFNPUHT650 HLT_FatDiPFJetMass750_DR1p1_Deta1p5. All versions of each of these triggers is used. None of these triggers are prescaled during the 2012 data taking period. HLT_PFNPUHT650 trigger is used for the data set after the RunC(including RunC), while HLT_PFHT650 trigger is only used for RunA and RunB data sets.

Figure ??, Figure ?? and Figure ?? shows the trigger efficiencies of the OR of the highest threshold HLT_PFHT650 trigger and the HLT_FatJetMass trigger w.r.t. an OR of the lower threshold HLT_HT550 trigger. From the plot, the trigger is 99% efficient above 890GeV for the untagged, single tagged and double tagged data.

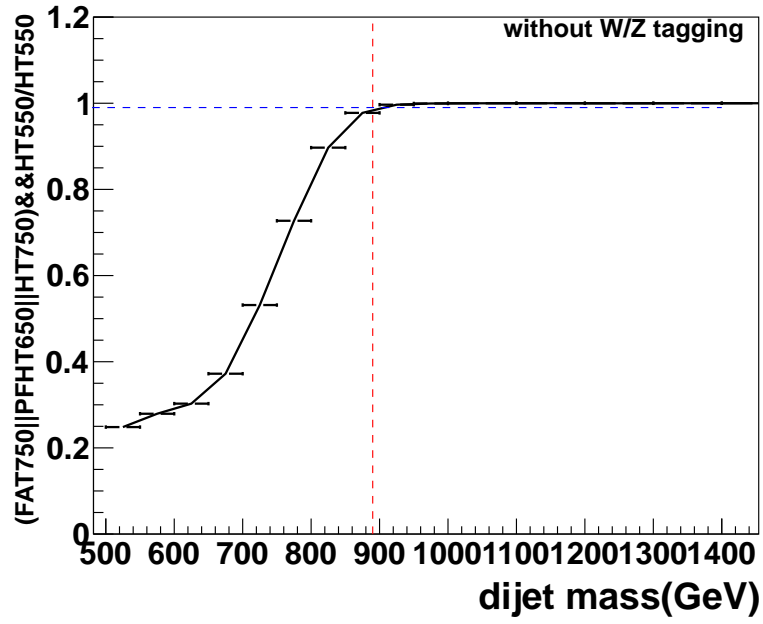


Figure 3.2: Trigger efficiency for untagged data of $\text{FAT}_{750}||\text{HLT_PF(NoPU)}\text{HT}_{650}||\text{HLT_HT}_{750}$ measured using data collected by lower threshold H_T550 trigger. The dash red line is positioned at m_{jj} equal 890GeV , the blue line is at efficiency at 99%.

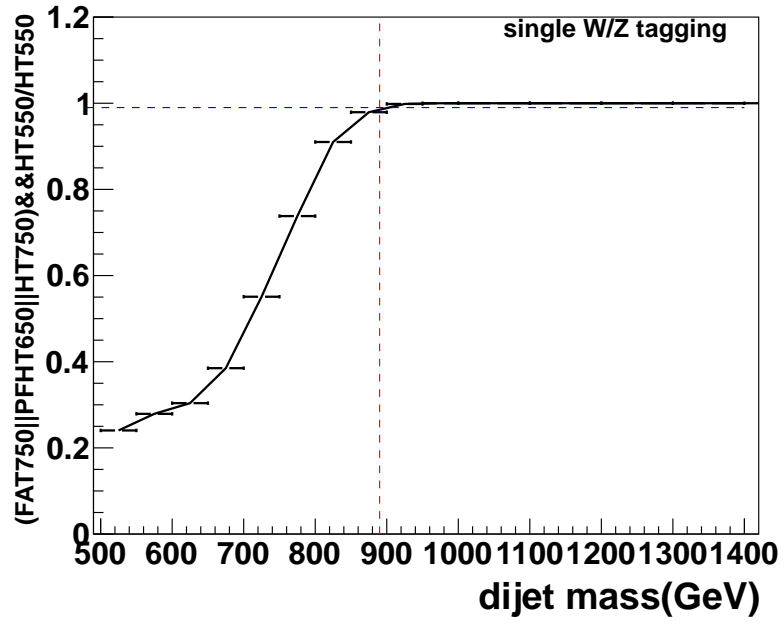


Figure 3.3: Trigger efficiency for single tagged data of $\text{FAT}_{750}||\text{HLT_PF(NoPU)HT650}||\text{HLT_HT750}$ measured using data collected by lower threshold H_T550 trigger. The dash red line is positioned at m_{jj} equal 890GeV , the blue line is at efficiency at 99%.

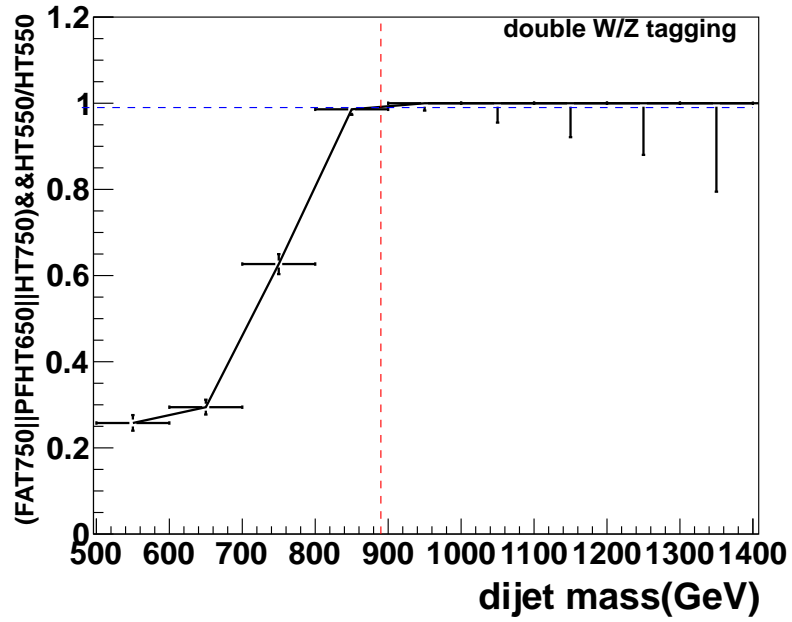


Figure 3.4: Trigger efficiency for double tagged data of $\text{FAT}_{750}||\text{HLT_PF(NoPU)}\text{HT}_{650}||\text{HLT_HT}_{750}$ measured using data collected by lower threshold H_T550 trigger. The dash red line is positioned at m_{jj} equal 890GeV , the blue line is at efficiency at 99%.

3.5 Data and MC comparisons

In this section, we compare some kinematic features of the jets between QCD MC and data, which are shown in Fig ??, ??,??, ??, ??, ??,??, ??, ??, ??, ??, ??, ??, ??, and ??. Predictions from Pythia6 with Tune $Z2^*$ and Herwig++ with Tune 23 are shown. The comparison is shown in the exclusive dijet category, low and high purity, single and double tagged events.. The distributions are shown after the event selection (in particular $|y| < 2.5$, $|\Delta\eta| < 1.3$, $m_{jj} > 890 GeV$) is applied. The number of data events in each mass bin are shown in Table ??. The MC is normalized to the number of data events in each category and the shapes are compared.

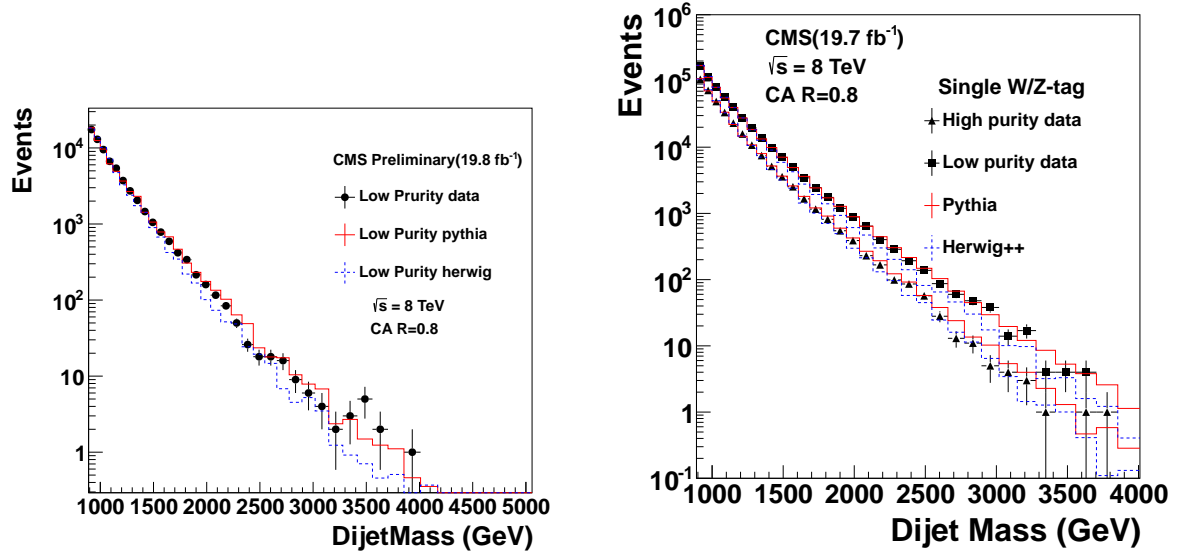


Figure 3.5: Comparisons between data and Monte Carlo for invariant mass of the two leading jets of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

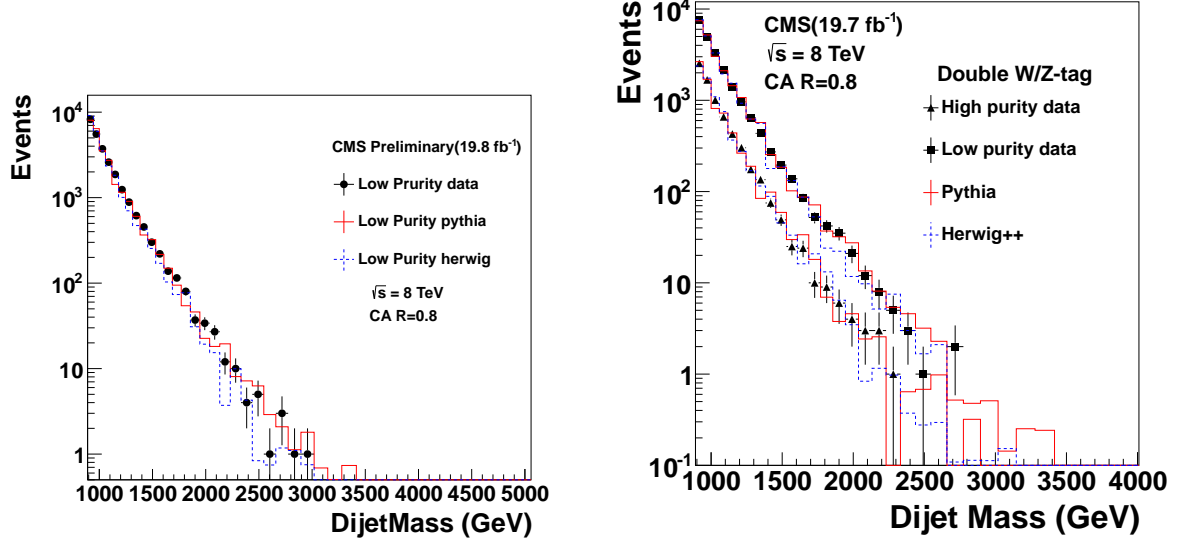


Figure 3.6: Comparisons between data and Monte Carlo for invariant mass of the two leading jets of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

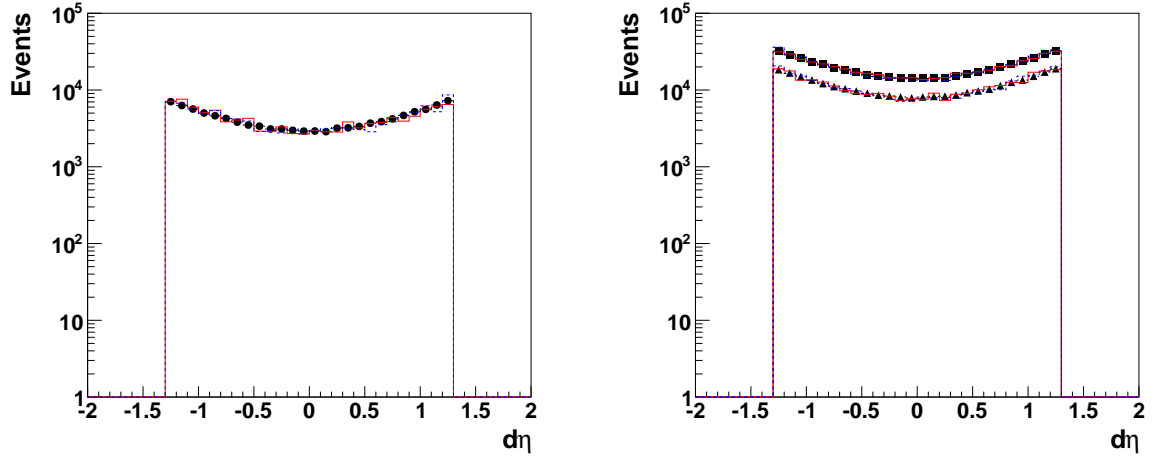


Figure 3.7: Comparisons between data and Monte Carlo for $\Delta\eta$ of the two leading jets of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

CHAPTER 3. EXO-12-024

lower mass bin border	low purity 1-tag events	high purity 1-tag events	low purity 2-tag events	high purity 2-tag events
890	165671	105892	7586	2544
944	115622	72007	4950	1673
1000	80537	48930	3311	1005
1058	56423	33398	2159	658
1118	39817	23086	1407	427
1181	27651	15817	962	302
1246	19531	10741	647	175
1313	13617	7477	434	135
1383	9880	5128	272	75
1455	6992	3578	195	49
1530	4939	2525	138	25
1607	3443	1658	86	24
1687	2454	1160	52	10
1770	1744	815	42	9
1856	1193	547	35	6
1945	881	389	21	4
2037	643	230	12	3
2132	402	167	8	3
2231	287	99	5	1
2332	193	86 ₃₆	3	
2438	138	57	1	
2546	87	28	0	

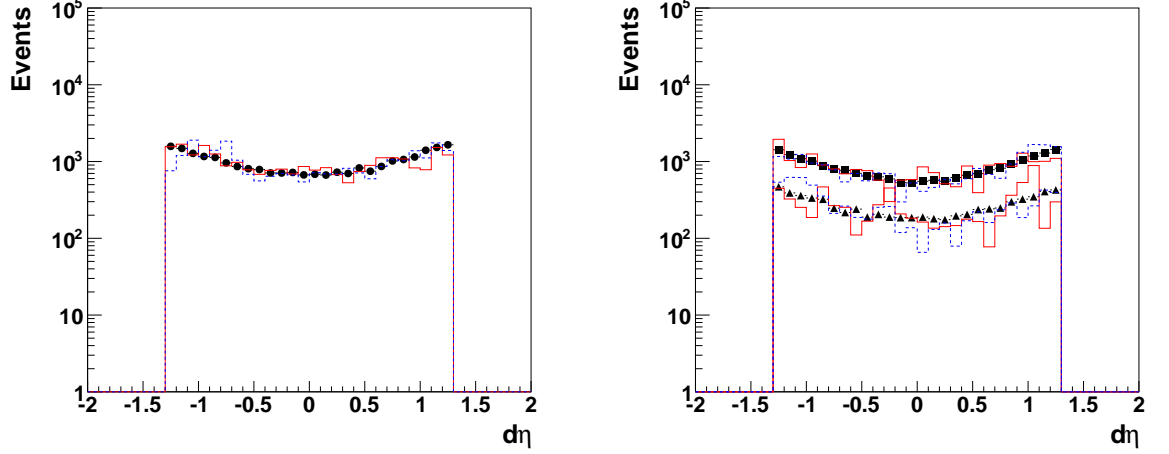


Figure 3.8: Comparisons between data and Monte Carlo for $\Delta\eta$ of the two leading jets of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

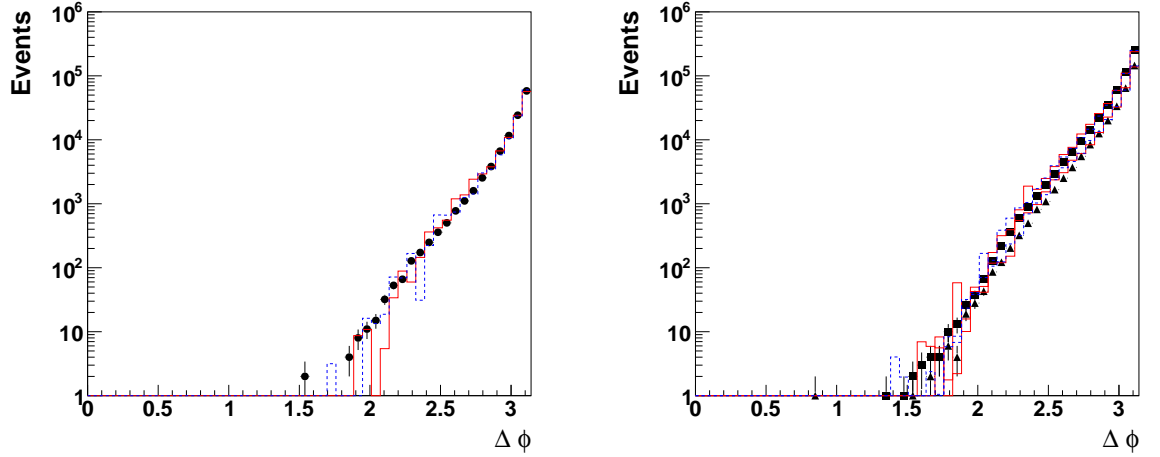


Figure 3.9: Comparisons between data and Monte Carlo for $\Delta\phi$ of the two leading jets of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

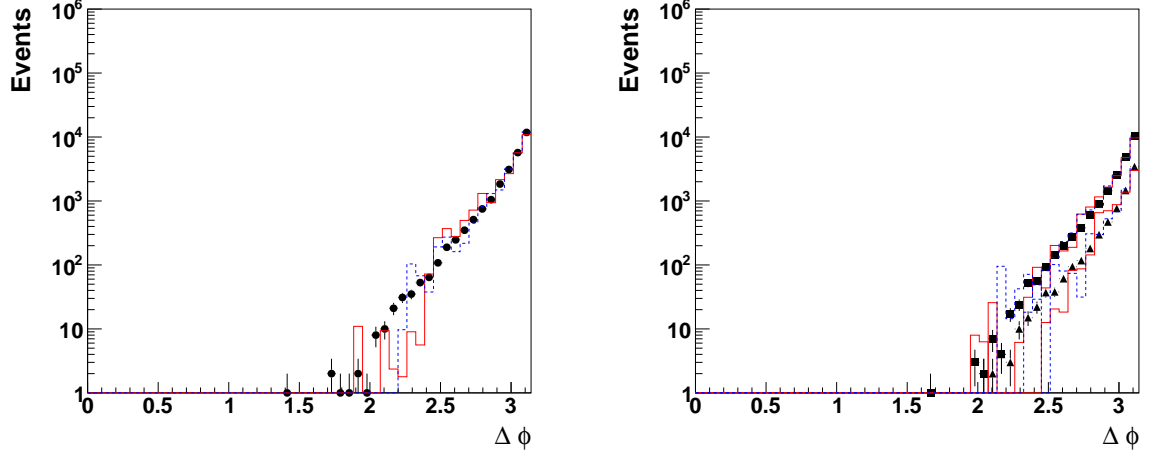


Figure 3.10: Comparisons between data and Monte Carlo for $\Delta\phi$ of the two leading jets of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

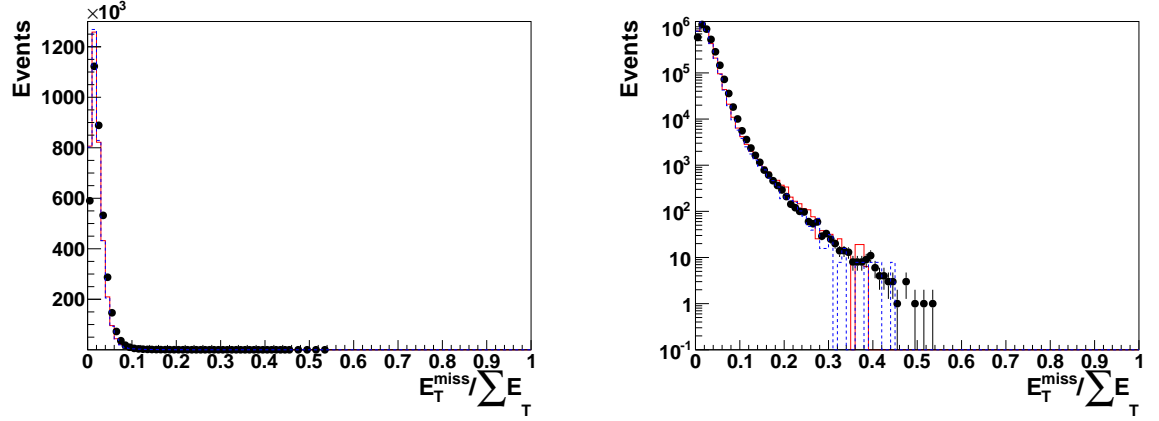


Figure 3.11: Comparisons between data and Monte Carlo for $E_T^{miss} / \sum E_T$. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

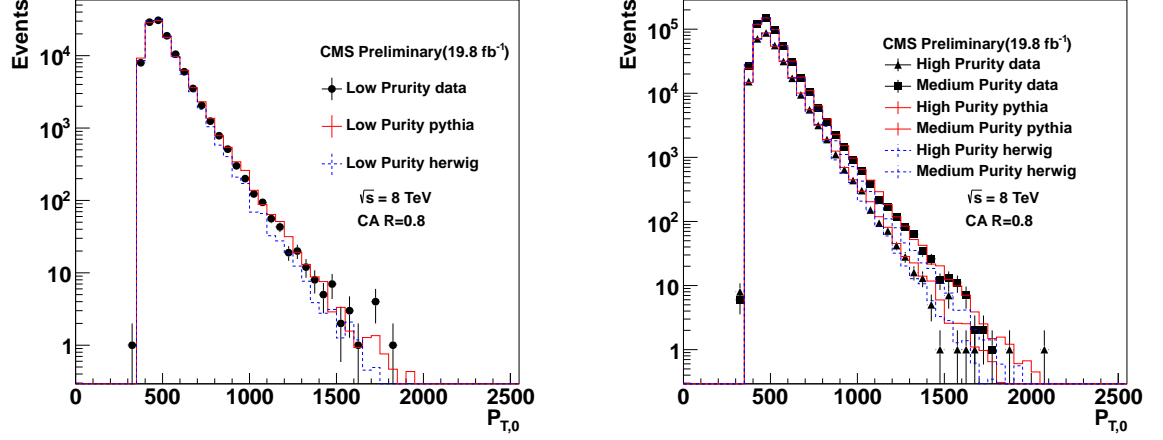


Figure 3.12: Comparisons between data and Monte Carlo for p_T of the leading jet of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

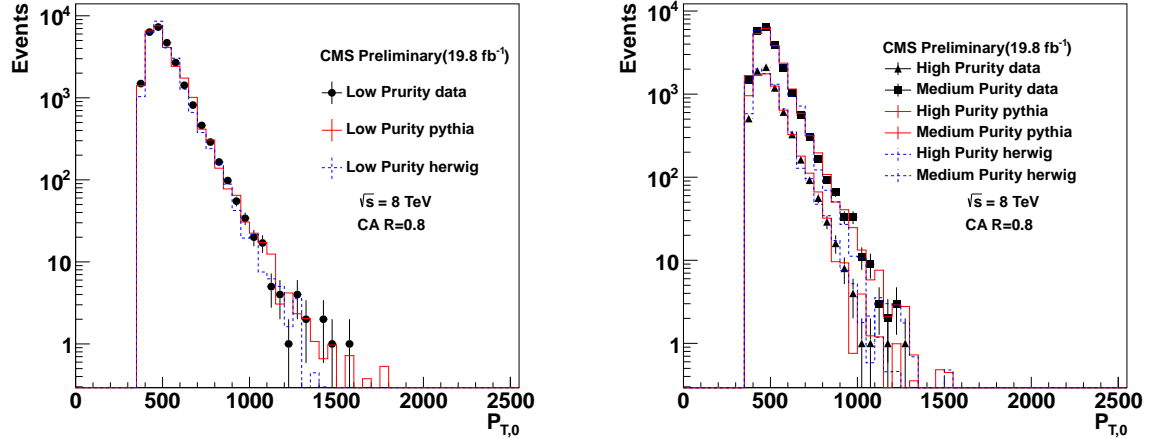


Figure 3.13: Comparisons between data and Monte Carlo for p_T of the leading jet of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

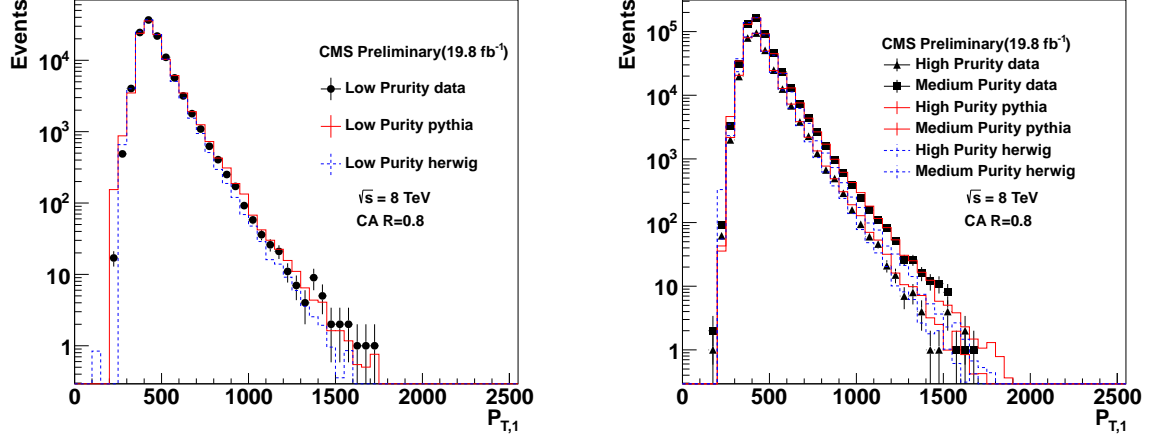


Figure 3.14: Comparisons between data and Monte Carlo for p_T of the second leading jet of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

