

---

# **Search for diboson resonances with CMS and Pixel Barrel Detector Calibration and Upgrade**

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde  
(Dr. sc. nat.)

vorgelegt der Mathematisch-naturwissenschaftlichen Fakultät  
der  
Universität Zürich



von  
Jennifer Ngadiuba

Promotionskomitee

Prof. Dr. Benjamin Kilminster  
Prof. Dr. Florencia Canelli  
Prof. Dr. Laura Baudis  
Prof. Dr. Ueli Straumann

Zürich 2016

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Standard Model and beyond</b>	<b>2</b>
2.1	The Standard Model . . . . .	2
2.1.1	Particles and interactions . . . . .	2
2.1.2	Spontaneous symmetry breaking . . . . .	2
2.1.3	The Higgs mechanism . . . . .	2
2.1.4	The Higgs boson discovery at LHC . . . . .	2
2.2	The hierarchy problem and other SM limitations . . . . .	2
2.3	Theories of new physics . . . . .	2
2.3.1	Warped Extra dimensions . . . . .	2
2.3.2	Compositeness . . . . .	2
2.3.3	Heavy vector triplet . . . . .	2
<b>3</b>	<b>The CMS Experiment at the LHC</b>	<b>3</b>
3.1	The Large Hadron Collider . . . . .	3
3.2	The CMS Detector . . . . .	6
3.2.1	Tracking detectors . . . . .	8
3.2.2	Calorimetry . . . . .	10
3.2.3	Muon detectors . . . . .	12
3.2.4	The trigger system . . . . .	13
<b>I</b>	<b>Search for diboson resonances with CMS</b>	<b>17</b>
<b>4</b>	<b>Diboson resonances as signature for new physics</b>	<b>18</b>
<b>5</b>	<b>Event simulation</b>	<b>19</b>
5.1	Monte Carlo event generators . . . . .	19
5.2	Simulation of physics processes . . . . .	19
5.2.1	Simulation of signal processes . . . . .	19
5.2.2	Simulation of background processes . . . . .	19
<b>6</b>	<b>Object and event reconstruction</b>	<b>20</b>
6.1	Tracks and vertices . . . . .	20
6.1.1	Track reconstruction . . . . .	20
6.1.2	Vertex reconstruction . . . . .	22
6.2	Electrons . . . . .	23
6.2.1	Electron reconstruction . . . . .	23
6.2.2	Electron trigger . . . . .	25
6.2.3	Electron identification . . . . .	28
6.3	Muons . . . . .	31
6.3.1	Muon reconstruction . . . . .	31
6.3.2	Muon trigger . . . . .	35
6.3.3	Muon identification . . . . .	37
6.4	Jets . . . . .	38

---

6.4.1	Jet clustering algorithms . . . . .	39
6.4.2	Jet reconstruction and calibration . . . . .	41
6.4.3	Identification of b jets . . . . .	46
6.5	Missing transverse energy . . . . .	51
6.6	$W \rightarrow \ell\nu$ reconstruction . . . . .	54
<b>7</b>	<b>Identification of highly boosted <math>W/Z \rightarrow q\bar{q}^{(\prime)}</math> and <math>H \rightarrow b\bar{b}</math></b>	<b>56</b>
7.1	Jet substructure observables . . . . .	56
7.1.1	Pruned jet mass . . . . .	56
7.1.2	N-subjettiness . . . . .	57
7.2	The V tagging algorithm . . . . .	59
7.3	The H tagging algorithm . . . . .	61
<b>8</b>	<b>Final event selection and categorization</b>	<b>64</b>
8.1	Search for a WH resonance in the $\ell\nu b\bar{b}$ final state at $\sqrt{s} = 8$ TeV . . . . .	64
8.1.1	$t\bar{t}$ background rejection . . . . .	64
8.1.2	Final selection and control plots . . . . .	64
8.2	Search for WW/WZ resonances in the $\ell\nu q\bar{q}^{(\prime)}$ final state at $\sqrt{s} = 13$ TeV . . . . .	64
8.2.1	$W/Z$ -jet mass categories . . . . .	64
8.2.2	Final selection and control plots . . . . .	64
<b>9</b>	<b>Background modeling</b>	<b>79</b>
9.1	$W+jets$ background estimate with alpha method . . . . .	79
9.1.1	Description . . . . .	79
9.1.2	Extraction of the $W+jets$ normalization . . . . .	79
9.1.3	Extraction of the $W+jets$ shape . . . . .	82
9.2	Top quark production . . . . .	90
9.3	Systematic uncertainties in the background estimation . . . . .	97
<b>10</b>	<b>Signal modeling and statistical treatment</b>	<b>98</b>
10.1	Signal modeling . . . . .	98
10.1.1	Parametrization of the resonance mass . . . . .	98
10.1.2	Signal efficiency . . . . .	98
10.2	Systematic uncertainties in the signal prediction . . . . .	98
10.3	Testing new resonance hypothesis . . . . .	98
10.3.1	Profile likelihood procedure . . . . .	98
10.3.2	The $CL_s$ method . . . . .	98
10.3.3	Treatment of uncertainties . . . . .	98
<b>11</b>	<b>Results with 8 TeV data</b>	<b>101</b>
11.1	Final $m_{WH}$ distribution . . . . .	101
11.2	Studies on the excess . . . . .	101
11.3	Significance of the data . . . . .	101
11.4	Cross section limits . . . . .	102
<b>12</b>	<b>Results with 13 TeV data</b>	<b>103</b>
12.1	Final $m_{WV}$ distribution . . . . .	103
12.2	Cross section limits . . . . .	103