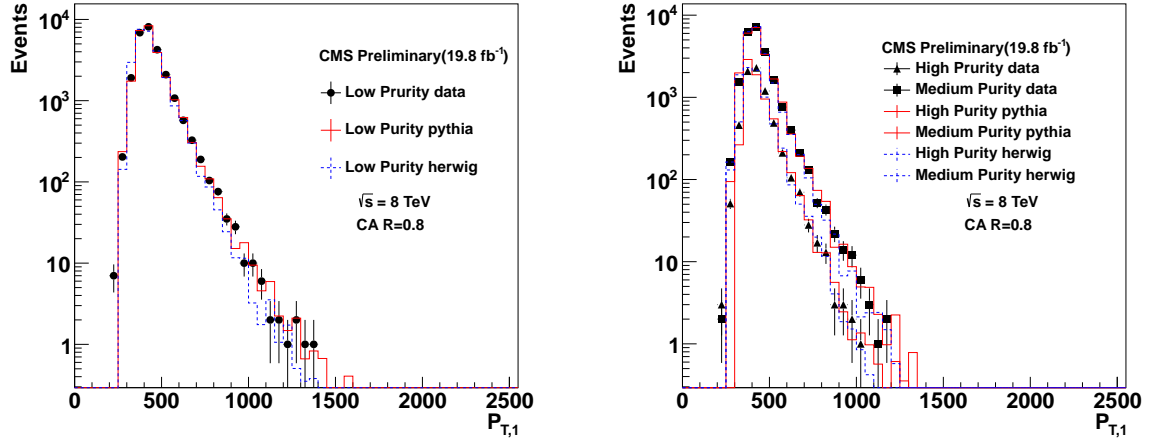


Figure 3.15: Comparisons between data and Monte Carlo for p_T of the second leading jet of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

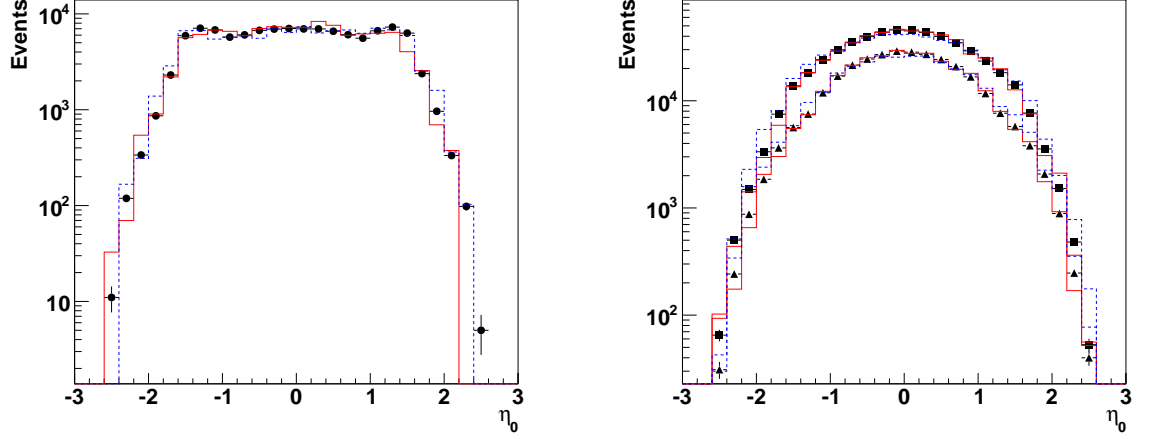


Figure 3.16: Comparisons between data and Monte Carlo for η of the leading jet of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

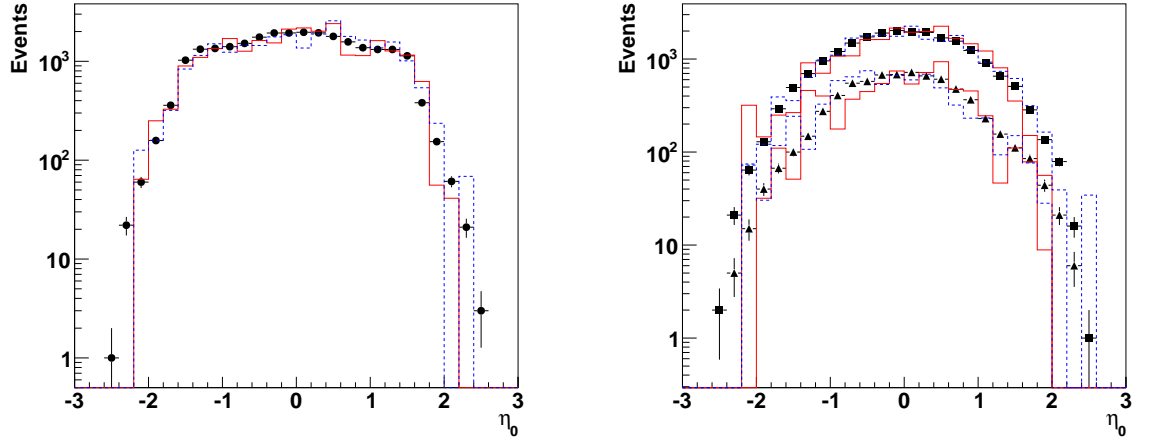


Figure 3.17: Comparisons between data and Monte Carlo for η of the leading jet of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

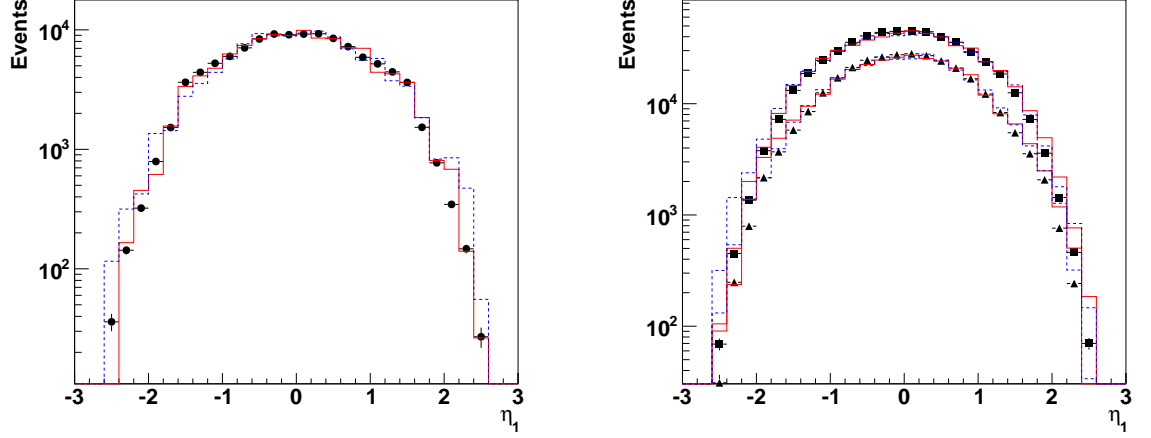


Figure 3.18: Comparisons between data and Monte Carlo for η of the second leading jet of low purity (left) and low-high purity (right) 1-tagged events. The MC is normalized to the number of data events in each category.

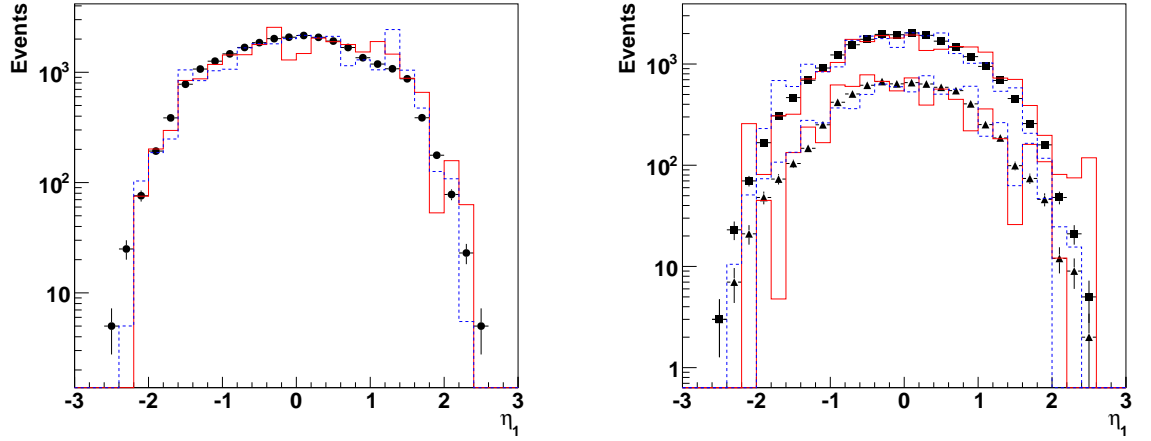


Figure 3.19: Comparisons between data and Monte Carlo for η of the second leading jet of low purity (left) and low-high purity (right) 2-tagged events. The MC is normalized to the number of data events in each category.

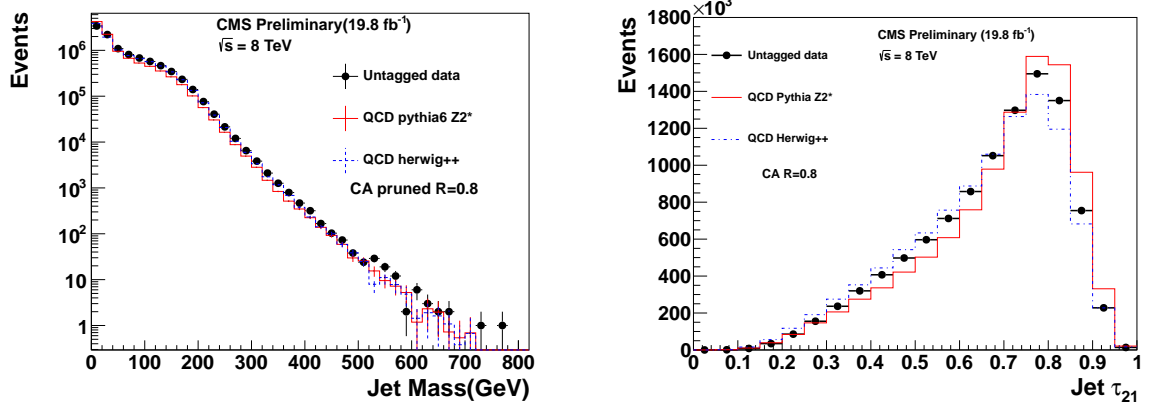


Figure 3.20: Comparisons between data and Monte Carlo for mass(left) and τ_{21} (right) of the leading two jets. The MC is normalized to the number of data events in each category.

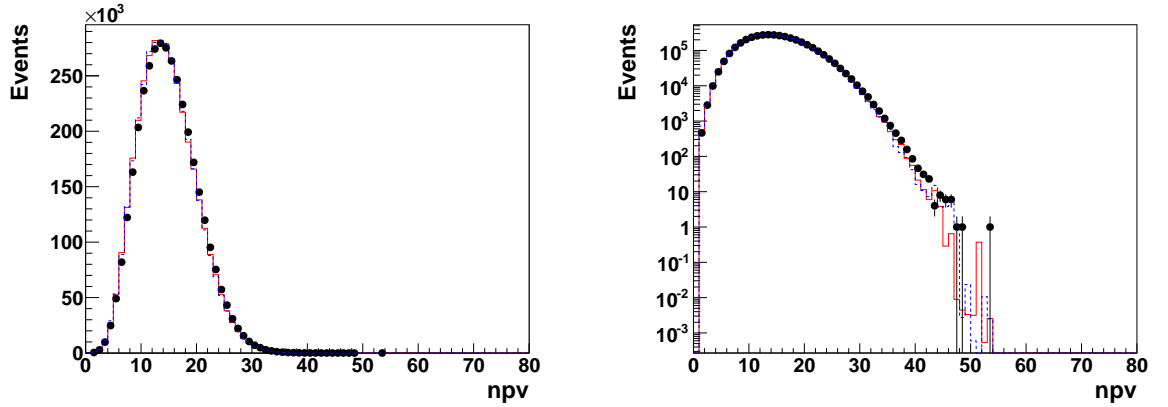


Figure 3.21: Comparisons between data and Monte Carlo for number of primary vertex to show the effect on Monte Carlo after pile up reweighting. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

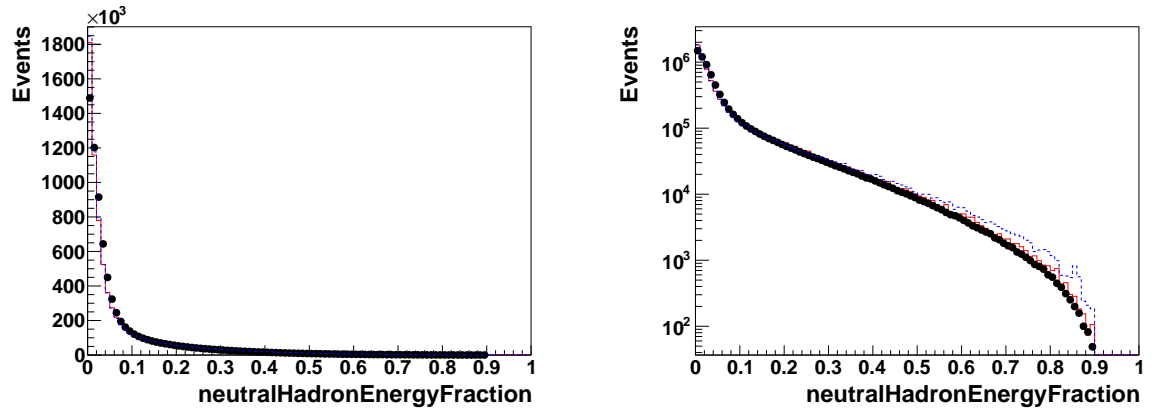


Figure 3.22: Comparisons between data and Monte Carlo for neutral hadron energy fraction. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

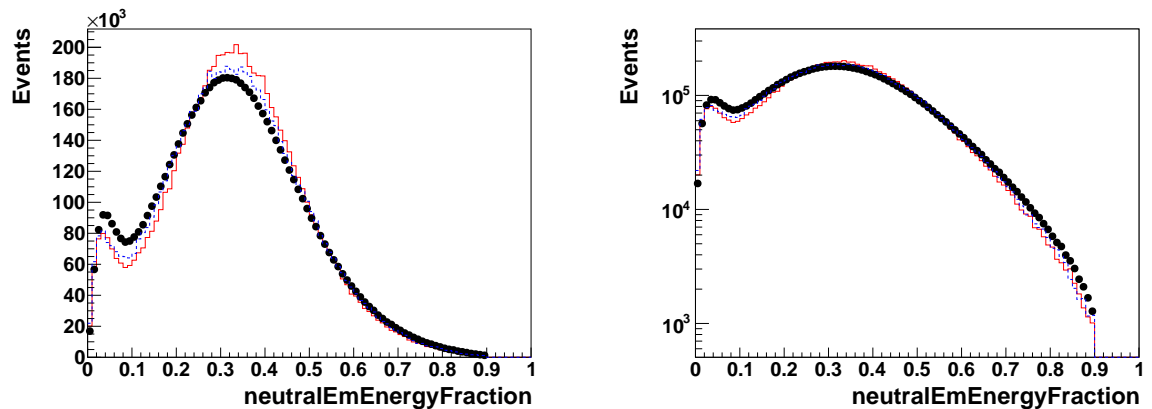


Figure 3.23: Comparisons between data and Monte Carlo for neutral electromagnetic energy fraction. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

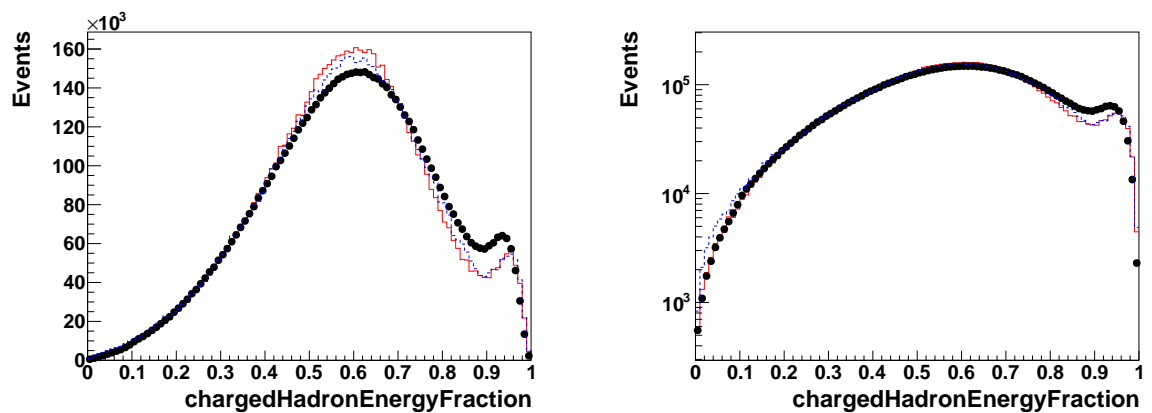


Figure 3.24: Comparisons between data and Monte Carlo for charged hadron energy fraction. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

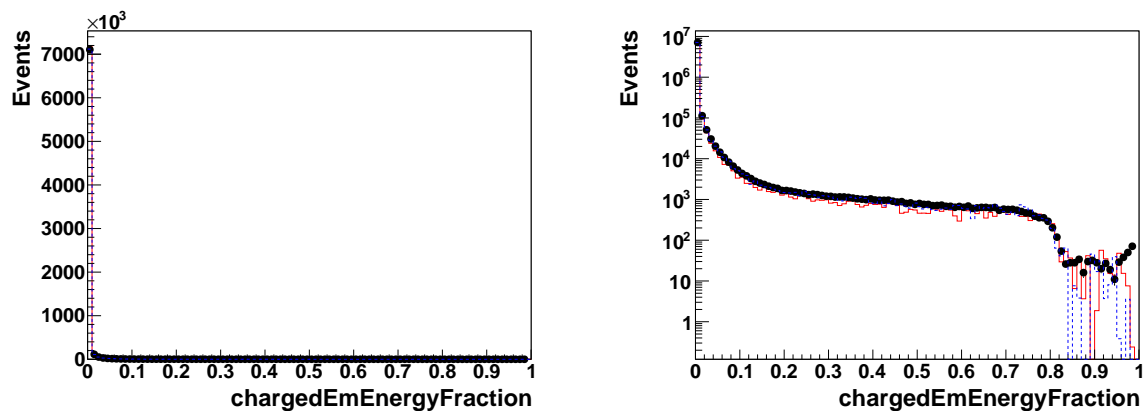


Figure 3.25: Comparisons between data and Monte Carlo for charged electromagnetic energy fraction. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

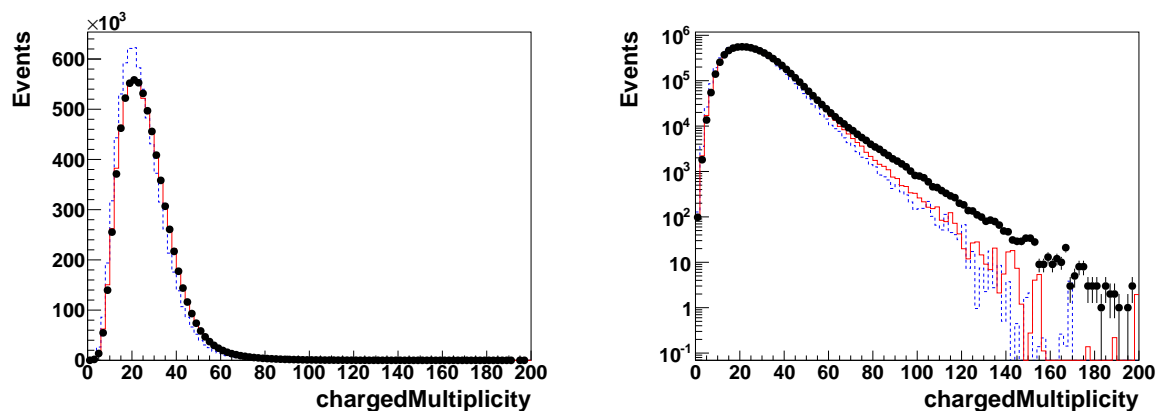


Figure 3.26: Comparisons between data and Monte Carlo for charged multiplicity. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

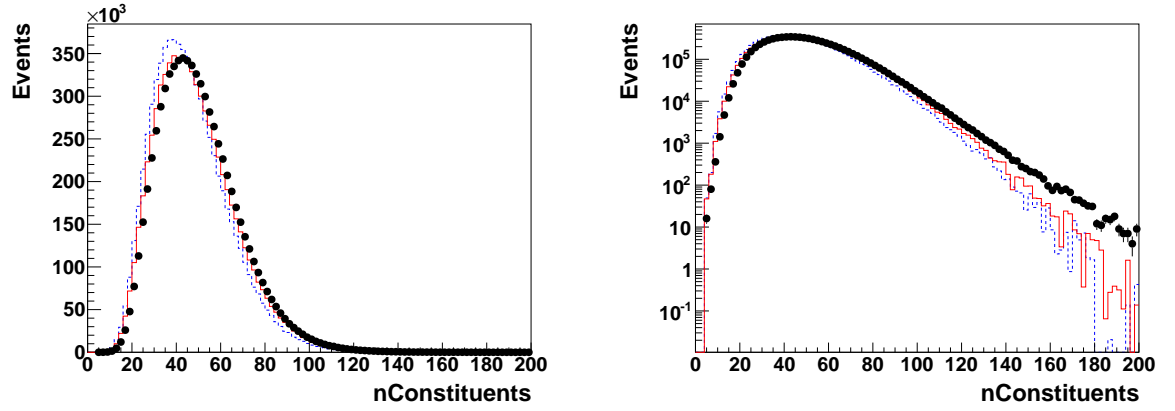


Figure 3.27: Comparisons between data and Monte Carlo for number of constituents. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

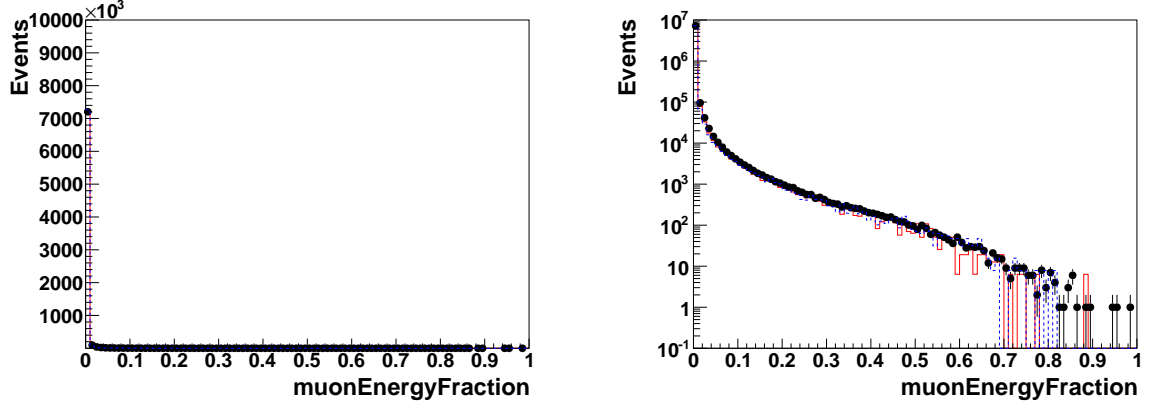


Figure 3.28: Comparisons between data and Monte Carlo for the muon energy fraction of the leading two jets. The MC is normalized to the number of data events. Plot on the right is the log scale plot. (The plot includes only a subset of the full data sample.)

We find that the QCD MC agrees with data, although not perfect. For the dijet kinematics and also the jet substructure variables, we observe about the same agreement of Pythia6 Z2* and Herwig++. For this analysis, we chose to model the background shape from the data itself (as described below) and depend on QCD MC only to provide us guidance and a cross check.

Fig ??, Fig ?? and Fig ?? are particularly useful to identify jets from calorimeter noise which would show up at low values of $\Delta\phi$ and high values of $E_T^{miss} / \sum E_T$. No enhancement in this region is observed which gives confidence that the applied noise filter cleaning and jet ID cuts leave no noise contamination within the two leading jets.

Fig ?? shows the number of primary vertices distribution after pile up reweighting on the MC. Fig ??, Fig ??, Fig ??, Fig ??, Fig ?? and Fig ?? show the jet ID variable

CHAPTER 3. EXO-12-024

distribution after the event selection, and Fig ??, the muon energy fraction of the leading two jets.

3.6 W-tagging scale factor

We derive the scale factor for the τ_2/τ_1 W tagger, in both the tight and loose regions, by comparing its efficiency in semileptonic $t\bar{t}$ events, both in data and Monte Carlo. We isolate the W candidates with kinematic cuts. We consider only muon events. We then apply the tagger and require the W mass to be within 70 and 100 GeV.

In Monte Carlo, we attempt to match the CA8 jets to real Ws by requiring that the daughters of a hadronic W from the particle generator lie within a cone of $\Delta R < 0.3$ of a jet's subjects. Jets that meet this requirement are "matched" jets. We can then classify all W candidates in the following ways:

1. Matched W jets which pass the tight τ_2/τ_1 cut.
2. Matched W jets which pass the loose τ_2/τ_1 cut.
3. Matched W jets which fail both τ_2/τ_1 cuts.
4. Unmatched W jets which pass the tight τ_2/τ_1 cut.
5. Unmatched W jets which pass the loose τ_2/τ_1 cut.
6. Unmatched W jets which fail both τ_2/τ_1 cuts.

The efficiency for either tight or loose can be extracted by counting the number of matched W jets in that region and dividing by the total number of matched W jets.

We derive the efficiency in data by simultaneously fitting the W mass distributions of the events in the tight, loose and failed events. The general shapes of the distributions are taken from MC. The efficiencies are explicit fit parameters which relate the normalizations of categories category. We must also take into account small background contributions from non- $t\bar{t}$ sources. These contributions are also parametrized as shapes and included in the fit.

3.6.1 Fit to Monte Carlo

We first find the PDFs associated with each of the event categories above by fitting their distributions from the Monte Carlo, as follows:

- **Tight, Matched Jets** We fit the tight matched events with the sum of two gaussians.
- **Loose, Matched Jets** We fir the loose matched events with a sum of the double-gaussian found in the tight selection and an exponential.
- **Failed, Matched Jets** We fit failed matched events with an exponential.
- **Tight, Unatched Jets** We fit tight unmatched events with the sum of a gaussian and a linear function with positive slope.
- **Loose, Unatched Jets** We fit loose unmatched events with a gaussian.
- **Failed, Unatched Jets** We fit failed unmatched events with an exponential.

- **Backgrounds** All backgrounds are fit with gaussians in the tight and loose regions, and an exponential in the failed region. These shapes are added to the respective unmatched shapes to derive a total non-matched shape.

Fits to matched and unmatched distributions are shown in figures ?? and ??.

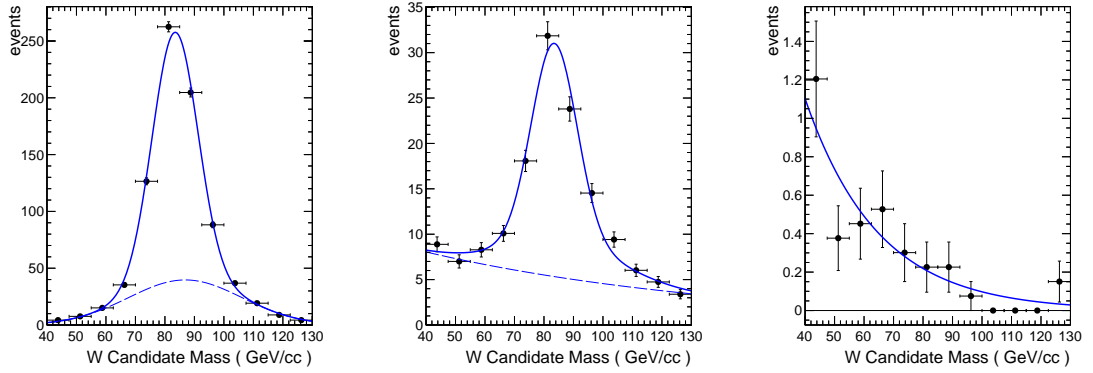


Figure 3.29: Fits to matched $t\bar{t}$ distributions. Left: tight region. Center: loose region. Right: failed events.

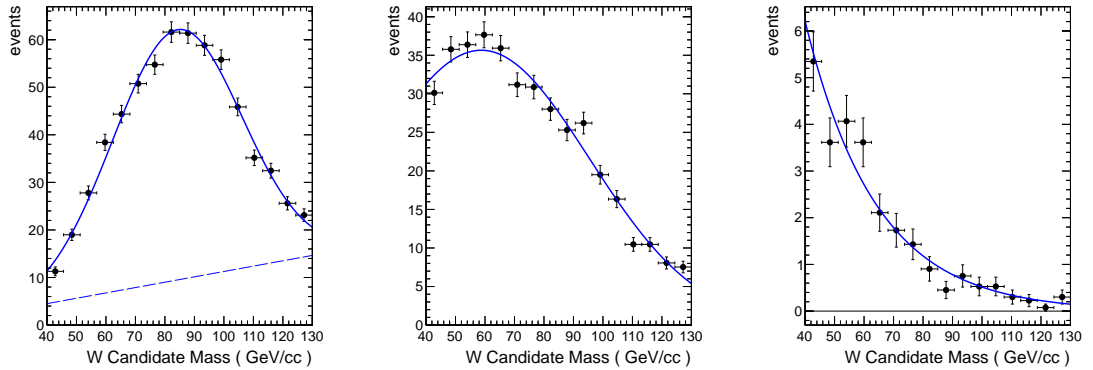


Figure 3.30: Fits to unmatched $t\bar{t}$ distributions. Left: tight region. Center: loose region. Right: failed events.

3.6.1.1 Fits to Data

We fit the above shapes to the data. The following parameters are kept constant, with their values taken from the MC fits described in the previous section:

1. In the unmatched and background distributions
 - (a) The relative normalization of the gaussian and linear components of the tight, unmatched shape.
 - (b) The position of the gaussian peak in the loose, unmatched distribution is fixed.
 - (c) The means and standard deviations, as well as the decay coefficients of all backgrounds are fixed, but their normalizations are not.
2. In the matched distributions
 - (a) The relative normalization of the two gaussians in the tight region is fixed.
 - (b) The relative position (as a multiplier) of the two gaussian peaks is fixed.
 - (c) The relative normalization of the exponential and double-gaussian shapes in the loose region is fixed.

All other parameters are allowed to float. All other normalizations are parametrized in terms of the efficiency. The results of the fits to data are shown in figure ???. As a consistency check, the same procedure is applied to the MC distribution. The result of that fit is shown in figure ???.

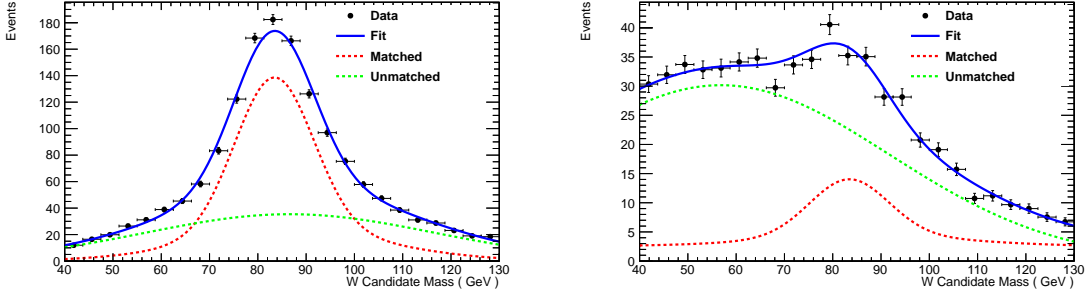


Figure 3.31: Distributions from data of W mass in the tight (left) and loose (right) τ_2/τ_1 regions, and resulting fits.

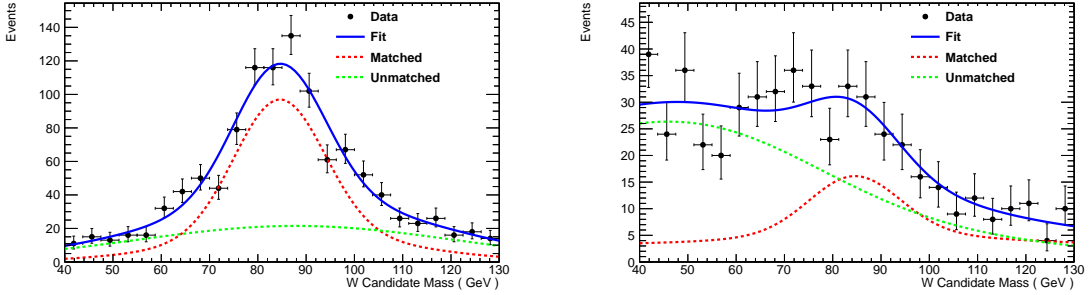


Figure 3.32: Distributions from MC of W mass in the tight (left) and loose (right) τ_2/τ_1 regions, and resulting fits.

More details on the fitting procedure and the $t\bar{t}$ selection can be found in the appendix of this note.

3.6.2 Scale Factor Measurement

We measure the scale factor of the τ_2/τ_1 cut efficiency and that of the W mass window 70 to 100 GeV efficiency. We find that the total scale factor for the two cuts is 0.860 ± 0.065 in the tight region and 138.5 ± 75.2 in the loose region.

3.6.2.1 Systematics

The errors in the scale factors are found from the fitting errors and systematic errors from the choice of fixed parameters. Each fixed parameter found in MC is varied by its fitting error and used to generate toy MC to which the fitting procedure is applied. The resulting offset is taken to be the systematic error from this fixed parameter. In addition, when the efficiency from a fit to MC differs with that from MC truth information, we take the difference to be an additional systematic. Additional systematic uncertainties on the W-tagging efficiency related to detector effects are discussed in the systematics section of this note.

3.7 The signal: dijet resonance

We search for dijet resonances corresponding to several models. Using the W/Z-tagging algorithm, we examine both single W/Z-tag and double W/Z-tag events.

The pruned jet mass and jet τ_{21} distributions in signal MC, data and background MC are shown in Fig. ???. Fully merged jets from Ws and Zs peak around 80-90 GeV in pruned jet mass while QCD jets and not fully merged Ws and Zs peak around 20 GeV. The discriminating power of the pruned jet mass and τ_{21} for the different signals is evident.

For both the pruned jet mass and τ_{21} , differences are observed between the HERWIG++ (G_{RS}) and PYTHIA6 (G_{Bulk} , q^* , W') distributions, which arise from differences in the polarization of the W/Z boson and the showering and hadronization models used by these generators. In particular this is the reason why the WZ prediction for τ_{21} is different from the WW , ZZ predictions. The differences due to showering and hadronization models are into account in the estimate of the systematic uncertainties on the tagging efficiency, as discussed below.

The full event selection efficiency is estimated using simulated signal samples. Less than 1% of the ZZ or WW events which pass the full selection are from $ZZ \rightarrow llqq$ or $WW \rightarrow l\nu qq$ decays, where l can be a muon or electron. While 3% of the selected ZZ events are from $ZZ \rightarrow \tau\tau qq$ decays, less than 1% of the selected WW events are from $WW \rightarrow \tau\nu qq$ decays. To within 10% accuracy the full selection efficiency can therefore be approximated by the product of the W/Z-tagging efficiency and an

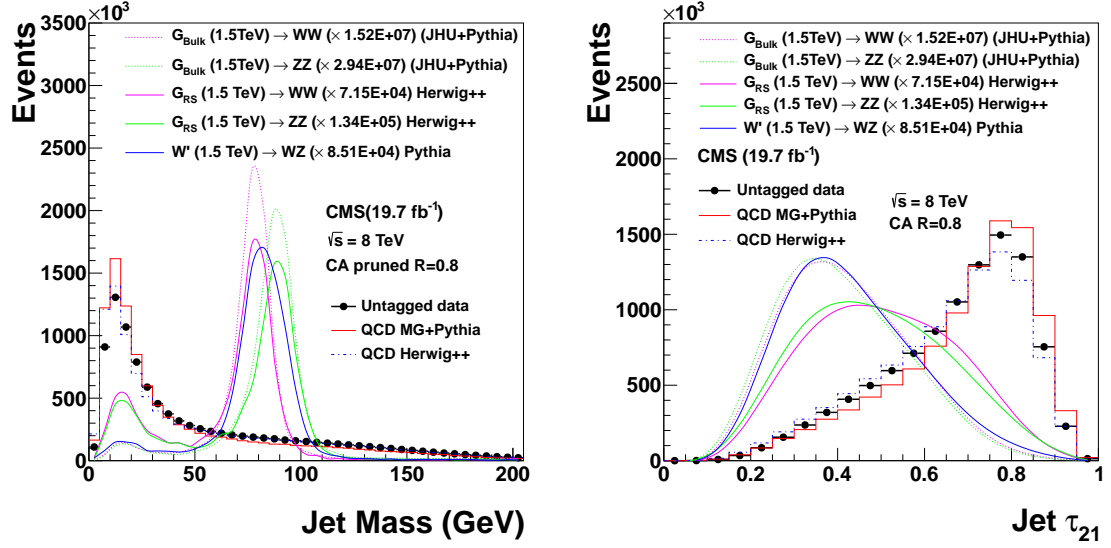


Figure 3.33: Pruned jet mass and τ_{21} in signal MC, data and background MC. All curves are plotted with the same binning. The signal MC distributions are plotted as smooth curves connecting the histogram entries. MC are normalized according to data.

approximate acceptance. This acceptance is shown in Fig. ?? and takes into account the angular acceptance ($|\eta| < 2.5$, $|\Delta\eta| < 1.3$), the branching into quark final states, $\text{BR}(W/Z \rightarrow \text{quarks})$ and a matching within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.5$ between the generated W/Z bosons and the reconstructed jets.

The W/Z-tagging efficiency is shown in Fig. ?? and Fig. ??.

The signal shapes for all five processes considered in this analysis are shown in Fig. ?? and Fig. ?. For the qW and qZ final states the shape with a single W/Z-tag required is shown, while for the other signals two W/Z-tags are required.

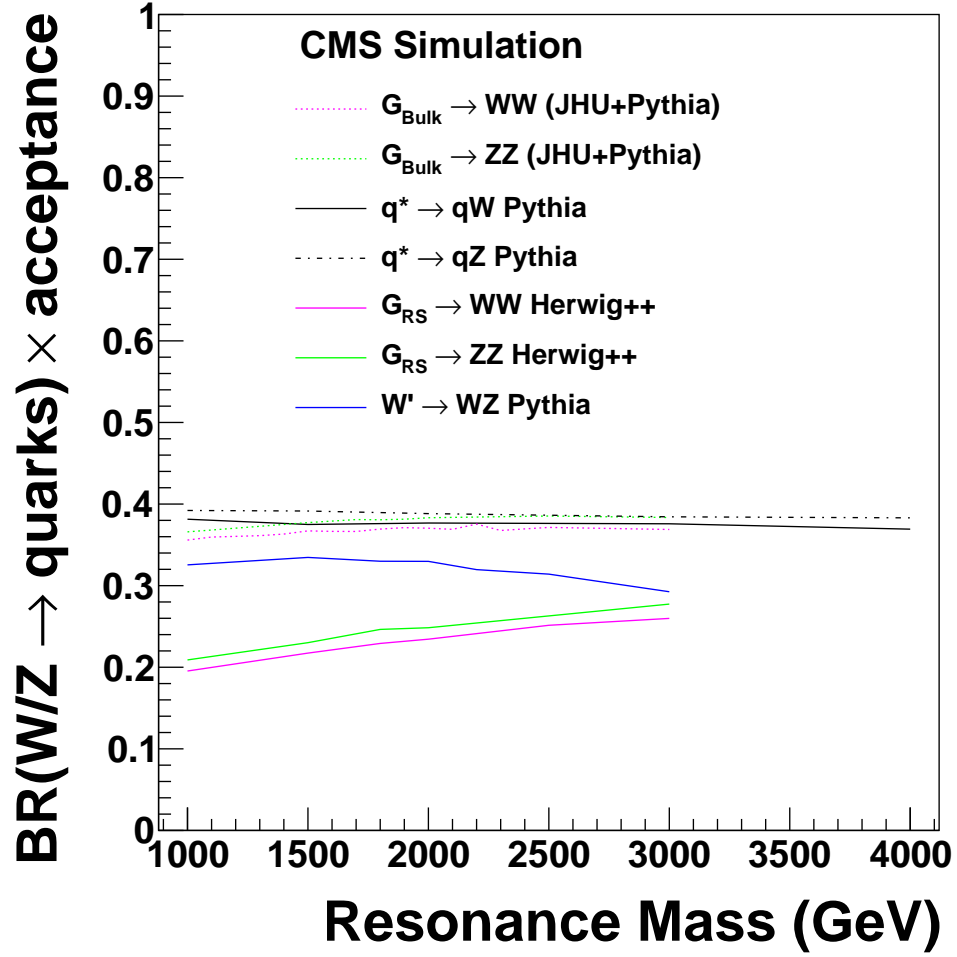


Figure 3.34: Fraction of events with branching into quark final states, $\text{BR}(W/Z \rightarrow \text{quarks})$, which are reconstructed as dijets (quarks \rightarrow jets) and pass the angular acceptance ($|\eta| < 2.5$, $|\Delta\eta| < 1.3$).

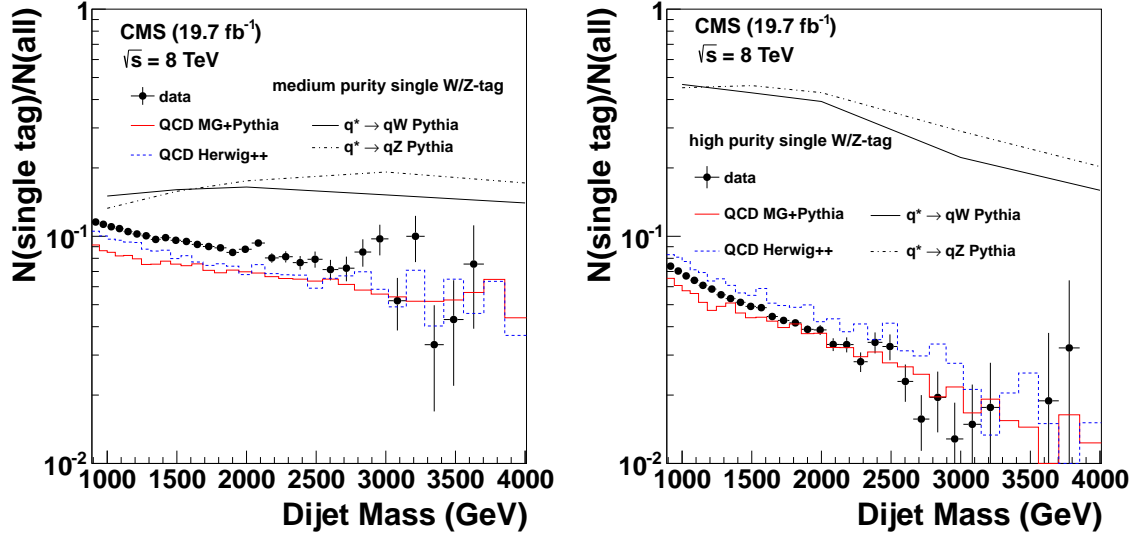


Figure 3.35: The fraction of singly-tagged events, requiring one medium purity (left) and high purity (right) W/Z -tag in data, signal and background simulations for events passing the angular acceptance requirement ($|\eta| < 2.5$, $|\Delta\eta| < 1.3$).

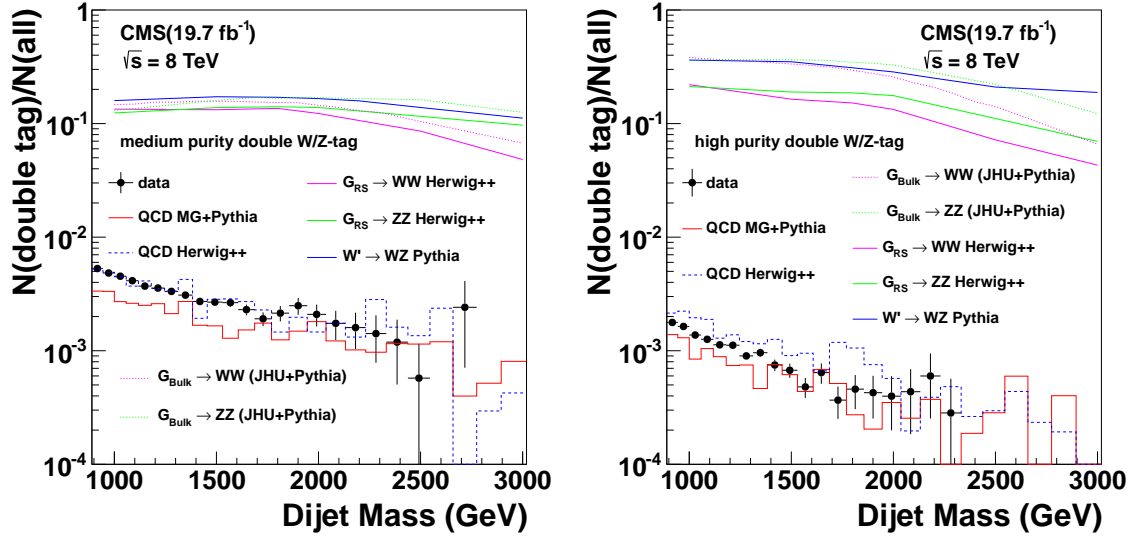


Figure 3.36: The fraction of doubly-tagged events, requiring two medium purity (left) and high purity (right) W/Z -tags in data, signal and background simulations for events passing the angular acceptance requirement ($|\eta| < 2.5$, $|\Delta\eta| < 1.3$).

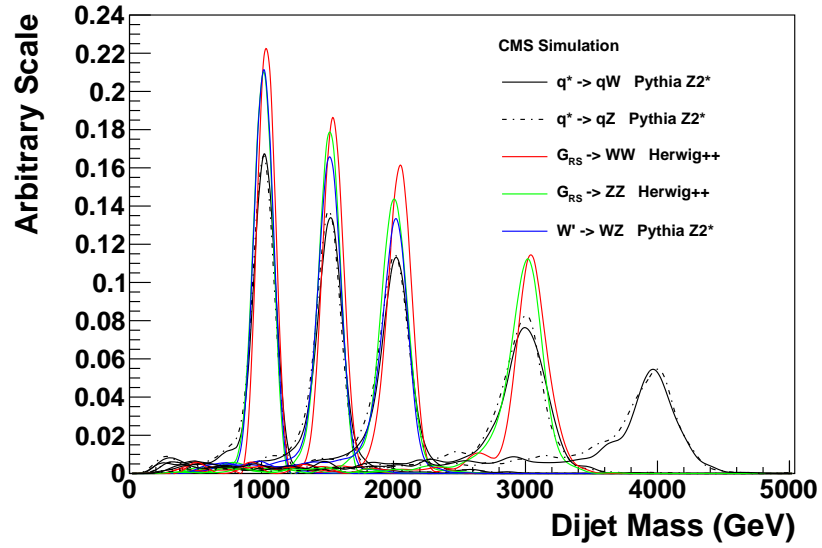


Figure 3.37: The normalized medium purity signal resonance distribution for $G_{RS} \rightarrow WW$, $G_{RS} \rightarrow ZZ$, $W' \rightarrow WZ$, $q^* \rightarrow qW$, and $q^* \rightarrow qZ$ resonances of dijet invariant mass 1.0TeV, 1.5TeV, 2.0TeV, 2.5 TeV, 3.0TeV, 4.0TeV.

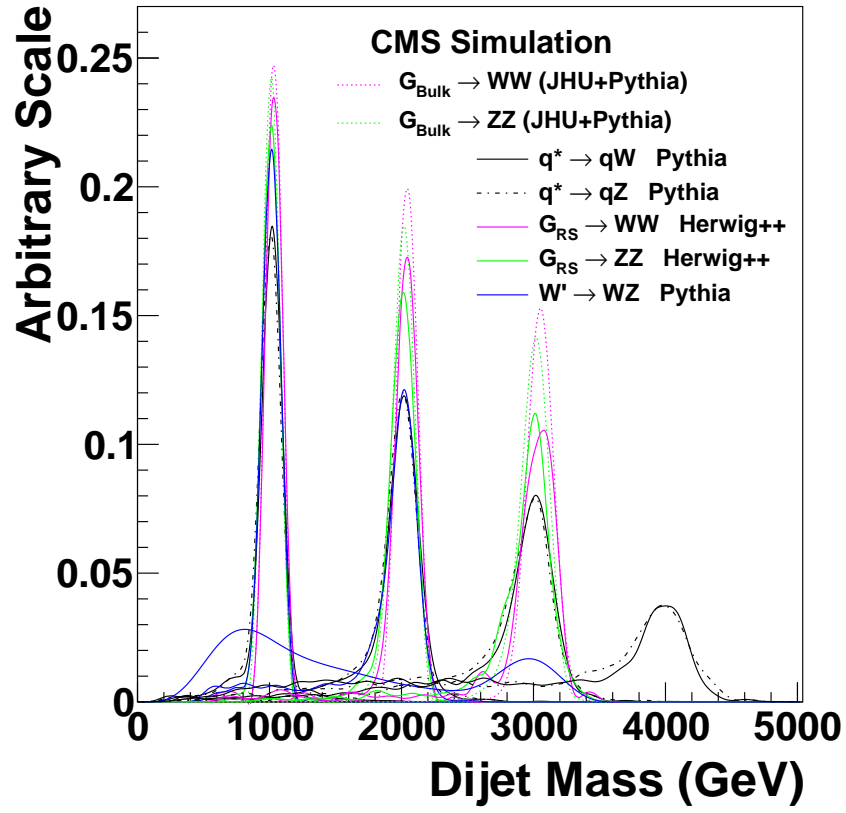


Figure 3.38: The normalized high purity signal resonance distribution for $G_{RS} \rightarrow WW$, $G_{RS} \rightarrow ZZ$, $W' \rightarrow WZ$, $q^* \rightarrow qW$, and $q^* \rightarrow qZ$ resonances of dijet invariant mass 1.0TeV, 1.5TeV, 2.0TeV, 3.0TeV, 4.0TeV.

3.8 Systematic uncertainties

The sources of systematic uncertainties are summarized as follows:

- Background-related systematic uncertainties: background parametrization.
- Signal-related systematics uncertainties: W/Z-tagging efficiency, Jet Energy Scale(JES), Jet Energy Resolution(JER), luminosity.

3.8.1 Background shape parametrization

We model the shape of the QCD background in the dijet spectrum using a simple parametrization which has been successfully deployed in previous searches in the dijet mass spectrum [?]. The background model is given in Eq. (??):

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2}}. \quad (3.1)$$

where m denotes the dijet mass and \sqrt{s} the pp center of mass energy. P_0 acts as a normalization parameter for the probability density function, and P_1, P_2 describe its shape. It was checked by a Fisher F-test that no additional parameter is not needed to describe the distributions.

Figure ?? and Figure ?? show the dijet mass spectra from single and double W/Z-tagged data fitted to Eq. (??) and the bottom panes show corresponding pull distributions, demonstrating the agreement between the background-only probability

density function and the data.

No sizeable deviation from the background-only hypothesis is seen, exclusion limits are set on the product of cross section, acceptance, and branching fraction for the five considered final states: qW , qZ , WW , WZ , and ZZ .

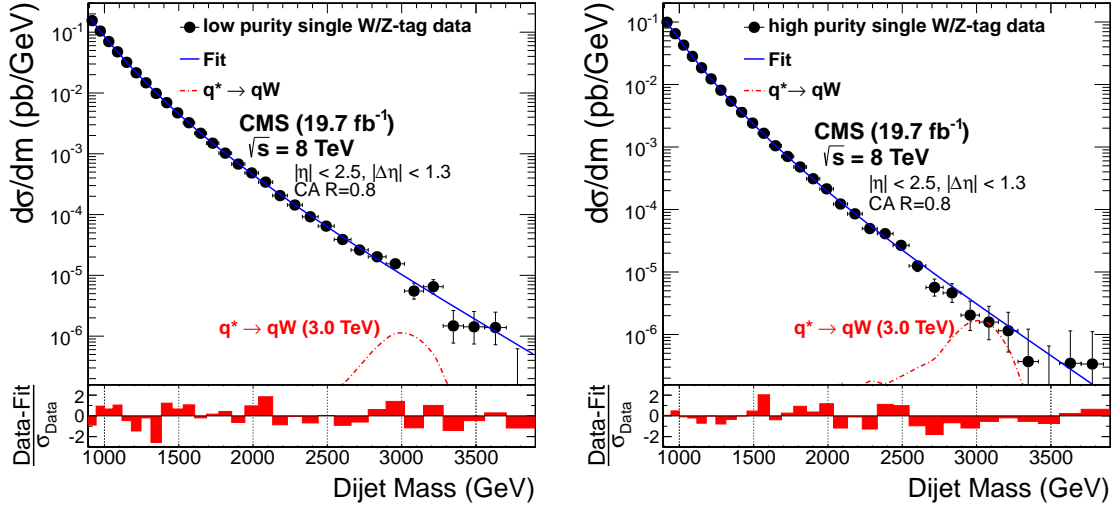


Figure 3.39: The medium purity (left) and high purity (right) single W/Z -tagged m_{jj} distributions (points) in data fitted with the QCD background parametrization (solid curve). Signal shape distribution for $q^* \rightarrow qW$ with its corresponding cross section is also shown. Bottom panes: the corresponding pull distributions ($\frac{\text{Data-Fit}}{\sigma_{\text{Data}}}$).

3.8.2 W/Z -tagging efficiency

The W/Z -tagging efficiency is determined from the Monte Carlo simulation. We cross-check the MC modelling of the signal efficiency by measuring the W/Z -tagging efficiency in semileptonic $t\bar{t}$ data, and compare it with the same efficiency obtained using identical procedure from $t\bar{t}$ Monte Carlo sample generated with MadGraph [?]