---

<b>13 Combination of searches for diboson resonances at <math>\sqrt{s} = 8</math> and 13 TeV</b>	<b>106</b>
13.1 Inputs to the combination . . . . .	106
13.1.1 8 TeV VV searches . . . . .	106
13.1.2 13 TeV VV searches . . . . .	106
13.1.3 8 TeV VH searches . . . . .	106
13.1.4 13 TeV VH searches . . . . .	106
13.2 Combination procedure . . . . .	106
13.3 Results . . . . .	106
13.3.1 Limits on $W'$ . . . . .	106
13.3.2 Limits on $Z'$ . . . . .	106
13.3.3 Limits on heavy vector triplet ( $W'+Z'$ ) . . . . .	106
13.3.4 Limits on Bulk Graviton . . . . .	106
13.3.5 Significance at 2 TeV . . . . .	106
<b>14 Conclusions</b>	<b>114</b>
<b>II Calibration and Upgrade of the CMS Pixel Barrel Detector</b>	<b>115</b>
<b>15 introduction chapter: why pixels are so important for physics</b>	<b>116</b>
<b>16 The CMS Pixel Barrel Detector</b>	<b>117</b>
16.1 Design of the CMS Pixel Barrel Detector . . . . .	117
16.2 Detector modules . . . . .	117
16.2.1 Sensor . . . . .	117
16.2.2 Readout Chip . . . . .	117
16.2.3 Token Bit Manager . . . . .	117
16.3 Readout and control system . . . . .	117
16.3.1 Analog readout chain . . . . .	117
16.3.2 Front End Driver . . . . .	117
16.3.3 Supply Tube . . . . .	117
16.3.4 Communication and Control Unit . . . . .	117
16.3.5 Front End Controller . . . . .	117
16.4 Pixel Online Software . . . . .	117
16.5 Performance at $\sqrt{s} = 8$ and 13 TeV . . . . .	117
<b>17 Optimization and commissioning for LHC Run II</b>	<b>118</b>
17.1 Radiation damage after LHC Run I . . . . .	118
17.2 Optimization for LHC Run II . . . . .	118
17.2.1 Overview of pixel calibrations . . . . .	118
17.2.2 Temperature dependence . . . . .	118
17.3 Commissioning for LHC Run II . . . . .	118
17.3.1 Installation into CMS . . . . .	118
17.3.2 Check out of optical connections . . . . .	118
17.3.3 Adjustment of readout chain settings . . . . .	118
17.3.4 Optimisation of signal performance . . . . .	118

<b>18 Phase I Upgrade of the CMS Pixel Barrel Detector</b>	<b>119</b>
18.1 Motivations . . . . .	119
18.2 Summary of changes . . . . .	119
18.3 The digital readout chain . . . . .	119
18.4 The Phase I supply tubes . . . . .	119
18.5 The test stand . . . . .	119
18.6 Supply tubes assembly and commissioning . . . . .	119
18.7 Detector commissioning . . . . .	119
<b>19 Conclusions</b>	<b>120</b>
<b>III Summary</b>	<b>121</b>
<b>Bibliography</b>	<b>123</b>