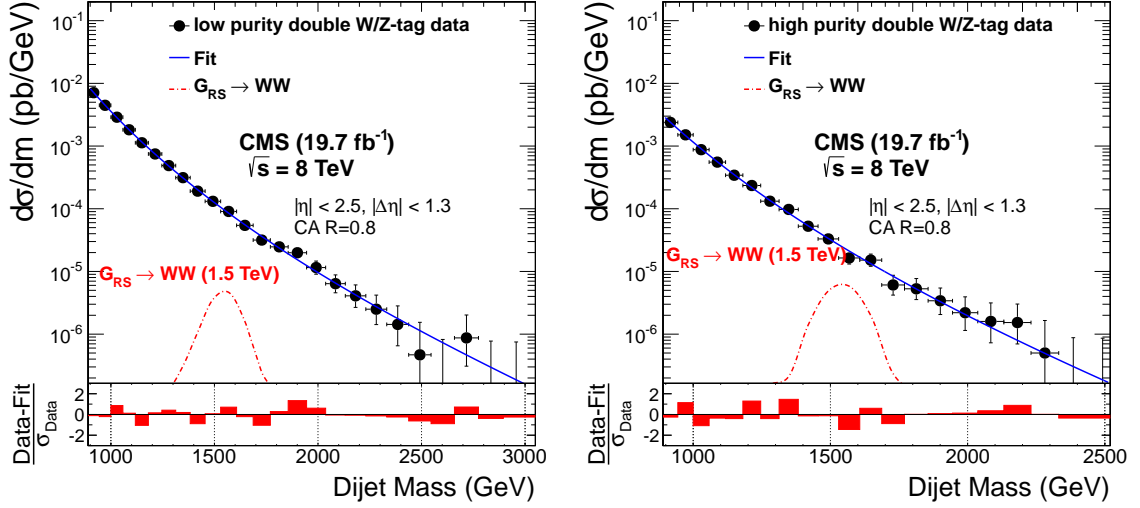


Figure 3.40: The medium purity (left) and high purity (right) double W/Z -tagged m_{jj} distributions (points) in data fitted with the QCD background parametrization (solid curve). Signal shape distribution for $G_{RS} \rightarrow WW$ with its corresponding cross section is also shown. Bottom panes: the corresponding pull distributions ($\frac{\text{Data-Fit}}{\sigma_{\text{Data}}}$).

and showered with Pythia6 Tune Z2*. The ratio of the two efficiencies defines a scale factor, which is then applied to the efficiencies for signals in the dijet data.

We follow the same procedure as described in Ref. [?]. Combining the efficiencies of the τ_{21} and jet mass cuts, a data-MC scale factor of 0.86 ± 0.07 (1.39 ± 0.75) for the high (low) purity selection for the W-tagging efficiency is determined. We assume that the same scale factor applies to Z-tagging as well. The errors on the scale factor are propagated into the systematic uncertainties on the overall signal efficiency.

The efficiency error on a single W/Z -tagging is estimated with a control sample of semileptonic $t\bar{t}$ events as described above. The uncertainties of 7.5% (54%) on the scale factors for high (low) purity tagging include sources from control sample

statistics, pruned jet mass scale and pruned jet mass resolution. Since we estimate the scale factor only in the kinematic regime of the $t\bar{t}$ sample where the W decay products merge, but the b -quarks are still reconstructed as separate jets, we need to rely on the simulation to extrapolate to higher jet p_T . Therefore, we estimate how the efficiency varies as a function of p_T for two different showering and hadronization models using G_{Bulk} samples generated with JHUGEN and interfaced with PYTHIA 6 and HERWIG++. We find that the differences are within 4% (12%) for the high (low) purity tagging and therefore smaller than the statistical uncertainty of the scale factor. Also, the dijet mass dependence of the W/Z -tagging efficiency for background events, shown in Fig. ?? and Fig. ??, is adequately described by the simulation. Other systematic errors on the tagging efficiency are small or negligible. Because of the rejection of charged particles not originating from the primary vertex and also the application of pruning, the pileup dependence on the W/Z -tagging efficiency is weak, and the uncertainty of the modeling of the pileup distribution is less than 3%. Modeling of the underlying event, estimated by switching it off in PYTHIA 6, impacts the tagging efficiency by less than 1%. These systematic errors refer to a single W/Z -tagged jet and are applied twice for double W/Z -tagged events.

3.8.3 Other uncertainties

In the jet p_T and η regions considered in this analysis, the Jet Energy Scale is known to a precision of 1-2% [?, ?]. The p_T and η dependent uncertainty is propagated to

the reconstructed dijet invariant mass, resulting in an uncertainty of 1% to a good approximation independent of the reconstructed dijet invariant mass. It is taken into account by shifting the resonance dijet mass in the statistical analysis. The Jet Energy Resolution(JER) is known to a precision of 10% and its tails are in agreement between data and MC [?]. It is taken into account in the statistical analysis by a variation of the resonance width by 10%. The luminosity has been measured with an uncertainty of 2.6% [?], and is also taken into account in the statistical analysis.

3.9 Limit setting procedure

For setting upper limits on the resonance production cross section the asymptotic approximation [?] of the LHC CL_s method [?, ?] is used. The binned likelihood, L , can be written as:

$$L = \prod_i \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} , \quad (3.2)$$

where

$$\mu_i = \sigma N_i(S) + N_i(B) , \quad (3.3)$$

n_i is the observed number of events in the i^{th} dijet mass bin, and $N_i(S)$ is the expected number of events from the signal in the i^{th} dijet mass bin, σ scales the signal amplitude, and $N_i(B)$ is the expected number of events from background in the i^{th} dijet mass bin. The background $N_i(B)$ is estimated as the background component of the best signal+background fit to the data points. The dominant sources of systematic uncertainties (the jet energy scale, the jet energy resolution, the integrated luminosity, and the W/Z -tagging efficiency) are considered as nuisance parameters associated to log-normal priors. The dependence of the likelihood on their value is removed through profiling. The ratio of the profiled likelihood at a given value σ^* over the maximum of the likelihood (for $0 < \sigma < \sigma^*$) is used as a test statistics to compute the CL_s value associated to σ^* . This allows to determine the 95% confidence-level (CL) limit.

3.10 Results

Fig. ?? , ?? and ?? shows the 95% CL cross section upper limits derived from the single and double W/Z tagged event samples. The predicted cross sections as a function of resonance mass for the considered benchmark models are overlaid. Table ?? shows the results of limits and also corresponding limits from 7TeVdata [?]. The observed local p-values for each model are shown in Fig. ?? , ?? , ?? and ??. The largest local significance for a qW(WW) signal of $2.0(1.3)\sigma$ is found at 1.5(1.9)TeV. An estimate of the look-else-where effect to transform the local p-values into global p-values is shown in Fig. ??.

Process	Observed Mass Exclusion(TeV)		Expected Mass Exclusion(TeV)	
	8 TeV	7 TeV	8 TeV	7 TeV
$q^* \rightarrow qW$	[1.00, 3.17]	[1.00, 2.38]	[1.00, 2.98]	[1.00, 2.43]
$q^* \rightarrow qZ$	[1.00, 2.88]	[1.00, 2.15]	[1.00, 2.63]	[1.00, 2.07]
$G_{RS} \rightarrow WW$	[1.00, 1.23]	-	[1.00, 1.33]	-
$G_{RS} \rightarrow ZZ$	-	-	-	-
$W' \rightarrow WZ$	[1.00, 1.23], [1.39, 1.52], [1.57, 1.61]	-	[1.00, 1.44]	-

Table 3.7: Summary of limit setting results.

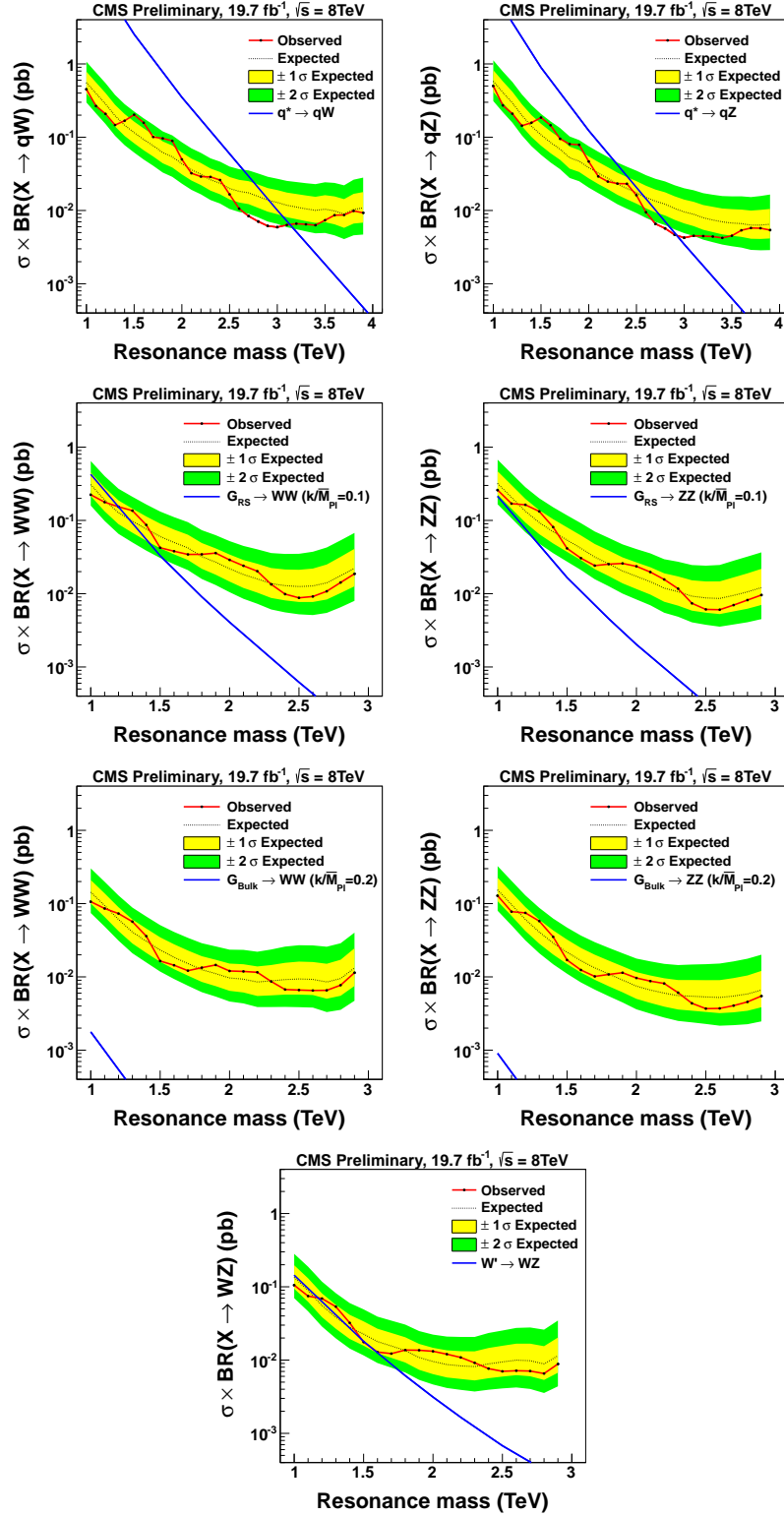


Figure 3.41: Expected and observed limits for qW (top-left), qZ (top-right), $G_{RS}WW$ (center-left), $G_{RS}ZZ$ (center-right), $G_{Bulk}WW$ (center-left), $G_{Bulk}ZZ$ (center-right) and WZ (bottom) resonances in the high-purity category. The predicted cross sections as a function of resonance mass for the considered benchmark models are overlaid.

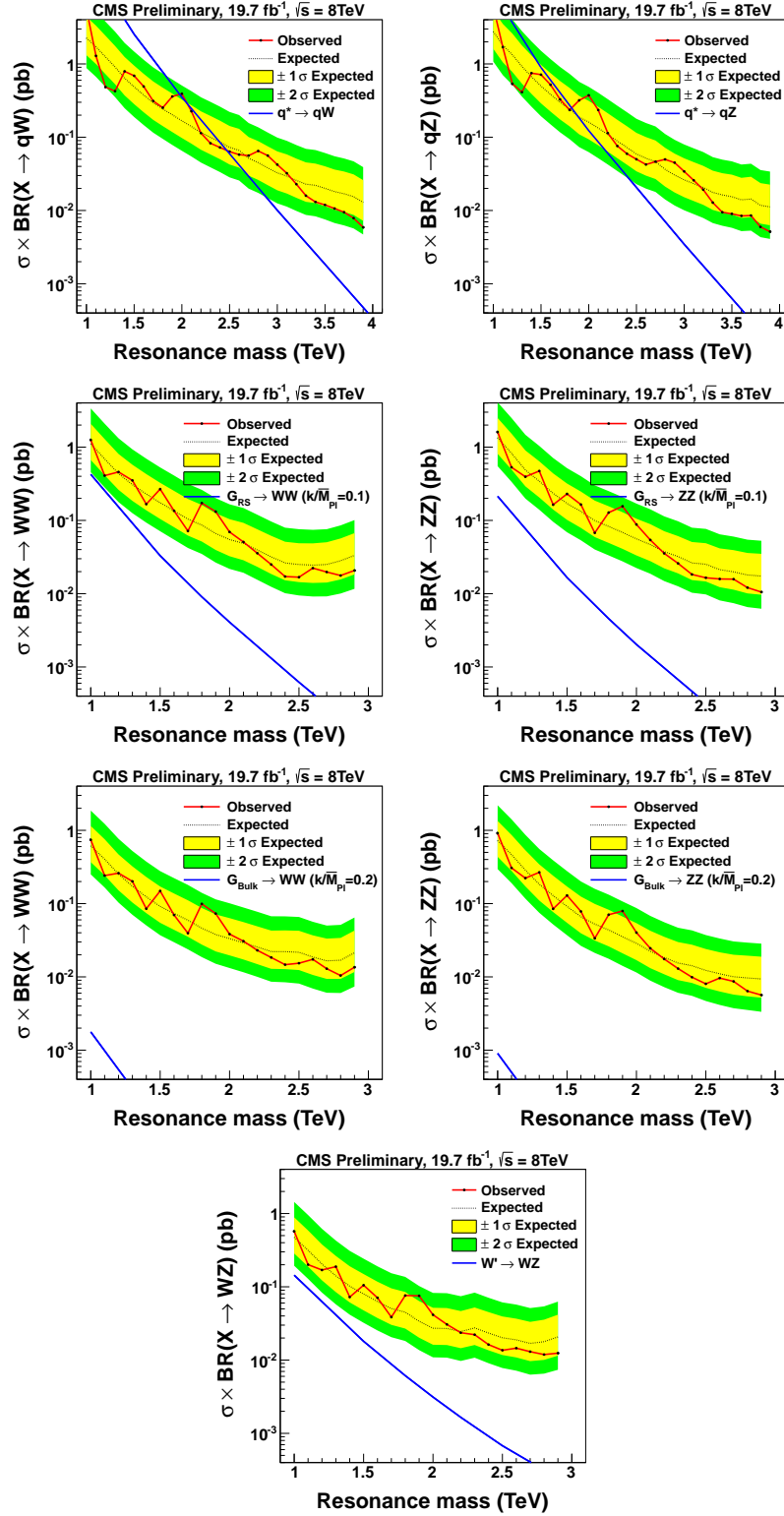


Figure 3.42: Expected and observed limits for qW (top-left), qZ (top-right), $G_{RS}WW$ (center-left), $G_{RS}ZZ$ (center-right), $G_{Bulk}WW$ (center-left), $G_{Bulk}ZZ$ (center-right) and $W' \rightarrow WZ$ (bottom) resonances in the low-purity category. The predicted cross sections as a function of resonance mass for the considered benchmark models are overlaid.

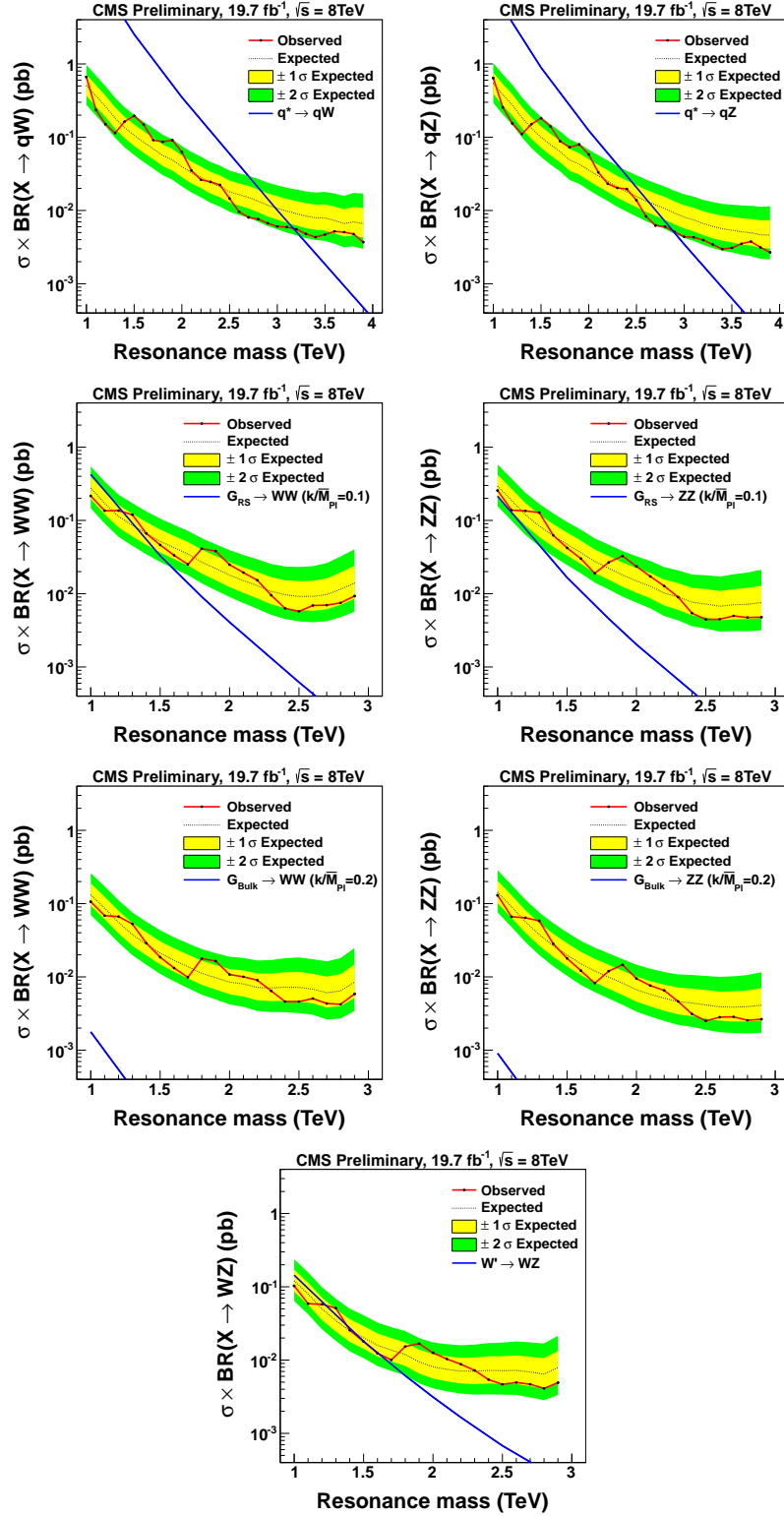


Figure 3.43: Expected and observed limits for qW (top-left), qZ (top-right), $G_{\text{RS}} WW$ (center-left), $G_{\text{RS}} ZZ$ (center-right), $G_{\text{Bulk}} WW$ (center-left), $G_{\text{Bulk}} ZZ$ (center-right) and WZ (bottom) resonances combining the low-purity and high-purity categories. The predicted cross sections as a function of resonance mass for the considered benchmark models are overlaid.

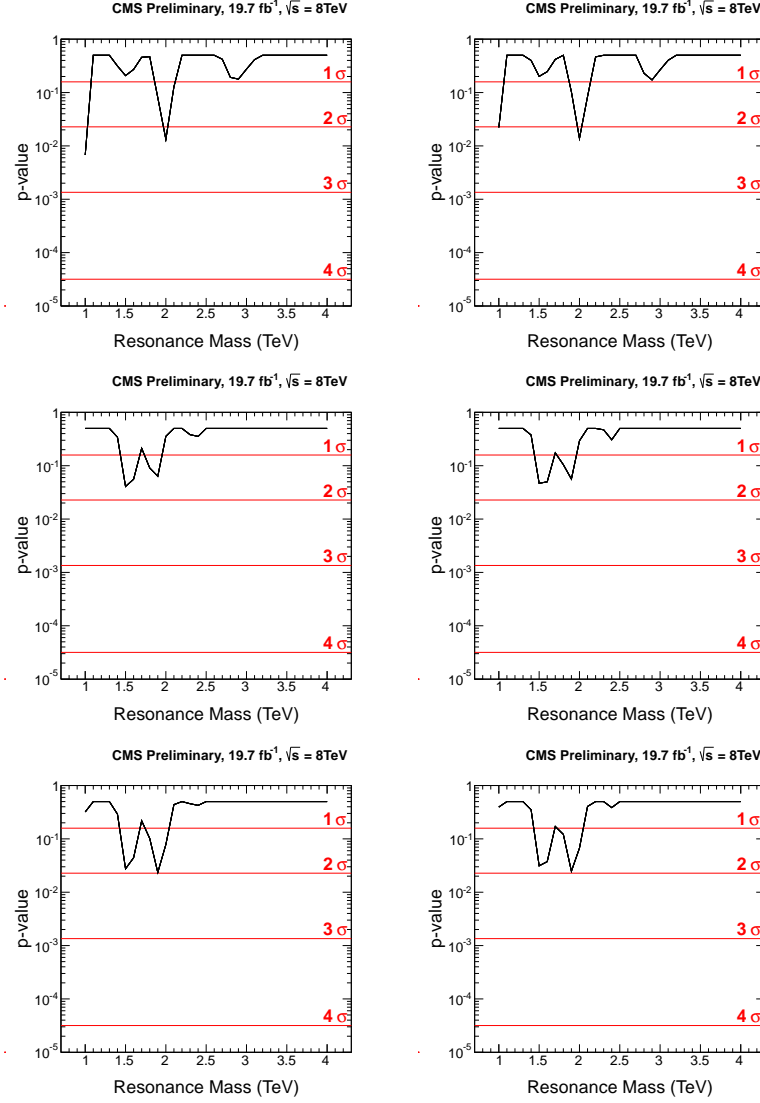


Figure 3.44: Observed local p-values assuming a qW (left) and qZ (right) signal model in the singly-tagged dijet mass spectrum in the low-purity (top), high-purity (middle) and combination (bottom).

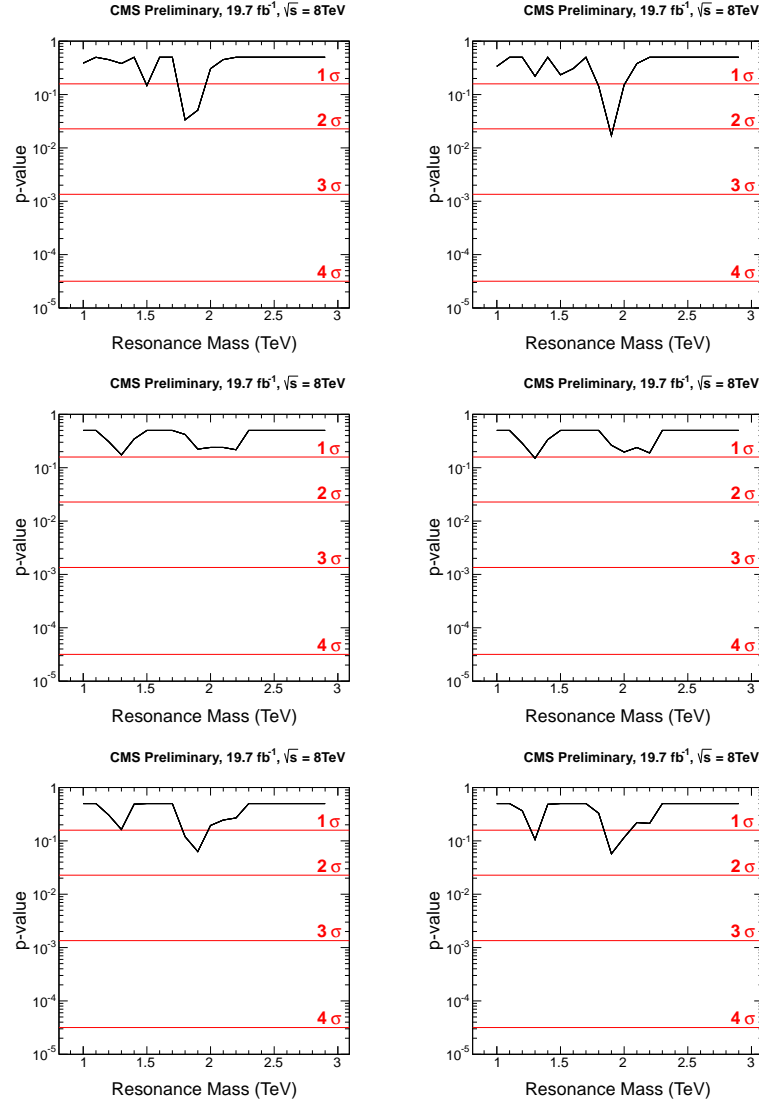


Figure 3.45: Observed local p-values assuming a $G_{\text{RS}} \text{WW}$ (left) and $G_{\text{RS}} \text{ZZ}$ (right) signal model in the doubly-tagged dijet mass spectrum in the low-purity (top), high-purity (middle) and combination (bottom).

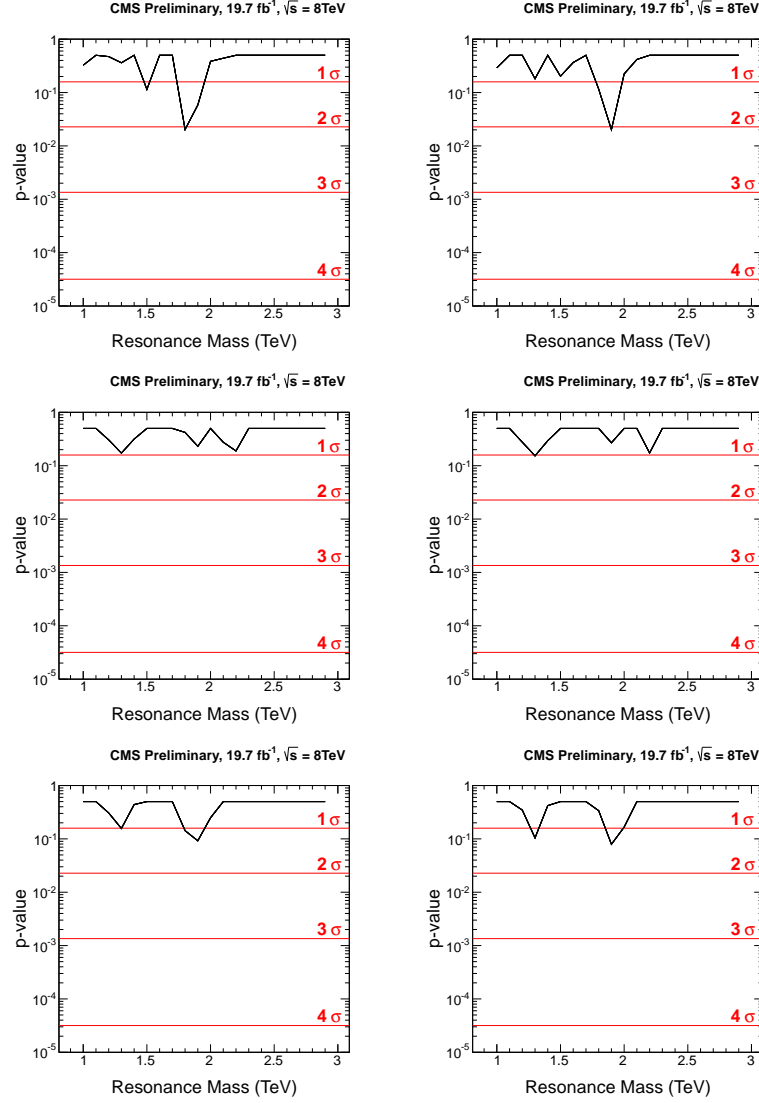


Figure 3.46: Observed local p-values assuming a $G_{\text{Bulk}} WW$ (left) and $G_{\text{Bulk}} ZZ$ (right) signal model in the doubly-tagged dijet mass spectrum in the low-purity (top), high-purity (middle) and combination (bottom).