CHAPTER 1

# Introduction

---

# The Standard Model and beyond

---

## 2.1 The Standard Model

### 2.1.1 Particles and interactions

### 2.1.2 Spontaneous symmetry breaking

### 2.1.3 The Higgs mechanism

### 2.1.4 The Higgs boson discovery at LHC

## 2.2 The hierarchy problem and other SM limitations

## 2.3 Theories of new physics

### 2.3.1 Warped Extra dimensions

### 2.3.2 Compositeness

### 2.3.3 Heavy vector triplet

# The CMS Experiment at the LHC

---

## 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [1] is a proton-proton (pp) collider located at the European Particle Physics Laboratory (CERN) near Geneva, Switzerland. It is situated in the former CERN Large Electron-Positron Collider (LEP) tunnel with a circumference of 27 km about 100 m under ground crossing the border between France and Switzerland. A hadron collider has been chosen to allow higher center-of-mass energies ( $\sqrt{s}$ ) compared to electron-positron colliders, the latter limited by synchrotron radiation due to the low mass of the particles to be accelerated. High centre of mass energies are required for the production of heavy SM particles such as the top quark and the Higgs boson, and to search for new BSM interactions at the TeV scale. For this purpose, the LHC is designed to produce pp collisions up to a center-of-mass energy of 14 TeV, superseding previous high energy hadron colliders by a factor of 7. In addition to colliding protons, the LHC is also capable of accelerating and colliding heavy nuclei, which is, however, not considered in this work.

The LHC is the final element in a succession of machines that accelerate protons to increasingly higher energies. Protons, obtained from a hydrogen source, are first accelerated by a linear accelerator (LINAC 2) to energies of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Finally, the beam is injected in the LHC ring, where it completes several revolutions to reach the targeted energy. The LHC ring and the acceleration chain are sketched in Fig. 3.1.

Inside the ring, the two proton beams circulate in opposite directions in two tubes kept at ultrahigh vacuum, referred as beam pipes. The acceleration of protons inside LHC is made by radio-frequency cavities (400 MHz), giving a 492 keV energy gain per revolution, with a 7 keV loss per turn due to synchrotron radiation. It takes 4 minutes and 20 seconds to fill each LHC ring, and 20 minutes for the protons to reach their maximum energy of 7 TeV. The maximum energy of the protons is limited by the strength of the magnetic field required for keeping the protons inside the ring. For 7 TeV-protons a magnetic field of 8.3 T has to be produced, which can only be reasonably obtained by superconducting magnets. The ring is equipped with 1232 dipole magnets for bending and 392 quadrupole magnets for focussing made of niobium-titanium (NbTi), which are cooled down to a temperature of 1.9 K with the help of super-fluid helium. After acceleration the protons move through the ring in separate bunches of protons with a fixed spatial separation.

The LHC ring has four interaction points at which the two counter rotating beams are made to cross and located in the center of the four LHC experiments. Just prior to collision, particles from the incoming beams must be squeezed closer together in order to maximize the chances of interaction. For this purpose, a system of three quadrupole magnets, so-called inner triplet, is located at both sides of each interaction point, which squeeze the beams and lead them to collisions in the center of the detector. Inner triplets tighten the beam, making it 12.5 times narrower from 0.2 mm down to 16  $\mu\text{m}$  across.

Besides the high center-of-mass energy required for the production of heavy particles, a high event rate has to be obtained to allow the discovery of processes with low production cross sections. The instantaneous luminosity  $\mathcal{L}$  characterizes the interaction rate. For a process with a cross section  $\sigma$ , the interaction rate is given by

$$\frac{dN_{ev}}{dt} = \sigma \mathcal{L}. \quad (3.1)$$

The instantaneous luminosity depends only on the beam parameters and can be written for a Gaussian beam distribution as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \sigma_x \sigma_y}, \quad (3.2)$$

where  $N_b$  is the number of particles per bunch,  $n_b$  the number of bunches per beam,  $f_{rev}$  the revolution frequency,  $\gamma_r$  the relativistic gamma factor, while  $\sigma_x$  and  $\sigma_y$  characterize the widths of the transverse beam profiles in the horizontal and vertical direction, respectively. The number of interaction events in a period of running time of the collider can be derived as