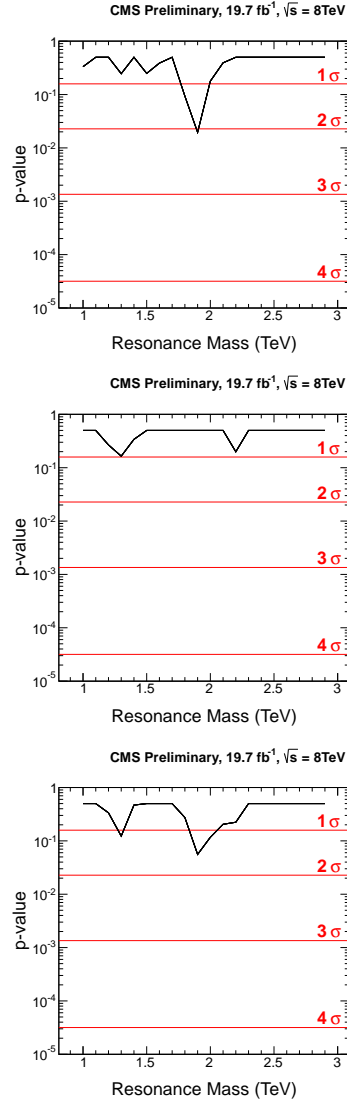


Figure 3.47: Observed local p-values assuming a WZ signal model in the doubly-tagged dijet mass spectrum in the low-purity (top), high-purity (middle) and combination (bottom).

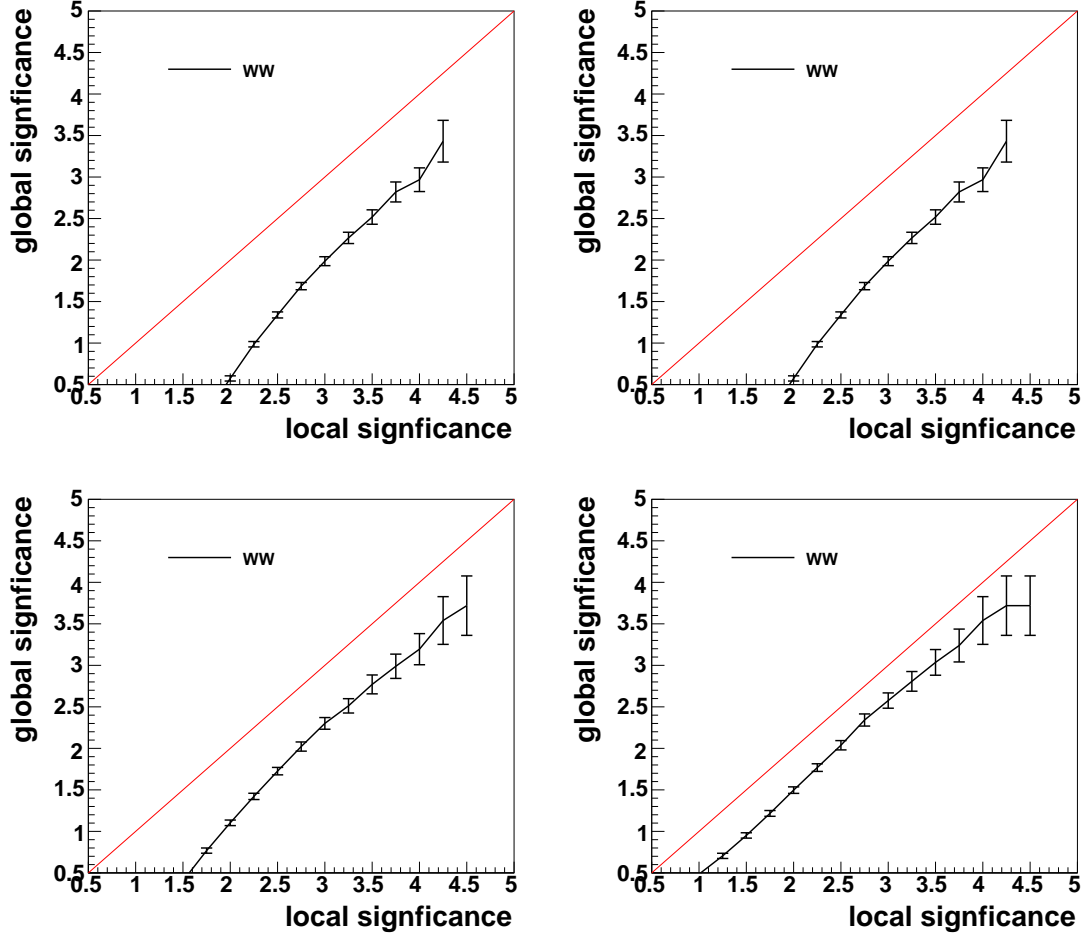


Figure 3.48: Estimate of the look-elsewhere effect for this search. Shown is the global significance as a function of the local significance which corresponds to the maximal significance in the dijet mass distribution in the mass ranges 1.0-2.3 TeV (top-left), 1.2-2.3 TeV (top-right), 1.5-2.0 TeV (bottom-left) and 2.0-2.3 TeV (bottom-right). The global significance is estimated using background-only toys and corresponds to the fraction of toys (translated from a p-value to significance) with at least a certain local significance in the dijet mass distribution.

3.11 Conclusions

A data sample corresponding to an integrated luminosity of 19.7 fb^{-1} collected in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ with the CMS detector has been used to measure the W/Z -tagged dijet mass spectrum using the two leading jets within the pseudorapidity range $|\eta| < 2.5$ and with pseudorapidity separation $|\Delta\eta| < 1.3$. The QCD background is suppressed using jet substructure tagging techniques, which identify vector bosons decaying into hadrons. In particular, we use the invariant mass of pruned jets and the N -subjettiness ratio τ_{21} to discriminate against the initially overwhelming QCD background. The remaining QCD background is estimated from a fit to the smooth parameterized shape. We have searched for the signal as a peak on top of the smoothly falling QCD background. No evidence has been found for new resonance production in the W/Z -tagged dijet spectrum. A 95% CL lower limit is set on the mass of excited quark resonances decaying into qW (qZ) at 3.17 (2.88) TeV. G_{RS} decaying into WW is excluded up to 1.24 TeV. W' decaying into WZ is excluded in the ranges $[1.00, 1.23]$, $[1.39, 1.52]$ and $[1.57, 1.61]$ TeV.

Chapter 4

EXO-14-009

4.1 Introduction

Several physics models beyond the standard model (SM) predicts the existence of vectorial resonances with masses above 1TeV that decay into a W or Z vector boson, denoted as V boson, and a Higgs boson [?, ?] . In proton-proton (pp) collisions at the energies reached at the Large Hadron Collider (LHC), bosons emerging from such decays usually would have sufficiently large momenta so that the hadronization products of their decays would merge into a single massive jet [?]. We present a search for events containing jets of this kind in pp collisions at a centre-of-mass energy of $\sqrt{s} = 8\text{TeV}$. The data sample, corresponding to an integrated luminosity of 19.7fb^{-1} , was collected with the CMS detector at the LHC.

The signal of interest is a heavy vector resonance X, produced via the process ,

$pp \rightarrow X \rightarrow HV$, as in Figure ??, where the V boson decay hadronically and the Higgs boson decay either to a b -quark pair or to a pair of fully hadronic W bosons¹.

The results are interpreted in the following models of $W' \rightarrow HW$ and $Z' \rightarrow HZ$ resonances. The Composite-Higgs [?, ?, ?] and Little Higgs models [?] provide a direct solution to the hierarchy problem and predict many new particles, including additional gauge bosons, e.g. heavy W' or Z' bosons. Models of such type are generalized in the Heavy Vector Triplet (HVT) model [?]. Of particular interest for this search is the so called HVT scenario B model. The W' and Z' decay to respective WH and ZH become dominant and almost degenerate with $W' \rightarrow WZ$, such that these parameter points are not constrained so far from experiments (see Fig 3.2 in Ref. [?]). For W' with the SM coupling, the most stringent limits are reported in searches with leptonic final states [?, ?], and the current lower limit on the W' mass is 2.9TeV. The limit varies by 0.1TeV, depending on the chirality of the W' couplings. Specific searches in the WZ final state have also been reported [?, ?, ?] setting a lower limit of 1.1TeV. For Z' with the SM coupling, the most stringent limits are reported in searches with leptonic final states [?, ?], and the current lower limit on the Z' mass is 2.8TeV.

The signal is characterized by a peak in the dijet invariant mass distribution m_{jj} over a continuous background from SM processes, comprised mainly of multijet events from quantum chromodynamic (QCD) processes. The sensitivity to jets from Higgs decays to b -quark pair is enhanced through subjects b -tagging [?]. While jets from

¹As the Higgs boson branching ratio to Z bosons is about one-tenth of its branching ratio to W bosons, we only consider the $H \rightarrow WW$ decay.

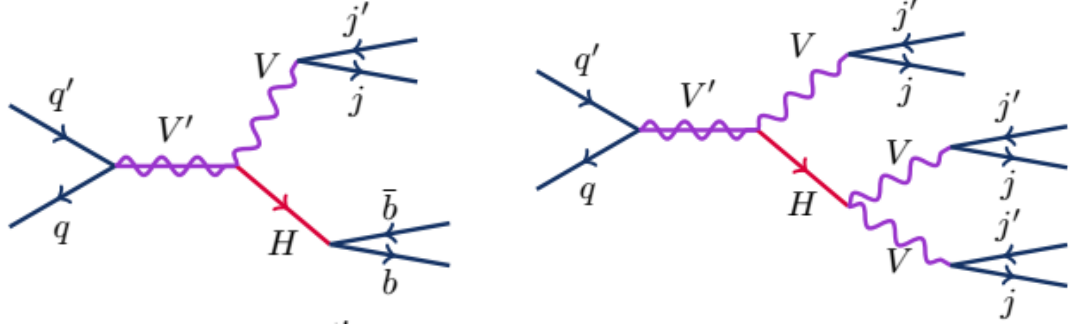


Figure 4.1: X decays.

Higgs decays to hadronic W-pairs, and also jets from W or Z bosons are identified with jet-substructure techniques [?, ?]. Results for W' and Z' , which decays to Higgs boson and one vector boson V , are presented in this analysis.

This is the first search for the heavy resonances decaying to all hadronic HV final states, as well as the first application of the highly boosted $H \rightarrow WW^* \rightarrow 4q$ decays.

This analysis proceeds via the following steps:

1. The search is performed in the dijet sample, using the same preselection as the standard search for resonances decaying to dijets [?, ?].
2. We identify events with W or Z: in each jet which is a candidate to originate from merging of V daughter jets:
 - we require a pruned jet mass cut, and
 - an N-subjettiness cut preferring two-prong decays

(This is identical to EXO-12-024 [?].)

CHAPTER 4. EXO-14-009

3. We identify events with a highly boosted Higgs boson:

- we require a pruned jet mass cut, and
- two b tagged subjects, or
- (when there are no two b tagged subjects) a N-subjettiness cut preferring four subjects

(The $H \rightarrow b\bar{b}$ tagging is synchronized with our sister analysis, the Radion search to the HH final state [?].)

4. After the full event selection, a potential signal would be characterized as a peak in the dijet invariant mass, on top of a falling background distribution.
5. We model the background distribution with a smoothly falling analytical function. (The functional form is identical to the one used in EXO-12-024.)
6. We form the joint likelihood of several dijet distributions of V tagged and H tagged jets. We include both two types of Higgs tags, and also low-purity Higgs and V taggers. The background estimate procedure is the same in all channels – analytical parameterization – but is performed separately for each channel.
7. Finally, we set the limits on the various simplified models for resonances decaying to HV final states.

4.2 Data and Monte Carlo samples

The data sample of proton-proton collisions at $\sqrt{s} = 8$ TeV was collected in 2012 and corresponds to an integrated luminosity of 19.7 fb^{-1} . The datasets and also the certifications used are summarized in Table ???. The dijet sample is dominated by light flavored and gluon jets, which we denote as the "QCD background".

We list part of our monte carlo simulated signal (from 1 TeV to 2.6 TeV) in Table ???. Signal samples are generated exclusively of the specific Higgs decay mode and W/Z decay mode. Model parameters and detailed cross sections are summarized in Appendix ???. The matrix element is calculated with Madgraph 5.1.5.11 [?]. The signals of interest, are showered and hadronized with PYTHIA 6.426 [?], and HERWIG++ 2.5.0 [?], using simulation of the CMS detector, based on GEANT4 [?]. Tune Z2* [?] is used in PYTHIA, while the version 23 tune [?] is used in HERWIG++. The CTEQ61L [?] parton distribution functions (PDF) are used for PYTHIA and the MRST2001 [?] leading-order (LO) PDF for HERWIG++. W' and Z' are generated with resonance widths at $\approx 4\%$ of the resonance mass, slightly smaller than the experimental resolution in m_{jj} for resonance masses considered in the analysis. Samples showered from PYTHIA are used in the analysis. While, samples from HERWIG++ are used to evaluate the systematic uncertainty by comparing the difference of hadronization from PYTHIA.

Dataset
/Jet/Run2012A-22Jan2013-v1/AOD
/JetHT/Run2012B-22Jan2013-v1/AOD
/JetHT/Run2012C-22Jan2013-v1/AOD
/JetHT/Run2012D-22Jan2013-v1/AOD

Table 4.1: Summary of 8 TeV collision data used in this analysis. The certification file used for these data is `Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt` .

Process	mass (GeV)	Events	X-sec[pb]
$Z' \rightarrow HZ$	1000	20000	8.56E-02
	1500	20000	1.19E-02
	2000	20000	1.93E-03
	2500	20000	3.39E-04
$W' \rightarrow HW$	1000	20000	1.71E-01
	1500	20000	2.55E-02
	2000	20000	4.25E-03
	2500	20000	7.31E-04

Table 4.2: Examples of the simulated Monte Carlo samples used in this analysis for process $V' \rightarrow VH$. Cross sections are calculated from production cross sections of V' times its $BR(W' \rightarrow HW$ or $Z' \rightarrow HZ)$. These samples are generated using Madgrap5 and hadronized with Pythia6.

4.3 Trigger

Events are selected if one of the following triggers has fired: HLT_HT750, HLT_PFHT650, HLT_PFNpUHT650 HLT_FatDiPFJetMass750_DR1p1_Deta1p5. All versions of each of these triggers is used. None of these triggers are prescaled during the 2012 data taking period. HLT_PFNpUHT650 trigger is used for the data set after the RunC(including RunC), while HLT_PFHT650 trigger is only used for RunA and RunB data sets.

Figs ??, ?? and ?? show the trigger efficiency. The trigger efficiency has been measured with respect to a lower-threshold, but prescaled, HLT_HT550 trigger. The trigger is 99% efficient above 890GeV for the untagged, HbbVqq-tagged, and HwwVqq-tagged data.

Fig ?? shows the turn-on curve of the reference trigger on the signal MC. 1.0 TeV signals are used here in Fig ??, and the plot shows HwwVqq and HbbVqq signals are, fully efficient for HLT_HT550 trigger, which is not prescaled in MC. So other signals, having resonance mass bigger than 1.0 TeV, will surely be fully efficient for the triggers.

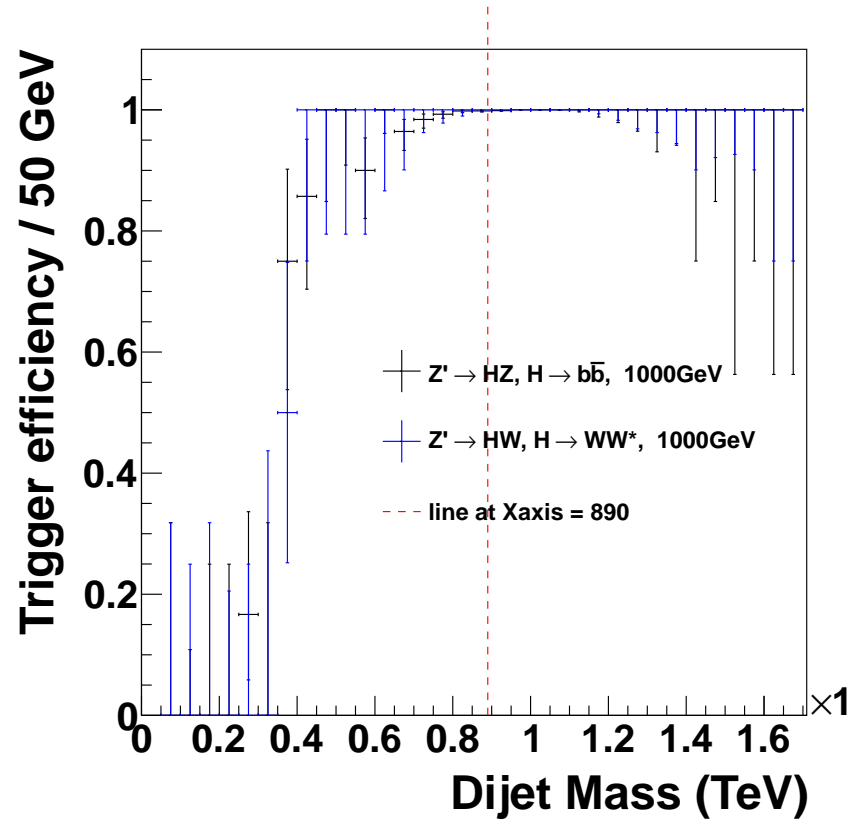


Figure 4.2: Reference trigger efficiency of signal MC.

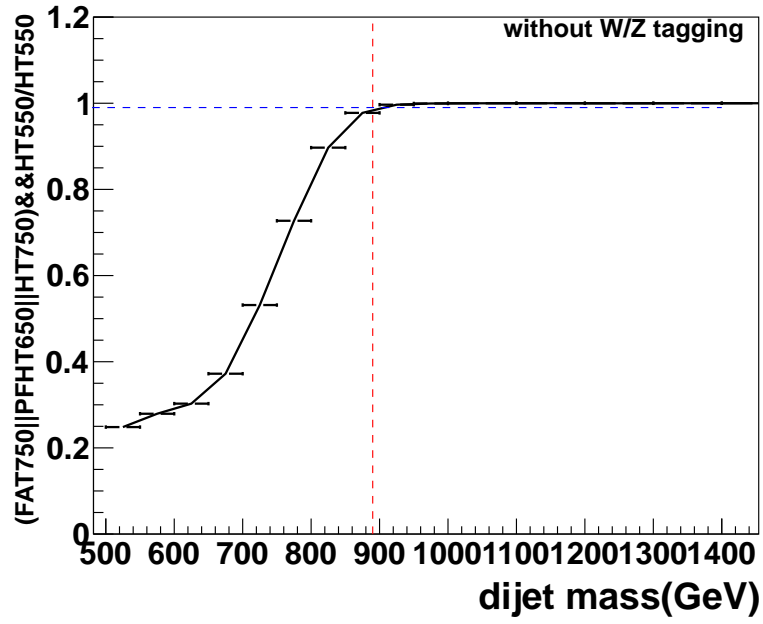


Figure 4.3: trigger efficiency for untagged data of $\text{fat}_{750}||\text{hlt_pf}(\text{nopu})\text{ht}_{650}||\text{hlt_ht}_{750}$ measured using data collected by lower threshold h_t550 trigger. the dashed red line is drawn at m_{jj} equal 890 GeV, the blue line is at efficiency at 99%.

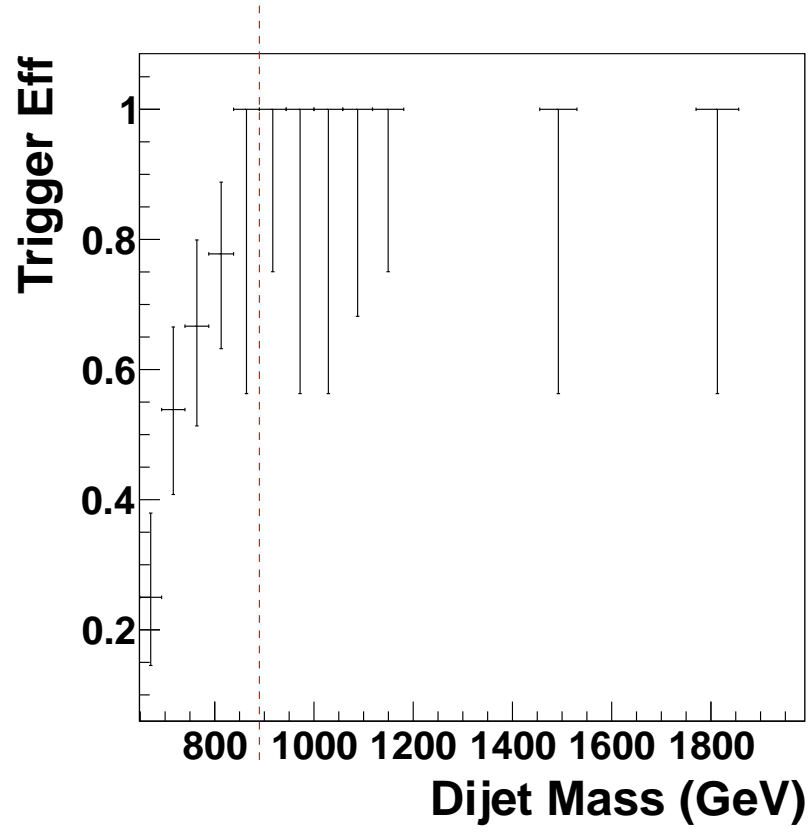


Figure 4.4: Trigger efficiency for HbbVqq tagged data of FAT_750||HLT_PF(NoPU)HT650||HLT_HT750 measured using data collected by lower threshold H_T550 trigger. The dashed red line is drawn at m_{jj} equal 890 GeV.

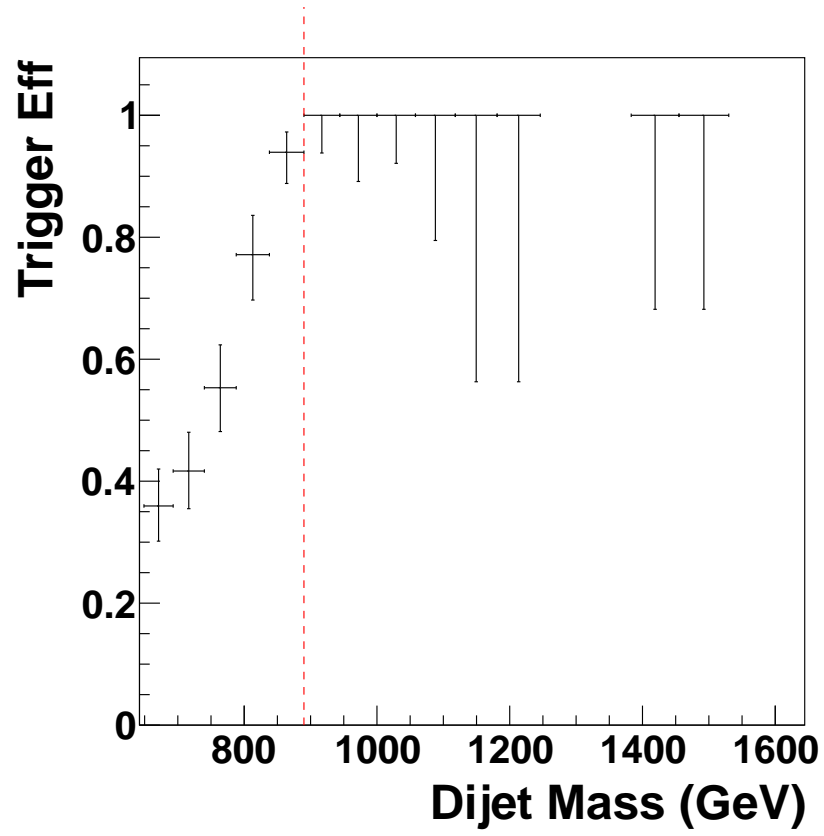


Figure 4.5: Trigger efficiency for HqqqqVqq tagged data of FAT_750||HLT_PF(NoPU)HT650||HLT_HT750 measured using data collected by lower threshold H_T550 trigger. The dashed red line is drawn at m_{jj} equal 890 GeV.

4.4 Event preselection

4.4.1 Jet reconstruction

Based on CMSSW 5.3.x software package, events are reconstructed using the particle-flow reconstruction algorithm [?] that attempts to reconstruct stable particles in an event by combining information from all subdetectors. The algorithm categorizes all particles into five types: muons, electrons, photons, charged and neutral hadrons. The resulting particle flow candidates are passed to Cambridge-Aachen (CA) [?, ?] jet clustering algorithm, as implemented in FastJet version 3.0.1 [?, ?], to create "particle flow jets". The CA clustering sequence is only determined by the distance between clusters and is not weighted by their momentum, as is done for the k_T and anti- k_T algorithms. A distance parameter of size $R = 0.8$ is used for the CA algorithm. These jets are hereafter referred to as CA8 jets.

Charged hadrons identified as pileup are removed from the list of PF candidates input to the jet clustering algorithms. The remaining neutral component of pileup is removed by applying a residual area-based correction as described in Ref. [?, ?]. The mean p_T per unit area is computed with the k_T algorithm with the "active area" method, with a distance parameter of 0.6, and the jet energy is corrected by the amount of pileup expected in the jet area. The amount of energy expected from the underlying event is added back into the jet. The pileup-subtracted jet four momenta are finally corrected for nonlinearities in η and p_T with simulated data, with

a residual η -dependent correction added to correct for the difference in simulated and true responses [?, ?].

The jet energy corrections for the CA8 jets are derived from studies using the anti- k_T $R = 0.7$ jet algorithm. Simulation studies confirm that these anti- k_T -derived jet corrections are adequate for the CA8 jet algorithm for the jet momenta considered here [?].

4.4.2 Event cleaning

Events are selected using the following cuts:

- The event must have a well reconstructed primary vertex as computed by a deterministic annealing filter (DAF) ($|z_{\text{Primary Vertex}}| < 24$ cm, $N_{\text{DOF}} > 4$).
- The following recommended noise event filters are used:
 - CSC tight beam halo filter
 - HBHE noise filter with isolated noise rejection
 - HCAL laser event filter (HBHE) and HCAL laser event filter 2012
 - ECAL dead cell trigger primitive (TP) filter
 - Bad EE supercrystal filter
 - The tracking failure filter
 - Good primary vertex filter

CHAPTER 4. EXO-14-009

- Tracking coherent noise filter
- Tracking TOBTEC fakes filter
- The events are required to have at least two ungroomed CA8 jets with
 - $p_T > 30 GeV$, $|\eta| < 2.5$
 - to have muon energy fraction < 0.8
 - pass tight particle flow jet ID. The tight PF jet ID is listed below:
 - * Neutral Hadron (EM) Fraction $< 0.90(< 0.90)$, for all jet η
 - * Number of Constituents > 1 , for all jet η
 - * Charged Hadron (EM) Fraction $> 0(< 0.99)$, for jet $|\eta| < 2.4$
 - * Charged Multiplicity > 0 , for jet $|\eta| < 2.4$
- Beam background events are removed using the following requirements:
 - In events with at least 10 tracks, a minimum of 25% of these tracks must be high purity tracks.
- The events must pass $|\Delta\eta| < 1.3$, $m_{jj} > 890 GeV$

This sample of dijet events is then searched for W/Z and Higgs bosons. One jet must be identified as a hadronically decaying W or Z boson, whereas the other jet must be identified as one of the two types of hadronic Higgs decays.

4.5 The H tagging and W/Z tagging algorithms

The products of hadronic decays of Higgs, W, and Z bosons can fall within a single jet if these particles are highly boosted. In this analysis, we aim to cover as much of the Higgs branching ratio as possible. The Standard Model Higgs with a mass of 125 GeV decays to $b\bar{b}$ with a branching fraction of 57.7%, and to WW^* with a branching fraction of 21.4%. [?]. Using these two decay modes in a VH search, where WW^* specifically decays to four quarks, is the main topic of this note. (The semileptonic decay mode $H \rightarrow WW \rightarrow 2q\ell\nu$ is viable, but its reconstruction is more involved and will be covered in a subsequent analysis.)

The algorithms to identify W/Z, $H \rightarrow b\bar{b}$ and $H \rightarrow WW^*$ jets are necessarily different, but they use similar jet-level variables: N-subjettiness (described in Section ??) and jet pruning (Section ??). The W/Z-tagger is described in Section ??, and the two H-taggers in Sections ?? and ??.

4.5.1 N-subjettiness

N-subjettiness $[\tau_1, \tau_2, \tau_3]$ exploits the fact that the pattern of the hadronic decay of a heavy object is reflected through the presence of distinctive energy lobes corresponding to the decay products, as opposed to QCD jets which present a more uniformly

spread energy configuration. The inclusive jet shape N-subjettiness is defined, in its generalized version [?], as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min((\Delta R_{1,k})^\beta, (\Delta R_{2,k})^\beta \dots (\Delta R_{N,k})^\beta) \quad (4.1)$$

where the index k runs over the jet constituents and the distances $\Delta R_{n,k}$ are calculated with respect to the axis of the n^{th} subjet. The normalization factor d_0 is calculated as $d_0 = \sum_k p_{T,k} R_0^\beta$, setting R_0 to the jet radius of the original jet. In the analysis, we use onepass_kt_axes definition of subjet axes and the N-subjettiness is calculated from the unpruned jets with the parameter $\beta = 1$. It has been shown in the literature [?] that the best way to separate (N+1)-prong from the N-prong decays merged into a single jet is to select jets with low value of the ratio τ_{N+1}/τ_N . In this vein, the variable able to best discriminate between W/Z jets and QCD jets is $\tau_{21} = \tau_2/\tau_1$. The distribution of τ_{21} for the VV signal and QCD background is given in Fig. ??.

4.5.2 Jet Pruning

Jet pruning consists of the removal of the softest components of the jets [?, ?]. The jet pruning algorithm uses the CA $R = 0.8$ jets as inputs. In the process, soft and wide-angle particles (relative to the parent in the clustering) are ignored and are not clustered. The same parameters are chosen for the jet pruning algorithm as in the original paper [?, ?].

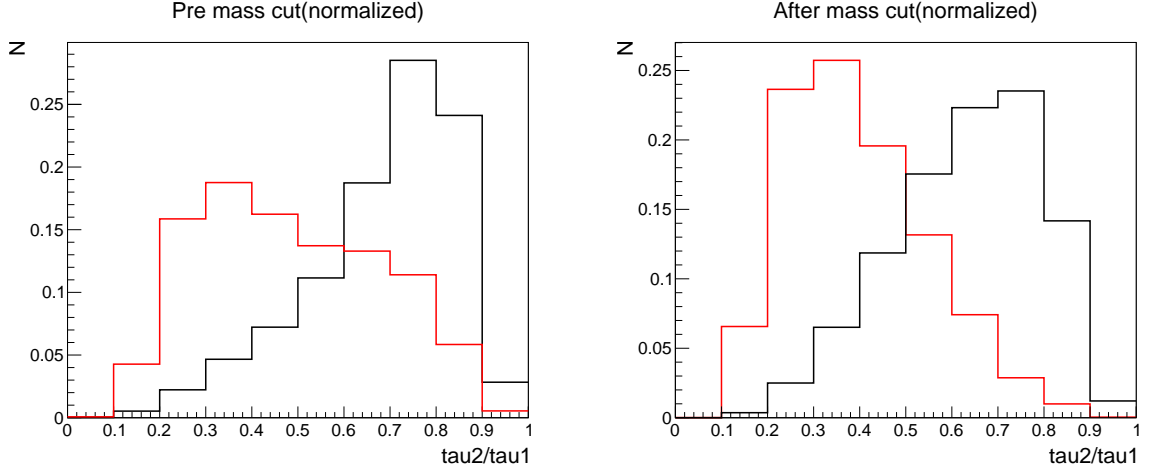


Figure 4.6: Comparison for τ_2/τ_1 distribution between signal (red) and background (black) before the jet mass cut (left) and after the jet mass cut of [70, 100] GeV applied (right). The signal MC used here is Herwig WW 1.5 TeV, and background is Herwig QCD.

The result of jet pruning on the CA8 jets is two fold, i.e., the invariant jet mass reconstruction and subjet identification. In all cases, we use the jet invariant mass computed from the whole (or “fat”) pruned jet. This quantity is referred below to as the pruned jet mass. For W/Z tagging, we use pruned jet mass between 70 and 100 GeV. For the identification of Higgs jets, we require the pruned jet mass to lie between 110 and 135 GeV. The distribution of the pruned jet mass of the Higgs candidate jet is shown on Fig. ??

The main role of jet pruning is to allow better delineation of subjets within the jet. In $H \rightarrow b\bar{b}$ tagger, the axes of the pruned subjets are used as the basis for b tagging.

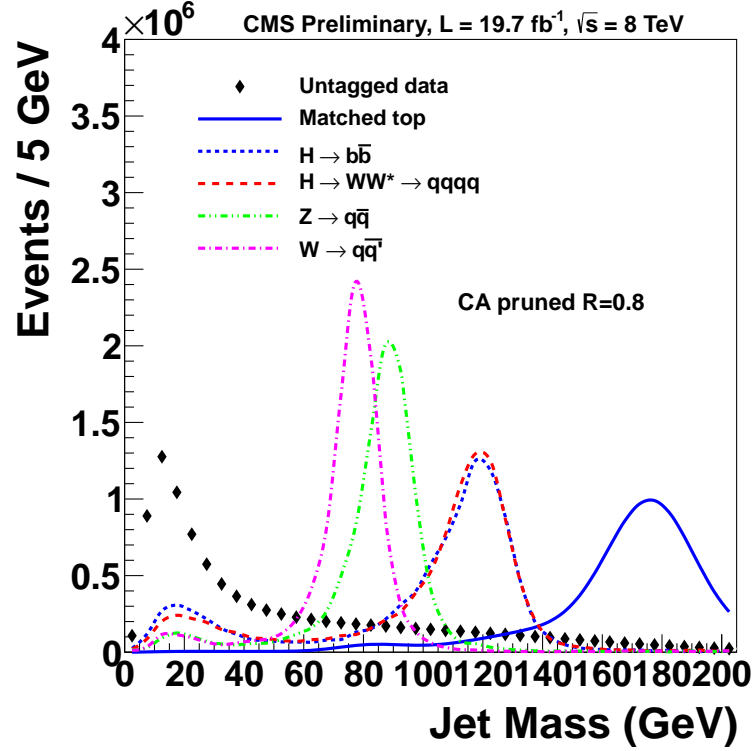


Figure 4.7: Pruned jet mass in signal MC, data and $t\bar{t}$ MC. MC samples are normalized to data. MC distributions are plotted as smooth curves connecting the histogram entries; the MC histograms have the same binning as the data. Higgs, W/Z and top jets are matched to their generator level particles, respectively.

4.5.3 W/Z tagging

For the identification of W/Z jets, we employ the same tagging algorithm previously used in published searches [?]. W/Z jets are selected using the following requirements:

- **Pruned jet mass m_{jet}** - Require the total pruned jet mass to satisfy $70\text{GeV} < m_{\text{jet}} < 100\text{GeV}$.

- **N-subjettiness** - We split the events into two categories, “high purity” W/Z jets by requiring $\tau_{21} \leq 0.5$, while $0.5 < \tau_{21} < 0.75$ defines the “low purity” W/Z jets. The thresholds are taken from the published VV search.

The performance of the W/Z tagger has been documented in detail in Ref. [?].

4.5.4 $H \rightarrow b\bar{b}$

To identify Higgs jets arising from the shower and hadronization of two collimated b quarks, we apply b tagging either on the two subjets or the fat jet, based on the angular separation of the two subjets, which is recommended by BTV-13-001 [?].

we use the following selection, synchronized with the radion search to $HH \rightarrow 4b$ [?] and the search for HW resonances in the semileptonic channel [?]:

- **Pruned jet mass m_{jet}** - Require the total pruned jet mass to satisfy $110\text{GeV} < m_{\text{jet}} < 135\text{GeV}$.
- **Subjet b -tagging**
 - if ΔR between the CA8 subjets is bigger than 0.3: *both* subjets must pass the CSV Loose working point.
 - if ΔR between the CA8 subjets is smaller than 0.3: require the *fat* CA8 jet to pass the CSV Loose working point.

4.5.5 $H \rightarrow WW^* \rightarrow 4q$

In this channel, Higgs decays to two W bosons, one real and one virtual (denoted with an asterisk). Given that this is effectively a three-body decay $H \rightarrow Wqq$, the jets from the four quarks are not on an even footing – the subjects from the real W are harder, and they also form a W mass. The subjects from the softer two quarks are less well defined.

A naive $H \rightarrow 4q$ tagger would require a fat jet with four subjects. However, a study done using the subjects as defined by the CMS Top Tagging algorithm (which reruns the CA8 jet clustering with additional weak pruning [?]) removes $\approx 90\%$ of the signal. Compounded with a decreasing angular separation between Higgs decay products, as a function of the Higgs p_T , at higher resonance masses, *e.g.* at 2TeV, only 1% of signal passes this selection. The distribution of the number of subjects of the reconstructed Higgs jets in signal MC (obtained from the CMS Top Tagging) is shown in Fig. ??

As an alternative, we explore the N-subjettiness, in particular the variables involving τ_4 . The ratio $\tau_{42} \equiv \tau_4/\tau_2$ has the best separation between the $H \rightarrow 4q$ signal and not only QCD background, but also Z and top jets. Figs ?? and ?? show the discriminating power of τ_{42} against $t\bar{t}$ and QCD, for 1 TeV and 2 TeV resonance masses respectively.

We also explore other combinations of $\tau_{NM} \equiv \tau_N/\tau_M$, which are listed in Appendix. ?. The ROC (receiver operating characteristic) curve of for several τ_{NM}

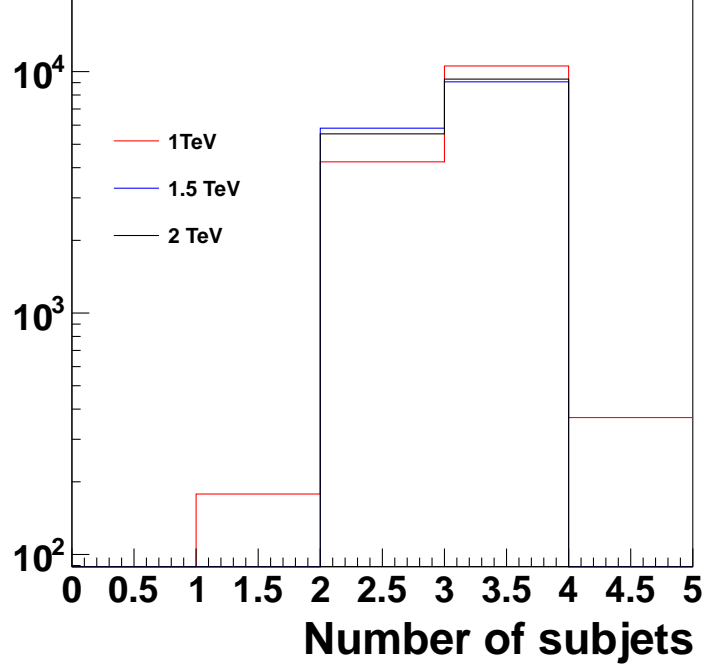


Figure 4.8: Number of subjects of the Higgs jets, in $H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$ signal MC. Subjects are obtained from the CMSTopTag jet collection.

cuts (but the same pruned jet mass cut) is shown in Fig. ???. The signal efficiency is evaluated using Higgs jets in 2 TeV signal MC, and the false positive rate (*i.e.*, mistag rate) is derived from QCDPT300to470 MC sample. From the figure, it is clear that τ_{42} outperforms any other single τ_{NM} variable.

After optimizing the cut on τ_{42} (documented in Sec. ?? below), the full selection of the $H \rightarrow WW^* \rightarrow 4q$ tagger is:

- **Pruned jet mass m_{jet}** - We require the total pruned jet mass to satisfy

$$110\text{GeV} < m_{\text{jet}} < 135\text{GeV}.$$

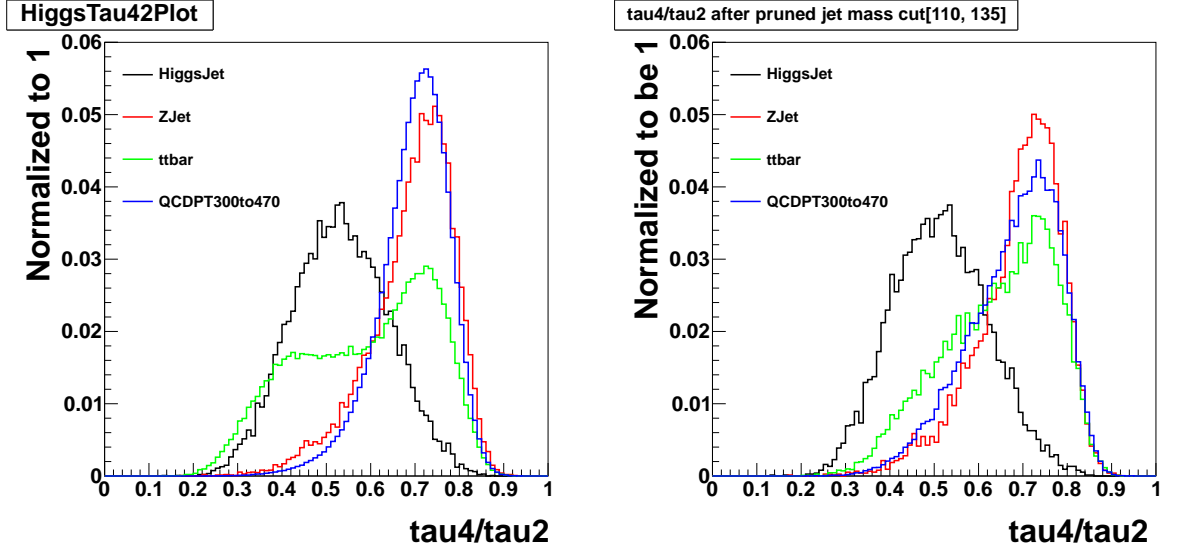


Figure 4.9: Distribution for τ_4/τ_2 in data and in simulations of signal (1.0 TeV) and background events. All simulated distributions are scaled to match the number of events in data, W/Z, matched top and Higgs jets are required to match their generator level particles, respectively.

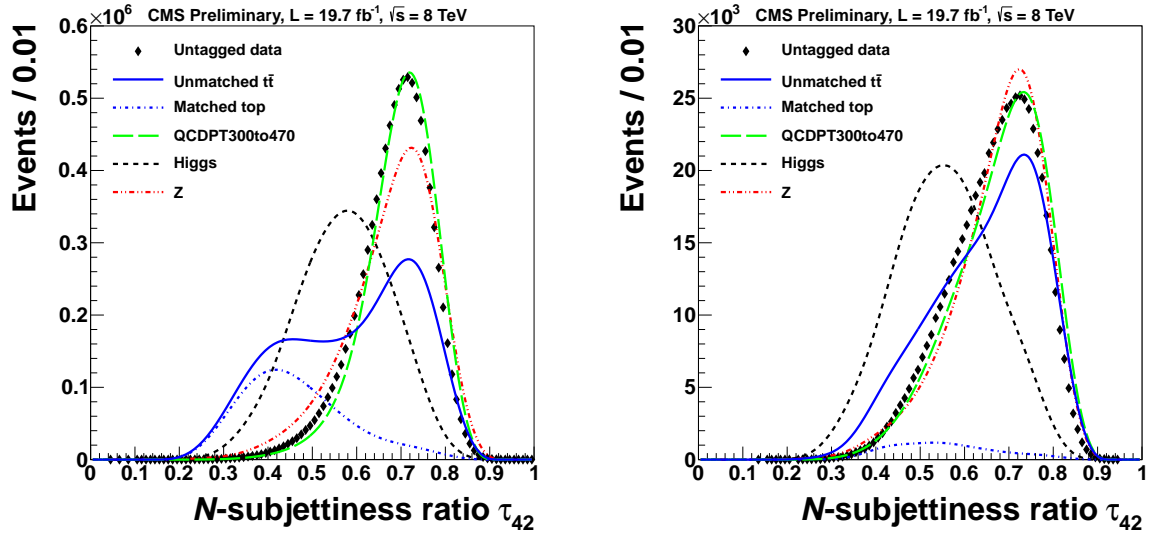


Figure 4.10: Distribution for τ_4/τ_2 in data and in simulations of signal (2.0 TeV) and background events. All simulated distributions are scaled to match the number of events in data, except that matched top is scaled to its fraction of unmatched $t\bar{t}$ times the number of data events. W/Z, matched top and Higgs jets are required to match their generator level particles, respectively.

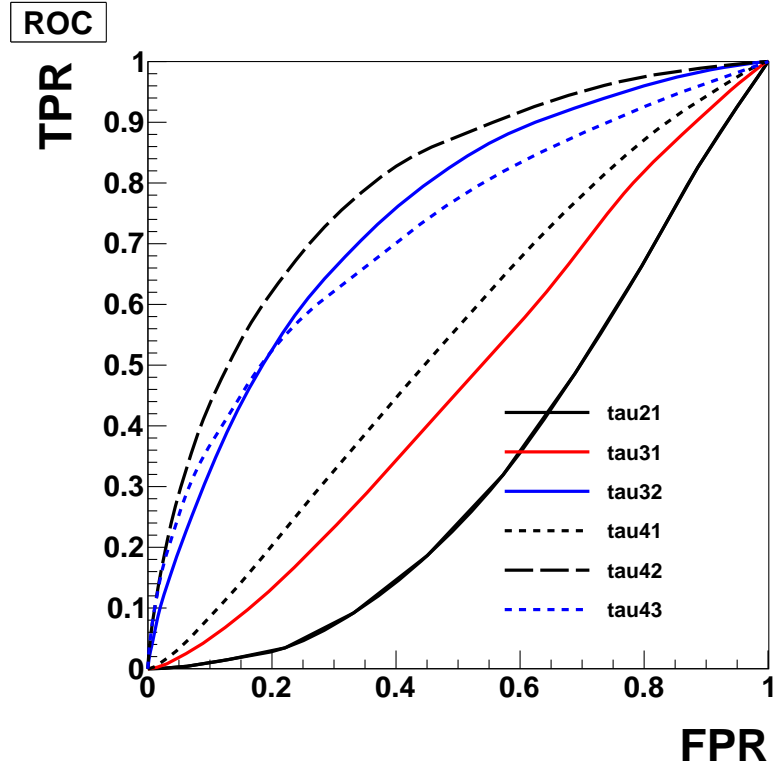


Figure 4.11: ROC curves for different τ_{NM} after the cut on the pruned jet mass. The false positive rate (FPR) is obtained from QCDPT300to470 and the true positive rate (TPR) from Higgs jets in 2 TeVsignal MC sample. Using τ_{42} to select Higgs jets outperforms all other τ_{NM} variables.

- **N-subjettiness** - We split the events into two categories, “high purity” Higgs jets by requiring $\tau_{42} \leq 0.55$, while $0.55 < \tau_{42} < 0.65$ defines the “low purity” Higgs jets.

4.5.5.1 Optimization of the τ_4/τ_2 threshold

Having selected τ_{42} as the discriminating variable, we next optimize its upper value. In this study, the jet mass is confined within $[110, 135]$ GeV. We use the limit setting method (described in Sec. ??) and evaluate the expected limits of several signal resonance masses at different τ_{42} working points. These expected limits are presented in Table. ?. Given our focus on the resonance masses above 1500 GeV, we choose to cut on $\tau_{42} < 0.55$. In the following analysis, to compensate the signal efficiency loss at higher resonance mass, we introduce an additional categories for $H \rightarrow WW^* \rightarrow 4q$ tagger as $0.55 < \tau_{42} < 0.65$. This is chosen from back-of-envelope calculation based on Figs ?? and ??, since this category provides very limited sensitivity.

Table 4.3: Upper limits (in units of 0.01 pb) for high purity HW and HZ signals at different resonance masses and also different τ_{42} working points.

HW / τ_{42}	0.45	0.5	0.55	0.6
1000	4.14	4.09	4.46	4.91
1500	0.97	0.88	0.86	0.91
2000	0.89	0.64	0.51	0.47
2500	1.36	0.82	0.53	0.40
HZ / τ_{42}				
1000	4.31	4.36	4.63	5.05
1500	0.98	0.89	0.86	0.90
2000	0.70	0.55	0.42	0.39
2500	0.96	0.61	0.41	0.32

4.6 Signal efficiency

We search for several models of heavy resonances decaying to a W or Z boson on one side, and a Higgs on the other, where both bosons decay to quarks producing merged jet. This analysis is focused on two channels:

- $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$, and
- $H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$

As previously discussed, we use one V-tagging and two Higgs tagging algorithms to identify such events. After subdividing the events according to high purity and low purity tags, we end up with five distinct categories, as shown in Table ??.

In this section, we discuss various issues related to the evaluation of the signal efficiency.

4.6.1 Cross-talk between the Higgs decay channels

In order to combine events from all categories into a single joint likelihood, the categories must be mutually exclusive. However, a cross-talk between the Higgs channels is nevertheless possible: for example, $H \rightarrow b\bar{b}$ tagger can identify other two-prong Higgs decay modes like $H \rightarrow gg$, $H \rightarrow \tau\tau$, or $H \rightarrow c\bar{c}$, although this kind of ‘false positive’ tag happens only rarely (the efficiency is $\lesssim 6\%$). Similarly, events from two-prong Higgs decay channels can also pass the τ_{42} cut in the $H \rightarrow WW^* \rightarrow 4q$ selection. In this case, the channel $H \rightarrow b\bar{b}$, because of its large branching ratio,

contributes a non-negligible number of events to the sample of 4q tags. This effect is illustrated by Fig. ??, where it can be seen that most of the low- τ_{42} tail of the $H \rightarrow b\bar{b}$ curve will be below the cut value of 0.55.

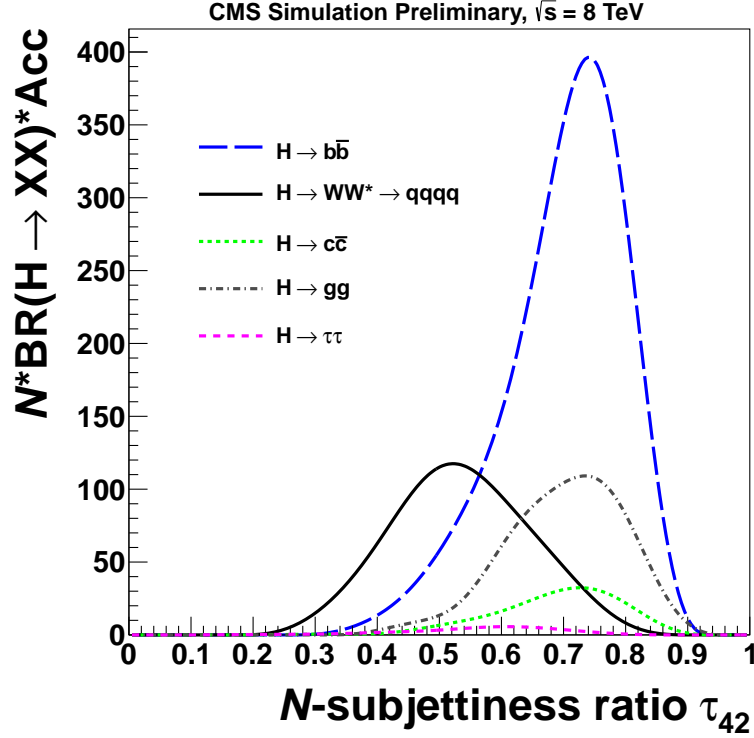


Figure 4.12: Comparison of τ_{42} distribution for $H(WW \rightarrow qq qq)$, $H(gg)$, $H(b\bar{b})$, $H(c\bar{c})$, $H(\tau\tau)$ channels. All curves are drawn for Higgs jets after pruned jet mass cut, and also normalized with respective branching ratios.

Table ?? provides an overview of the cross-talk between the various channels. The Higgs branching ratios correspond to the Higgs mass of 125 GeV. For $H \rightarrow WW^* \rightarrow 4q$, the branching ratio of the hadronic decay of (real) W boson is already included, so that the final state is four quarks. The table is normalized to 100,000 standard model Higgs bosons, and the numbers in the table show the number of Higgs decays that pass the

tagger for each channel, with the branching ratio taken into account. For example, let us consider $H \rightarrow c\bar{c}$ channel. At the Z' resonance mass of 1 TeV, out of 100,000 Higgs decays, 104 events are tagged by the $H \rightarrow b\bar{b}$ tagger and 82 pass $H \rightarrow WW^* \rightarrow 4q$ tagger but fail $H \rightarrow b\bar{b}$ tagger. For $H \rightarrow ZZ$ decays, we take its tagging efficiency the same as $H \rightarrow WW^* \rightarrow 4q$ signals. So the number of $H \rightarrow ZZ$ to pass $H \rightarrow b\bar{b}$ and $H \rightarrow WW^* \rightarrow 4q$ tagger is estimated by efficiency of $H \rightarrow WW^* \rightarrow 4q$ signal times $BR(H_{ZZ}) * BR(Z_{qq}) * BR(Z_{qq})$ divided by $BR(H_{WW}) * BR(W_{qq}) * BR(W_{qq})$.

From Table ??, it can be seen that for various signal resonance masses, the contribution of other decay channels to the sample of $H \rightarrow b\bar{b}$ tags never exceeds 6%. We will assign additional systematics for the cross-talk.

Since the $H \rightarrow b\bar{b}$ tagger has significantly lower background than $H \rightarrow WW^* \rightarrow 4q$, it takes precedence in selecting events: we first identify the events that pass the $H \rightarrow b\bar{b}$ tagger, and only if they fail, we test them for the presence of the $H \rightarrow WW^* \rightarrow 4q$ tag.

The effect of the $H \rightarrow b\bar{b}$ tagger veto on the $H \rightarrow WW^* \rightarrow 4q$ tagged dijet mass distribution background (data) is shown in Appendix ??.

CHAPTER 4. EXO-14-009

Table 4.4: Number of Higgs jets falls into two exclusive categories, assuming we have 100,000 SM Higgs (125 GeV) decays to all channels. $H \rightarrow ZZ^* \rightarrow 4q$ signals are estimated by its branching ratio times the efficiency of $H \rightarrow WW^* \rightarrow 4q$ signals divided by the branching ratio of $H \rightarrow WW^* \rightarrow 4q$ channel.

	Branching ratio (%)	Pass $H \rightarrow b\bar{b}$	Fail $H \rightarrow b\bar{b}$, pass $H \rightarrow WW^* \rightarrow 4q$
1.0 TeV			
$H \rightarrow b\bar{b}$	57.70	11871	804
$H \rightarrow WW^* \rightarrow 4q$	9.94	86	2360
$H \rightarrow ZZ^* \rightarrow 4q$	1.30	11	309
$H \rightarrow c\bar{c}$	3.00	104	82
$H \rightarrow \tau\tau$	6.30	17	37
$H \rightarrow gg$	10.00	14	314
1.5 TeV			
$H \rightarrow b\bar{b}$	57.70	11444	755
$H \rightarrow WW^* \rightarrow 4q$	9.94	228	1916
$H \rightarrow ZZ^* \rightarrow 4q$	1.30	29	250
$H \rightarrow c\bar{c}$	3.00	121	88
$H \rightarrow \tau\tau$	6.30	12	57
$H \rightarrow gg$	10.00	69	174
2.0 TeV			
$H \rightarrow b\bar{b}$	57.70	13816	551
$H \rightarrow WW^* \rightarrow 4q$	9.94	449	1435
$H \rightarrow ZZ^* \rightarrow 4q$	1.30	58	187

4.6.2 Summary of Higgs and W/Z tagging categories

The W or Z jets from the signal are selected by the V-tagger, and the Higgs candidates are selected by an OR of the two Higgs taggers, $H \rightarrow b\bar{b}$ and $H \rightarrow WW^* \rightarrow 4q$. Both V-tagger and $H \rightarrow WW^* \rightarrow 4q$ taggers have high-purity and low-purity categories. The latter are added to increase the sensitivity of the analysis at high resonance masses, where the QCD background is low, and a higher signal efficiency is at the premium.

We first identify the events that pass the $H \rightarrow b\bar{b}$ tagger, and only if they fail, we test them for the presence of the $H \rightarrow WW^* \rightarrow 4q$ tag. Thus we arrive at the final division of events into mutually exclusive categories listed in Table ?? . For the $H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$ channel, we drop the low-purity Higgs and low-purity V-tagging category, because it adds only a negligible sensitivity.

Table 4.5: The five exclusive event categories used in this analysis. We also assign specific names (in parenthesis) for each category, which will be used in following sections.

$H \rightarrow b\bar{b}, W/Z \rightarrow qq'$	$H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$
high-purity V-tag (Hbb1)	high-purity H-tag, high-purity V-tag (Hww1)
low-purity V-tag (Hbb2)	high-purity H-tag, low purity V-tag (Hww2)
	high-purity V-tag, low purity H-tag (Hww3)

The events from the $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$ decay could contribute to all the five

categories, due to its large branching ratio. The $H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$ signal events contribute only in events that fail $H \rightarrow b\bar{b}$ but pass $H \rightarrow WW^* \rightarrow 4q$ tagger; their contribution to $H \rightarrow b\bar{b}$ tagged sample is negligible. The contributions from other Higgs decay modes to all these five categories is tiny compared to $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$ and $H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$ yields. We will not specifically study them, but include them as systematic uncertainties.

4.6.3 Tagging efficiency for $H \rightarrow b\bar{b}$ jets

We study the Higgs tagging efficiency in MC by matching the jet to the Higgs generator-level particle. This jet is referred to as the Higgs jet. The Higgs tagging efficiency is obtained from the MC simulation as the fraction of the Higgs jets that passes the given H-tagging selection. It is given in Fig. ???. The same Figure shows the W/Z tagging efficiency for the other jet in the event. The total event efficiency is a product of these two efficiencies.

In the $H \rightarrow b\bar{b}$ channel, the $H \rightarrow b\bar{b}$ tagging efficiency start rising after ~ 1.6 TeV. The reason is explained as follows. For the resonance masses above 1.6 TeV, the Higgs jets are sufficiently boosted that the ΔR between the two b subjets is ≤ 0.3 . When $\Delta R \leq 0.3$, we are switching from 2 subjets b tagging to CSV loose fat jet b tagging. This causes the rising tagging efficiency.

Since the CSV tagging uses the cone of 0.3 to associate the candidate tracks to the jet, when the two subjets are at angular distance of 0.6, they begin sharing tracks.

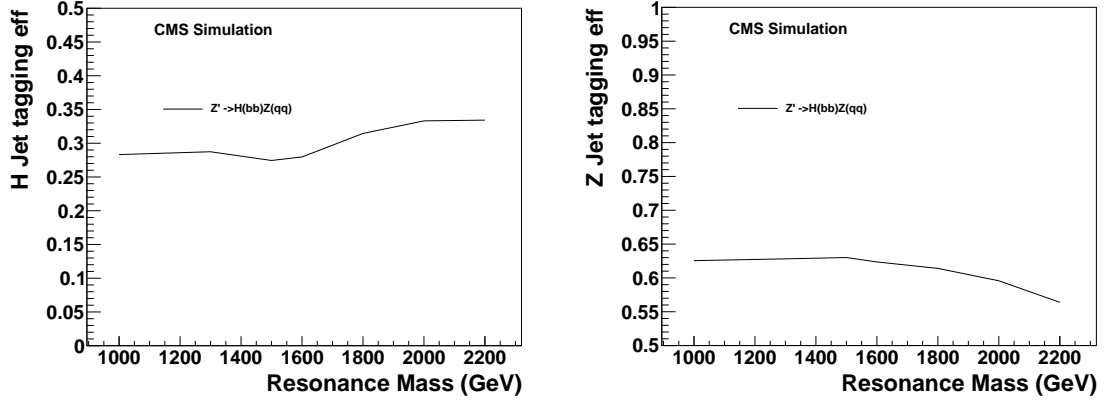


Figure 4.13: Higgs jets and Z jets tagging efficiencies in signal MC simulation. Left: $H \rightarrow b\bar{b}$. Right: $Z \rightarrow q\bar{q}$. The total event efficiency is a product of the two, resulting in a relatively flat efficiency for reconstructing $X \rightarrow HV$.

This effect becomes important for $\Delta R \leq 0.3 \sim 0.4$. For this reason, when the subjects are closer than 0.3, following the BTV POG recommendation, we switch to using the CSV b tagging decision for the fat jet. (We use CSVL, the loose operating point.)

Note that if for the fat jet b tagging CSVM operating point is used, the $H \rightarrow b\bar{b}$ tagging efficiency is smooth, as shown in Fig. ???. For the $H \rightarrow b\bar{b}, Z \rightarrow q\bar{q}$ analysis, we have compared the limits of these two different fat jet b tagging methods (CSVL *vs.* CSVM). Unsurprisingly, we have found that using the more-efficient CSVL b tagging results in better expected limits in this background-poor region than using the CSVM b tagging operating point. (The details of this study are given in the Appendix.)

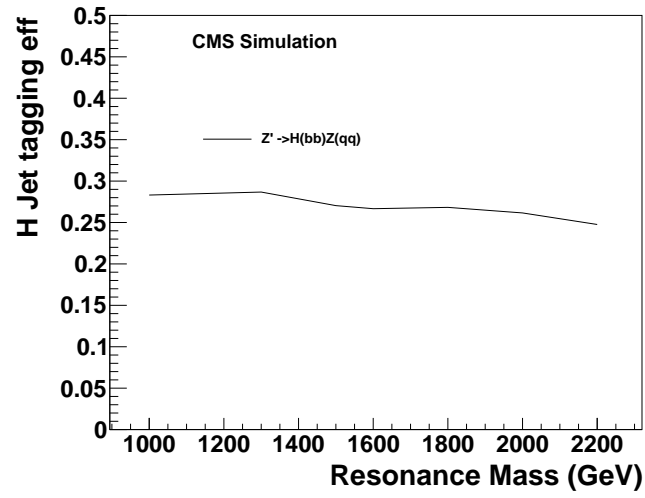


Figure 4.14: Higgs tagging efficiency in signal MC, for $H \rightarrow b\bar{b}$ channel. Changing fat jet b tagging to CSVM instead of CSVL.

4.6.4 Signal acceptance and total efficiency for $H \rightarrow$

$b\bar{b}, Z \rightarrow q\bar{q}$ channel

Signal acceptance is defined as the number of signal events pass all the kinematic event selection (that is, without the two jet-tagging algorithms) divided by the number of generated events. The signal acceptance for $H \rightarrow b\bar{b}, Z \rightarrow q\bar{q}$ channel is shown on Fig. ??.

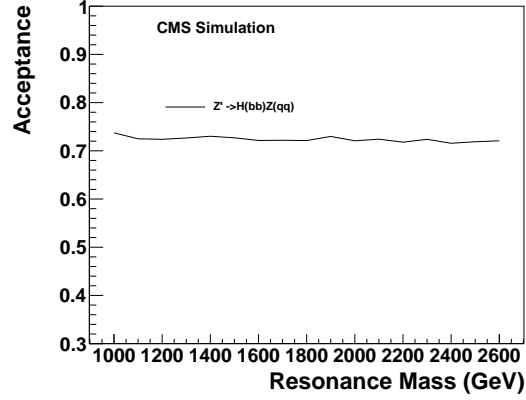


Figure 4.15: Acceptance in $H \rightarrow b\bar{b}, Z \rightarrow q\bar{q}$ signal.

The combined tagging rate of H and Z tagging, is defined as the number of events pass the HZ-tagging divided by the number of events after events selection, which is shown in Fig. ?? for $H \rightarrow b\bar{b}, Z \rightarrow q\bar{q}$ tagging.

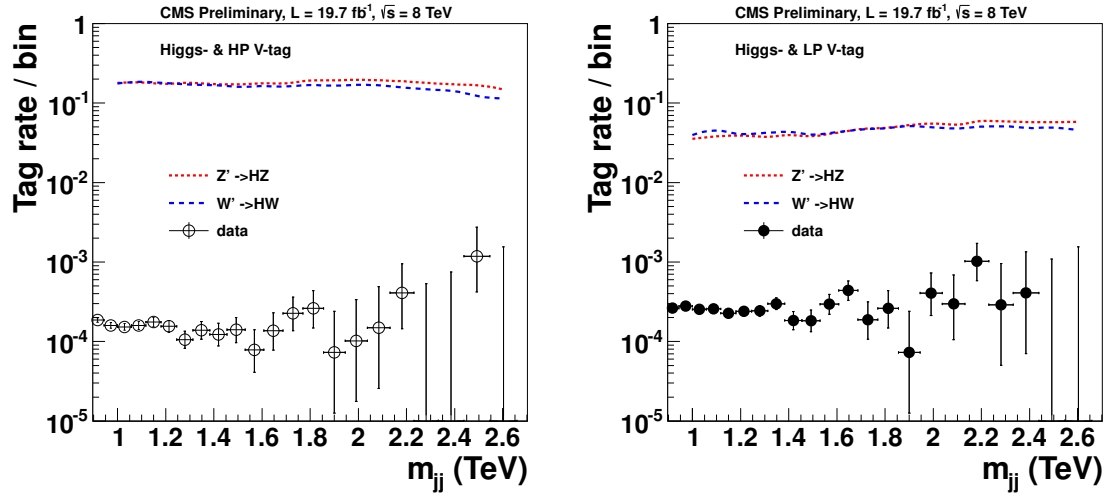


Figure 4.16: Tagging rates in $H \rightarrow b\bar{b}, Z \rightarrow q\bar{q}$ and $H \rightarrow b\bar{b}, W \rightarrow qq'$ signal channels and data. Horizontal bars in data indicates variable binning size.

4.6.5 Tagging efficiency for $H \rightarrow WW^* \rightarrow 4q$ jets

The tagging efficiency for $H \rightarrow WW^* \rightarrow 4q$ jets, as a fraction of getJets that passes the selection, is shown in Fig. ??.

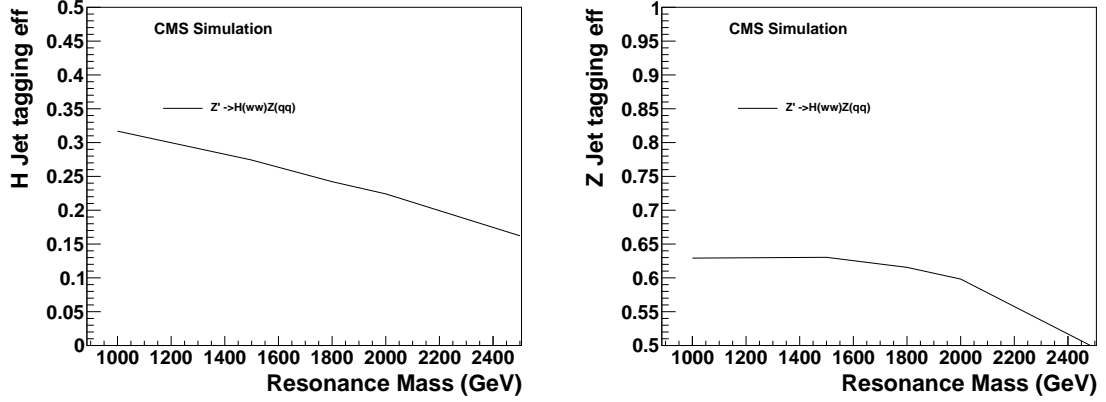


Figure 4.17: Higgs jets and Z jets tagging efficiencies in signal MC simulation. Left: $H \rightarrow WW^* \rightarrow 4q$. Right: $Z \rightarrow q\bar{q}$. The total event efficiency is a product of the two, resulting in a following spectrum efficiency for reconstructing $X \rightarrow HV$.

In the $H \rightarrow WW^* \rightarrow 4q$ all-hadronic channel, to compensate the efficiency loss in the high resonance mass, we also add two low purity categories, low purity H-tagging and low purity V-tagging. The tagging efficiency of low purity H/V-tagging on the H/Z jets is shown on Fig. ?. And low purity Higgs is defined as $0.55 < \tau_{42} < 0.65$, pruned jet mass in $[110, 135]$ GeV.

For low purity W/Z tagging, τ_{21} must be in $[0.5, 0.75]$, and the pruned jet mass in the window $[70, 100]$ GeV.

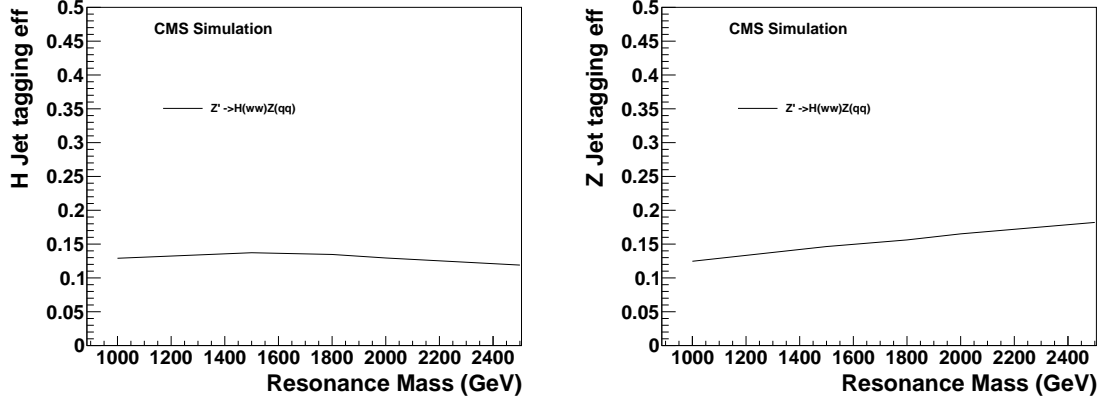


Figure 4.18: Two purity Higgs jets and Z jets tagging efficiencies in signal MC simulation. Left: $H \rightarrow WW^* \rightarrow 4q$. Right: $Z \rightarrow q\bar{q}$.

4.6.6 Signal acceptance and total efficiency for $H \rightarrow$

$WW^* \rightarrow 4q, Z \rightarrow q\bar{q}$ channel

The acceptance of the $H \rightarrow WW^* \rightarrow 4q, Z \rightarrow q\bar{q}$ channel is shown on Fig. ??

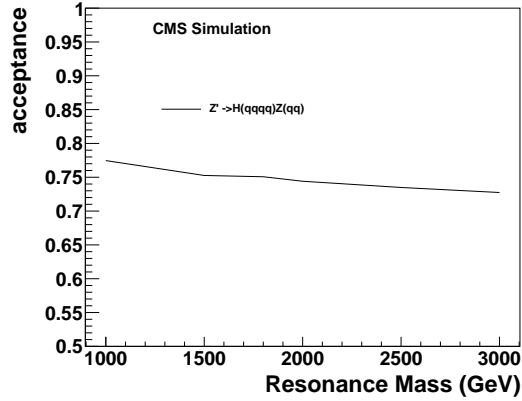


Figure 4.19: Acceptance in signal.

The combined tagging rates of H and Z tagging, for $H \rightarrow WW^* \rightarrow 4q, Z \rightarrow q\bar{q}$

channel in signal and data are shown on Fig. ?? .

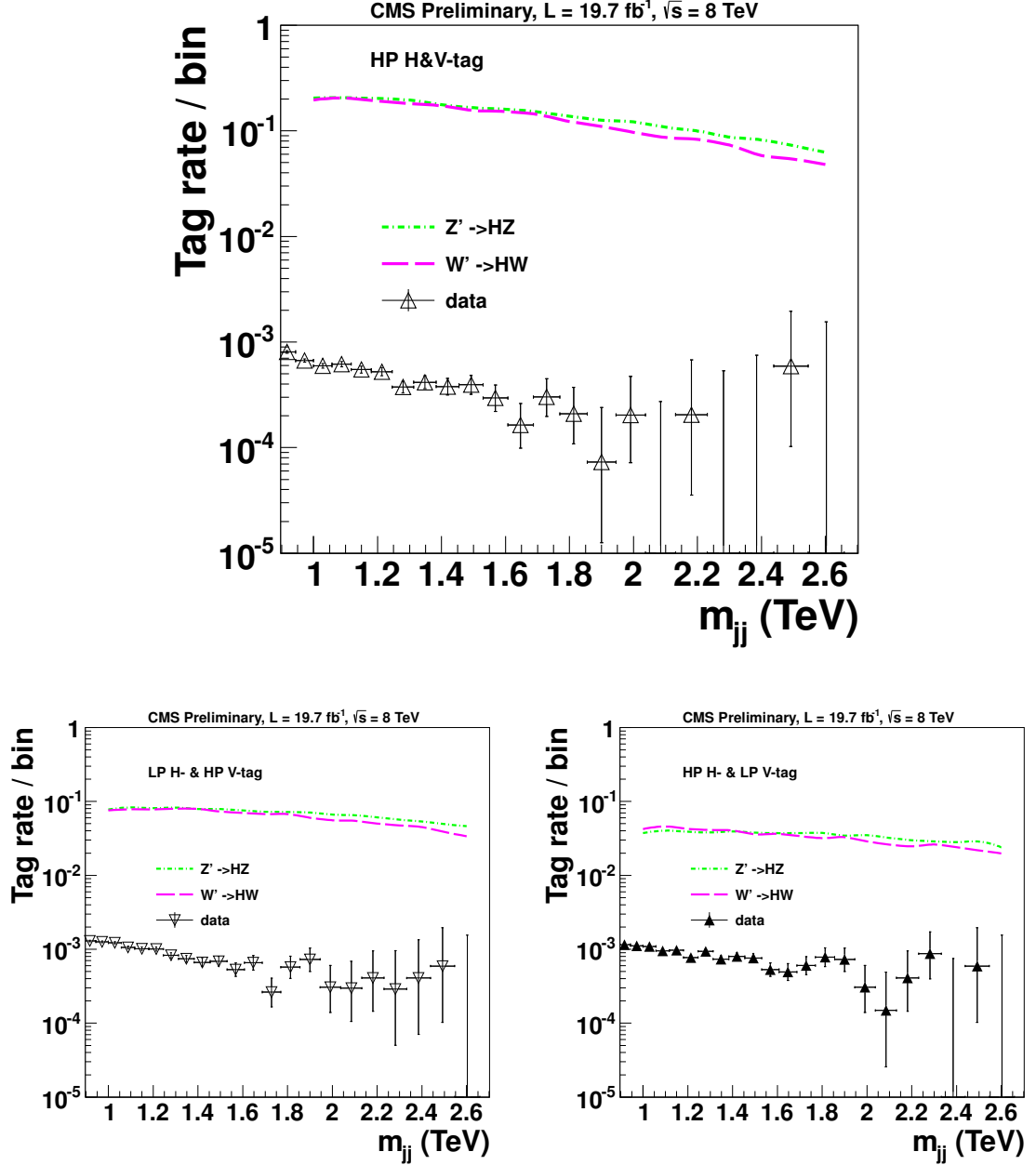


Figure 4.20: Tagging rates in $H \rightarrow WW^* \rightarrow 4q, Z \rightarrow q\bar{q}$ and $H \rightarrow WW^* \rightarrow 4q, W \rightarrow qq'$ signal channels and data. Horizontal bars in data indicates variable bins.

The signal shapes are shown in Fig. ??

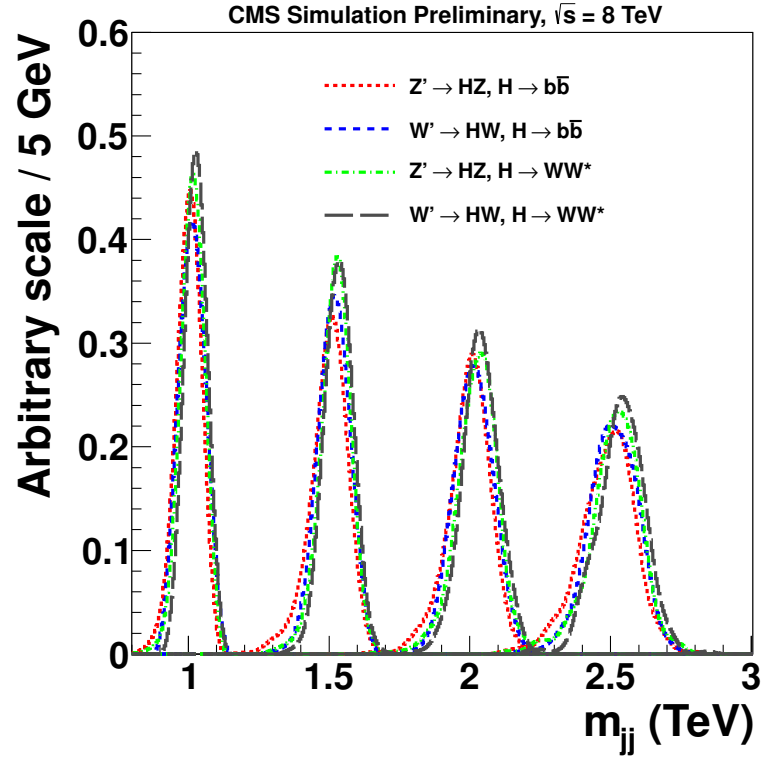


Figure 4.21: Signal shapes for Z' and W' signals at 1.0, 1.5, 2.0 and 2.5 TeV resonances.

4.7 Background shape parametrization

Background from multijet events is modelled by a smoothly falling distribution for each event category, given by the empirical probability density function

$$P_D(m_{jj}) = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}} . \quad (4.2)$$

For each category, the normalization factor P_0 and the two parameters P_1 and P_2 are treated as uncorrelated. This parameterization was deployed successfully in searches in dijet mass spectra [?]. A Fisher F-test [?] is used to check that no additional parameters are needed to model the individual background distribution, for each of the four cases considered.

The background-only fit plot is only to demonstrate that the fit function works nice if no signal is assumed. However we don't use that function. We search for a peak on top of the falling background spectrum by means of a maximum likelihood fit to the data.

Figure ?? and Figure ?? show the dijet mass spectra from H(bb)V-tagged and H(ww)V-tagged data fitted to Eq. (??), respectively. The bottom panes show corresponding pull distributions, demonstrating the agreement between the background-only probability density function and the data.

The largest resonance mass in each channel is described in Table ??

No sizeable deviation from the background-only hypothesis is seen, exclusion limits

Table 4.6: Highest resonance mass in each channel.

Category	Resoanance(GeV)
Hbb1	2481
Hbb2	2354
Hww1	2449
Hww2	2448
Hww3	2471

are set on the product of cross section, acceptance, and branching fraction for V' to HV search.

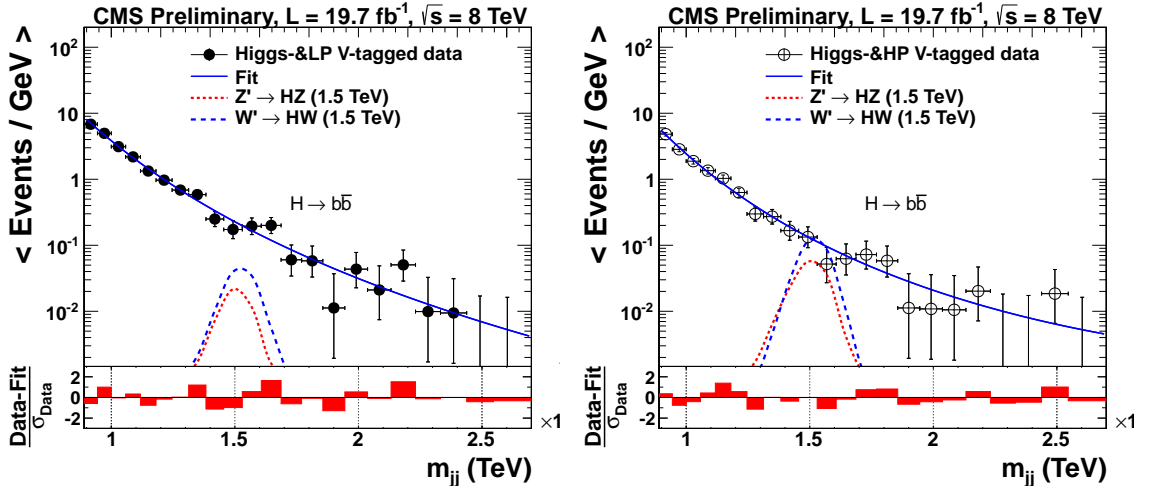


Figure 4.22: Distributions in m_{jj} , respectively, for LP V-tag(left), HP V-tag(right). The solid curves represent the results of fitting Eq. (??) to the data. The distributions for $H \rightarrow b\bar{b}$, $Z \rightarrow q\bar{q}$ and $H \rightarrow b\bar{b}$, $W \rightarrow qq'$ contributions, scaled to their corresponding cross sections, are given by the dash-dotted curves. Horizontal bars in data indicates variable binning size. The corresponding pull distributions ($\frac{\text{Data-Fit}}{\sigma_{\text{Data}}}$, where σ_{Data} represents the statistical uncertainty in the data in a bin in m_{jj}) are shown below each m_{jj} plot.

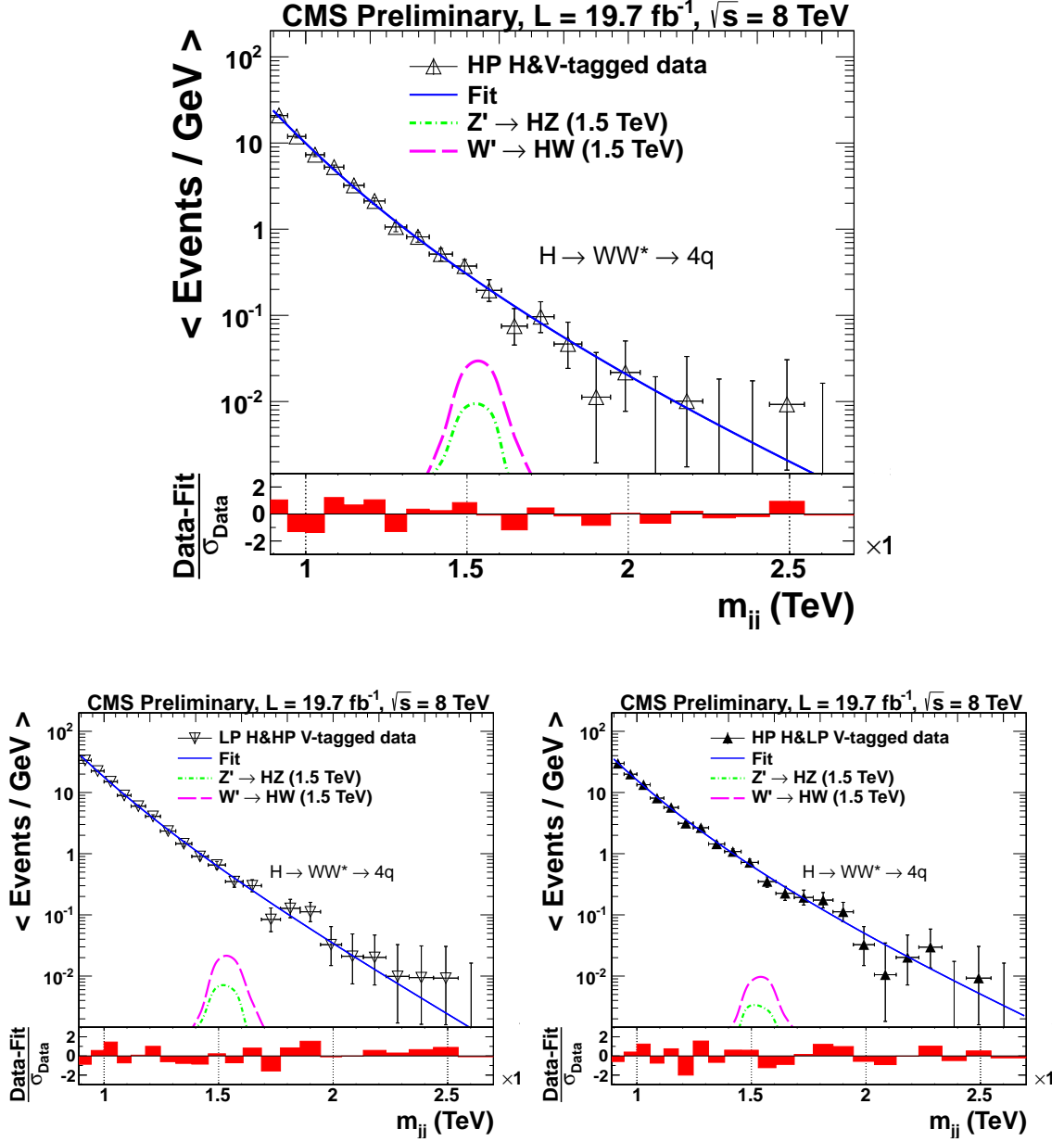


Figure 4.23: Distributions in m_{jj} , respectively, for HP(top), LP H-tag(bottom left) and LP V-tag(bottom right). The solid curves represent the results of fitting Eq. (??) to the data. The distributions for $H \rightarrow WW^* \rightarrow 4q$, $W \rightarrow qq'$ and $H \rightarrow WW^* \rightarrow 4q$, $Z \rightarrow q\bar{q}$ contributions, scaled to their corresponding cross sections, are given by the dash-dotted curves. Horizontal bars in data indicates variable binning size. The corresponding pull distributions ($\frac{\text{Data-Fit}}{\sigma_{\text{Data}}}$, where σ_{Data} represents the statistical uncertainty in the data in a bin in m_{jj}) are shown below each m_{jj} plot.

4.8 Systematic uncertainties

The sources of systematic uncertainties are summarized here: γ -H-tagging efficiency, b-tagging scale factor, PDF uncertainties, W/Z-tagging efficiency, Jet Energy Scale(JES), Jet Energy Resolution(JER), luminosity, cross-talk of various signals.

4.8.1 b-tagging scale factor

We use method 1c recommended by the BTV group to apply b-tagging scale factor and uncertainties, which is https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMMethods#1c_Event_reweighting_using_scale.

if ΔR of the two subjects is bigger than 0.3, we apply the b-tagging scale factor on the two subjects. While ΔR is between 0.3 to 0.4, we double the b-tagging scale factor uncertainties. When ΔR is smaller than 0.3, we apply the b-tagging scale factor on the fat jet. The uncertainty from b-tagging scale factor is evaluated by taking the largest difference in tagging efficiency by shifting b-tagging scale factor 1σ , which results in 15% uncertainty.

4.8.2 W/Z-tagging efficiency

W/Z-tagging efficiency scale factor is studied in hadronic VV search [?, ?], details following. The W/Z-tagging efficiency is determined from the Monte Carlo simulation. We cross-check the MC modelling of the signal efficiency by measuring the

W/Z-tagging efficiency in semileptonic $t\bar{t}$ data, and compare it with the same efficiency obtained using identical procedure from $t\bar{t}$ Monte Carlo sample generated with MadGraph [?] and showered with Pythia6 Tune Z2*. The ratio of the two efficiencies defines a scale factor, which is then applied to the efficiencies for signals in the dijet data.

Combining the efficiencies of the τ_{21} and jet mass cuts, a data-MC scale factor of 0.86 ± 0.07 (1.39 ± 0.75) for the high (low) purity selection for the W-tagging efficiency is determined. We assume that the same scale factor applies to Z-tagging as well. The errors on the scale factor are propagated into the systematic uncertainties on the overall signal efficiency.

The efficiency error on a single W/Z-tagging is estimated with a control sample of semileptonic $t\bar{t}$ events as described above. The uncertainties of 7.5% (54%) on the scale factors for high (low) purity tagging include sources from control sample statistics, pruned jet mass scale and pruned jet mass resolution. Since we estimate the scale factor only in the kinematic regime of the $t\bar{t}$ sample where the W decay products merge, but the b-quarks are still reconstructed as separate jets, we need to rely on the simulation to extrapolate to higher jet p_T .

Therefore, we estimate how the W/Z jets tagging efficiency varies as a function of p_T for two different showering and hadronization models using PYTHIA 6 and HERWIG++. We find that the differences are within 4% (12%) for the high (low) purity tagging [?], significantly smaller than the statistical uncertainties in the scale

factors.

4.8.3 H-tagging efficiency for $H \rightarrow WW^* \rightarrow 4q$ tagger

We extrapolate the H-tagging efficiency scale factor from the W/Z-tagging efficiency scale factor, more details are in Appendix ???. In $H \rightarrow WW^* \rightarrow 4q$, Higgs will decay to one on-shell W and one off-shell W. The on-shell W decays similarly as a single W/Z jet. while the off-shell W, is soft. So the overall Higgs jet could be viewed as a real W/Z jet plus soft part. To the first order approximation, we apply the same scale factor to H-tagging as W/Z-tagging, but with an additional uncertainty.

For W/Z tagging, which uses τ_{21} , we know exactly the scale factor at low p_T and need no additional uncertainty from Pythia/Herwig. So the Herwig efficiency as a function of p_T can be normalized in such a way that Pythia and Herwig agree at low p_T , but are different at high p_T .

For τ_{42} we don't know if it is correct at low p_T and high p_T . Therefore, we should be more conservative and take the Pythia-Herwig difference without normalizing them at low p_T .

We estimate how the Higgs jet tagging efficiency varies as a function of p_T for two different showering and hadronization models using PYTHIA 6 and HERWIG++. We find that the differences are within 7% (7%) for the high (low) purity tagging,

the results are summarized in Table ??.

Table 4.7: The difference of $H \rightarrow WW^* \rightarrow 4q$ jet tagging efficiency in signal MC by showering and hadronization with Pythia and Herwig. Higgs jets are the jets matched to Higgs generator particles.

Resonance	1000GeV	1500GeV	2000GeV	2500GeV
High Purity Higgs jets	4.75%	2.28%	1.52%	6.31%
Low Purity Higgs jets	5.91%	6.33%	0.79%	4.39%

4.8.4 Other uncertainties

Because of the rejection of charged particles not originating from the primary vertex and also the application of pruning, the pileup dependence on the $H/W/Z$ -tagging efficiency is weak, and the uncertainty of the modeling of the pileup distribution is less than 3%. Modeling of the underlying event, estimated by switching it off in PYTHIA 6, and also by comparing different tunes of PYTHIA 6, impacts the tagging efficiency by less than 1%.

In the jet p_T and η regions considered in this analysis, the Jet Energy Scale is known to a precision of 1-2% [?, ?]. For JES, p_T and η dependent uncertainty is propagated to the reconstructed dijet invariant mass, and taken into account by shifting the resonance dijet mass in the statistical analysis.

The Jet Energy Resolution(JER) is known to a precision of 10% and its tails are in agreement between data and MC [?]. The JER is taken into account in the statistical

analysis by a variation of the resonance width by 10%.

The JES and JER of b jets are studied by JETMet, showing a smaller uncertainty than the standard QCD mixture of light quarks and gluon jets, whose JES and JER are applied for all flavor jets in this analysis. So we don't apply additional specific JES and JER uncertainty for b jets. The uncertainty on the pruned jet mass cut of $H \rightarrow WW^* \rightarrow 4q$ is included in the scale factor extrapolated from W-tagging. For $H \rightarrow b\bar{b}$ tagger, pruned jet mass uncertainty is evaluated as 2.6%, synchronized with EXO-12-053.

The uncertainty related to the PDF used to model the signal acceptance is estimated from the eigenvectors of the CT10, MRST2008 and NNPDF sets of PDF. The envelope of the upward and downward variations of the estimated acceptance of the three sets is assigned as uncertainty and found to be 5%-15% in the resonance mass range of interest.

The luminosity has been measured with an uncertainty of 2.6% [?], and is also taken into account in the statistical analysis.

For the cross-talk of Higgs signals, we assign them as systematic uncertainty. In this analysis, signals are generated in exclusive decay channels, for example, $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$ signals only have $H \rightarrow b\bar{b}$ decays, no other Higgs decays. So in category Hbb1 and Hbb2, the only signal is $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$, other Higgs decays passing the $H \rightarrow b\bar{b}$ tagger will be assigned as systematic uncertainties, which is

evaluated according to Equation ??.

$$Uncertainty_{Cross-talk} = \frac{NumberofExpectedNuisanceSignals + NumberofSignalOfInterest}{NumberofSignalOfInterest} - 1 \quad (4.3)$$

For $H \rightarrow b\bar{b}$ tagger, signal of interest is $H \rightarrow b\bar{b}$ signals. All other Higgs decays are taken as nuisance signals. For $H \rightarrow WW^* \rightarrow 4q$ tagger, signal of interest would be $H \rightarrow b\bar{b}$ plus $H \rightarrow WW^* \rightarrow 4q$ and all other Higgs decays are taken as nuisance signals. This number is evaluated across various resonance signal masses, resulting in 7% for category Hbb1 and Hbb2. This uncertainty is 31% for Hww1, Hww2 and Hww3, including 9.4% uncertainty from $H \rightarrow ZZ^* \rightarrow 4q$ decays, and 21% from Hcc,Hgg, etc. $H \rightarrow ZZ$ is taken as having the same efficiency as $H \rightarrow WW^* \rightarrow 4q$, as discussed in Table ??.

Table ?? shows a summarization of all the systematics applied.

Table 4.8: Summarization of systematics. Numbers in parenthesis are for low purity categories.

Systematics/Signals	Relevant quantity	$H \rightarrow b\bar{b}, W/Z \rightarrow qq'$ signals		$H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$ signals	CHAPTER 4. EXO-14-009
		Hbb1, Hbb2	Hww1, Hww2, Hww3		
Background fit	Resonance shape	shape	shape	shape	
Integrated Luminosity	Yield (per event)	2.6%	2.6%	2.6%	
cross-talk	Yield (per event)	7.0%	31%	31%	
Dijet mass shift due to JES	Resonance shape	1.0%	1.0%	1.0%	
Dijet mass shift due to JER	Resonance shape	10.0%	10.0%	10.0%	
PileUp	Efficiency (per jet)	1.5%	1.5%	1.5%	
PDF	Yield (per event)	5-15.0%	5-15.0%	5-15.0%	
B-tagging SF	Yield (per event)	15.0%	1.0%	15.0%	
Higgs mass	Efficiency (per jet)	2.6%	2.6%	-	
W-tagging tau21	Efficiency (per jet)	7.5%(54%)	7.5%(54%)	7.5%(54%)	
W-tagging tau21 shower/hadronization	Efficiency (per jet)	4%(12%)	4%(12%)	4%(12%)	
H-tagging tau42	Efficiency (per jet)	-	7.5%(54%)	7.5%(54%)	
H-tagging tau42 shower/hadronization	Efficiency (per jet)	-	7%(7%)	7%(7%)	

4.9 Limit setting procedure

We search for a peak on top of the falling background spectrum by means of a maximum likelihood fit to the data. The likelihood \mathcal{L} , computed using events binned as a function of m_{jj} , is written as

$$\mathcal{L} = \prod_i \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}, \quad (4.4)$$

where $\lambda_i = \mu N_i(S) + N_i(B)$, μ is a scale factor for the signal, $N_i(S)$ is the number expected from the signal, and $N_i(B)$ is the number expected from multijet background. The parameter n_i quantifies the number of events in the i^{th} m_{jj} mass bin. The background $N_i(B)$ is described by the functional form of Eq. (??). While maximizing the likelihood as a function of the resonance mass, μ as well as the parameters of the background function are left floating.

The asymptotic approximation [?] of the LHC CL_s method [?, ?] is used to set upper limits on the cross sections for resonance production. The dominant sources of systematic uncertainties are treated as nuisance parameters associated with log-normal priors in those variables. For a given value of the signal cross section, the nuisance parameters are fixed to the values that maximize the likelihood, a method referred to as profiling. The dependence of the likelihood on parameters used to describe the background in Eq. (??) is removed in the same manner, and no additional systematic uncertainty is therefore assigned to the parameterization of the

CHAPTER 4. EXO-14-009

background.

4.10 All combined limits

Figure ?? shows the limits for combining $H \rightarrow b\bar{b}$ and $H \rightarrow WW^* \rightarrow 4q$ decaying channels together in the five categories. The Higgs and V bosons branching ratios are already taken into account. In HVT B model, W' and Z' are degenerate, having about the same mass. So we also show the combined limit of W' and Z' here.

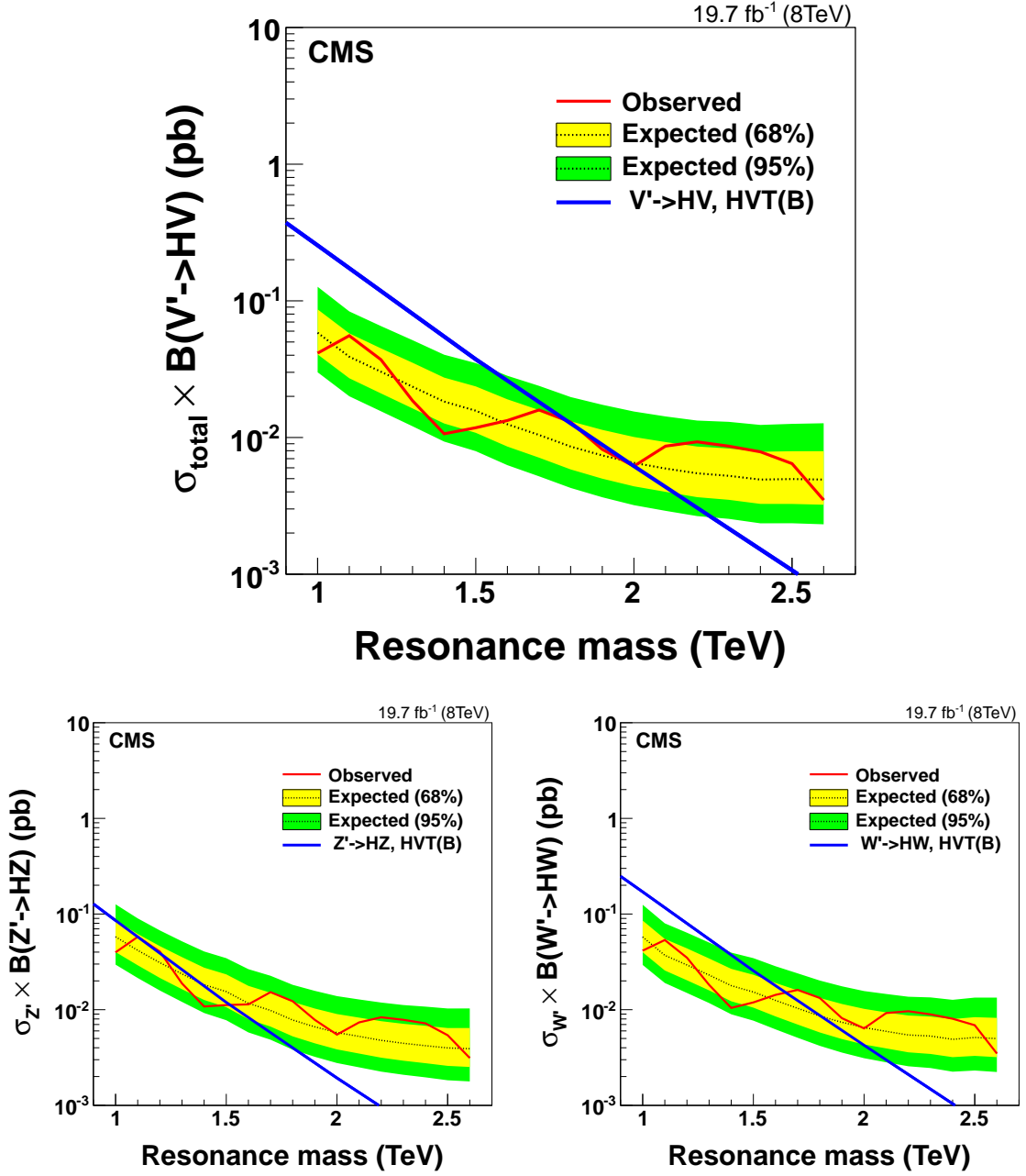


Figure 4.24: W' and Z' combined limit is the top plot, by considering them having the same mass, including all the categories in this analysis. Combined expected and observed limits for $Z' \rightarrow HZ$ (bottom left) and $W' \rightarrow WH$ (bottom right), including $H \rightarrow b\bar{b}$ and $H \rightarrow WW^* \rightarrow 4q$ channels. Branching ratios of Higgs and V decays are already taken into account. Theory model used here is HVT scenario B, arXiv:1402.4431.

4.11 $H \rightarrow b\bar{b}$ tagger limits (categories Hbb1, Hbb2)

Figure ?? shows the limits for $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$ signals passing the $H \rightarrow b\bar{b}$ tagger. Limits of combining categories Hbb1 and Hbb2 are presented. The $H \rightarrow b\bar{b}$ and V bosons branching ratios are already taken into account.

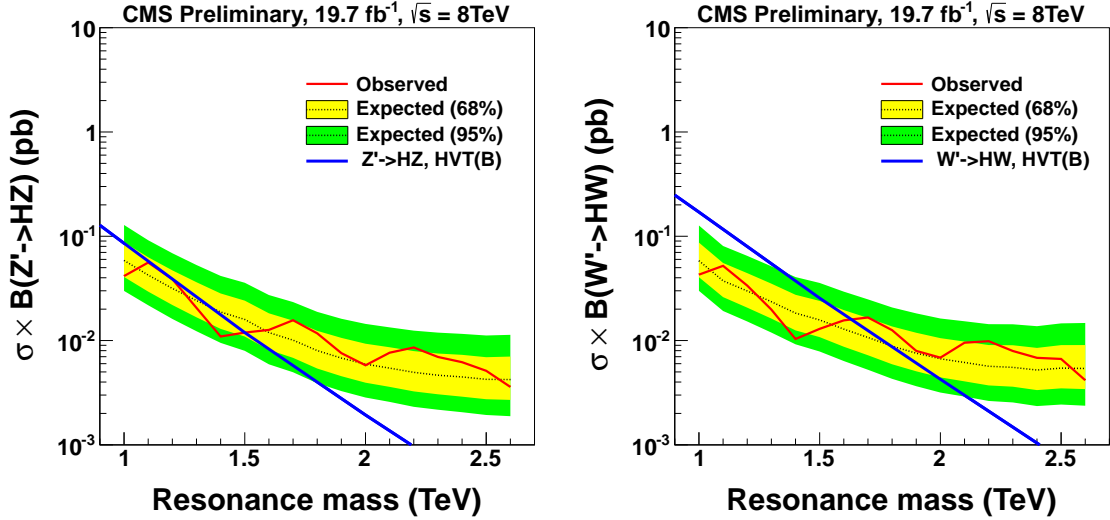


Figure 4.25: Expected and observed limits for $Z' \rightarrow HZ$ (left) and $W' \rightarrow WZ$ (right) search, in $H \rightarrow b\bar{b}$ decay mode. Branching ratios of $H \rightarrow b\bar{b}$ and V decays are taken into account. Theory model used here is HVT scenario B, arXiv:1402.4431.

4.12 $H \rightarrow WW^* \rightarrow 4q$ tagger limits (categories Hww1, Hww2, Hww3)

Figure ?? shows the combined limits for $H \rightarrow WW^* \rightarrow 4q, W/Z \rightarrow qq'$ and $H \rightarrow b\bar{b}, W/Z \rightarrow qq'$ signals failing the $H \rightarrow b\bar{b}$ tagger but passing the $H \rightarrow WW^* \rightarrow 4q$ tagger. Limits of combining category Hww1, Hww2 and Hww3 are presented. The $H \rightarrow WW^* \rightarrow 4q, H \rightarrow b\bar{b}$ and V bosons branching ratios are already taken into account.

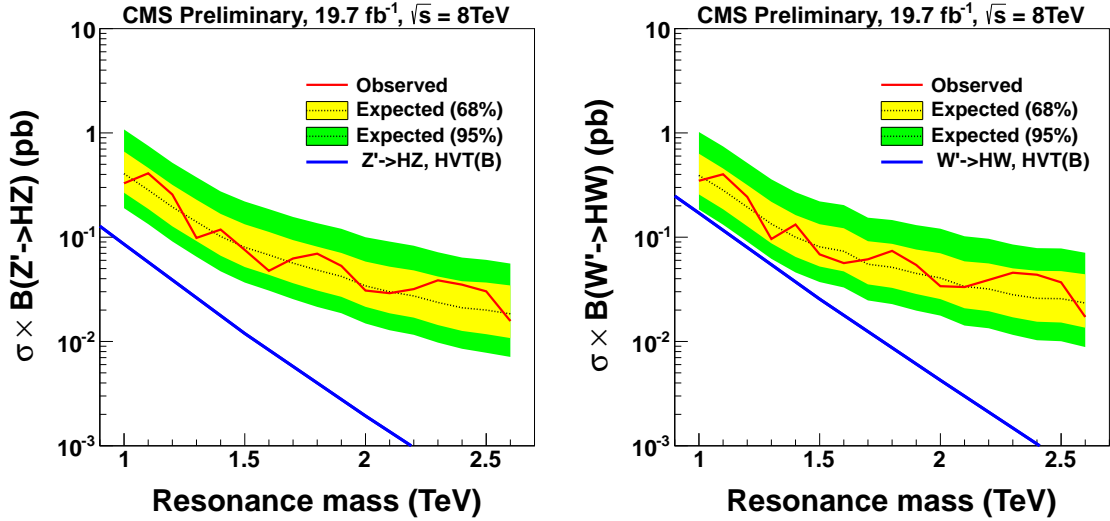


Figure 4.26: Expected and observed limits for $Z' \rightarrow HZ$ (left) and $W' \rightarrow WZ$ (right) search for categories Hww1, Hww2 and Hww3. Branching ratios of $H \rightarrow WW^* \rightarrow 4q, H \rightarrow b\bar{b}$ and V decays are taken into account. Theory model used here is HVT scenario B, arXiv:1402.4431.

4.13 Results and conclusions

Our limits results are summarized in Table ??.

A data sample corresponding to an integrated luminosity of 19.7 fb^{-1} collected in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ with the CMS detector has been used to measure the W/Z - and H-tagged dijet mass spectrum using the two leading jets within the pseudorapidity range $|\eta| < 2.5$ and with pseudorapidity separation $|\Delta\eta| < 1.3$. The QCD background is suppressed using jet substructure tagging techniques and/or b-tagging, which identify boosted bosons decaying into hadrons. In particular, we use the invariant mass of pruned jets and the N -subjettiness ratios τ_{21} and τ_{42} , as well as b tagging applied to the subjects of the Higgs jet, to discriminate against the initially overwhelming QCD background. The remaining QCD background is estimated from a fit to the parameterized shape. We have searched for the signal as a peak on top of the smoothly falling QCD background. Z' is excluded in resonance mass regions, $[1.0, 1.1]$ and $[1.3, 1.5]$ TeV. While W' is excluded in resonance mass regions $[1.0, 1.6]$ TeV. In HVT B model, W' and Z' are degenerate, so we combine them together. The exclusion limit is set on $[1.0, 2.1]$ TeV.

Table 4.9: Summary of observed limits on resonance masses at 95% CL and their expected values, assuming a null hypothesis. The analysis is sensitive to resonances heavier than 1TeV.

Process	Observed	Expected
	excluded mass limit (TeV)	excluded mass limit (TeV)
$V' \rightarrow VH$	[1.0, 1.7], [1.9, 2.0]	2.0
$Z' \rightarrow HZ(qq)$	[1.0, 1.1], [1.3, 1.5]	1.3
$W' \rightarrow HW(qq)$	[1.0, 1.6]	1.7