$$N_{ev} = \sigma \int \mathcal{L} dt = \sigma L, \quad (3.3)$$

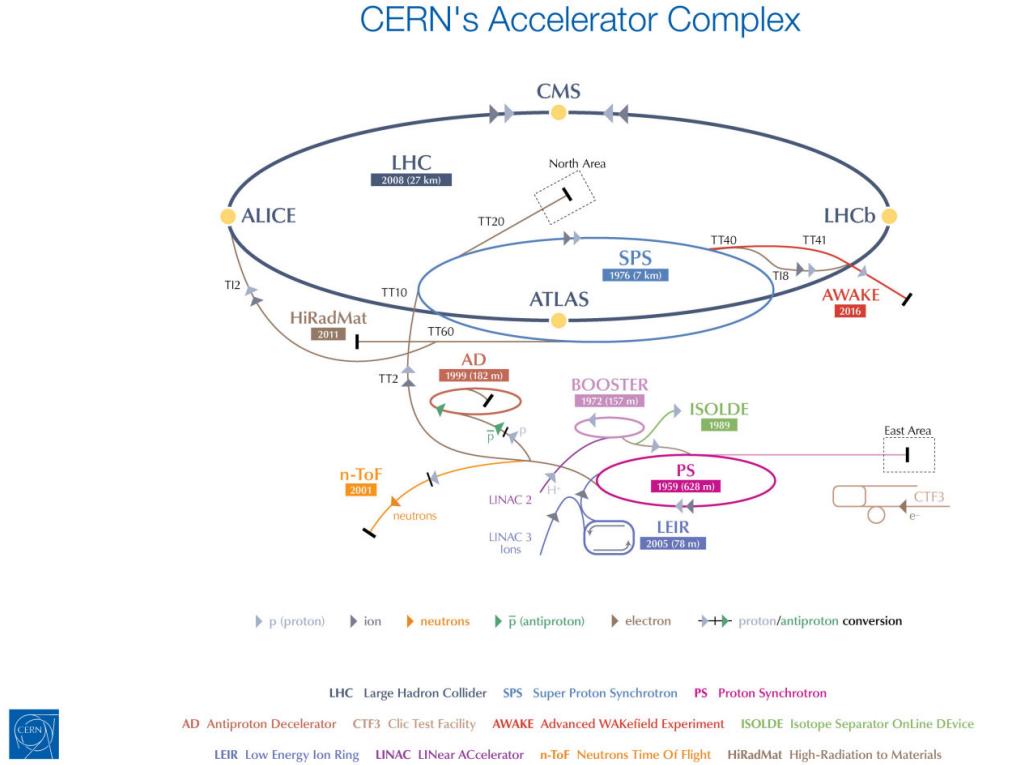
where  $L$  is called the integrated luminosity. It is a measurement of the collected data size and it is usually expressed in inverse of cross section.

The LHC beams can reach very high luminosity with a high frequency bunch crossing and a high density of protons per bunch. In the ring, 2808 bunches of  $1.15 \cdot 10^{11}$  protons are circulated, with an average length of 7.5 cm, a width of about 16  $\mu\text{m}$  and a bunch spacing of 25 ns (collision frequency of 40 MHz). This corresponds to the design instantaneous luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  for pp collisions, which supersedes by a factor of 100 the luminosity reached by previous hadron colliders.

Proton collisions take place in four points of the LHC tunnel where the four main experiments are located: ATLAS (*A Toroidal LHC ApparatuS*) [2], CMS (*Compact Muon Solenoid*) [3], LHCb (*LHC beauty experiment*) [4] and ALICE (*A Lead Ion Collider Experiment*) [5]. ATLAS and CMS are general purpose experiments, designed to get an extensive study of SM and BSM physics and to operate at the design luminosity. The LHCb experiment is instead optimized for bottom quark physics studies while the ALICE experiment is dedicated to the study of the lead-lead collisions at the design luminosity of  $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ .

LHC operation officially started at the beginning of September 2008 but it was interrupted after a short period, due to the breakdown of superconducting magnets. The collider has been reactivated in November 2009 with first pp collisions at  $\sqrt{s} = 900$  GeV, officially starting a new era in the particle physics experiments. The operating center-of-mass energies in pp collisions have so far been 7 TeV in 2010-2011, 8 TeV in 2012 and 13 TeV in 2015-2016. The 7 and 8 TeV periods together make out the *LHC Run 1*, while the 13 TeV period is called the *LHC Run 2*. The work presented in this document is based on both dataset collected at 8 TeV in 2012 and at 13 TeV in 2015.

In the whole Run 1, the LHC operated with a 50 ns bunch spacing. The peak of instantaneous luminosity in 2011 has been  $\sim 0.4 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with a total delivered integrated luminosity of  $6.1 \text{ fb}^{-1}$  [7]. In 2012 the beam energy increased to 4 TeV per beam with a peak luminosity of  $\sim 0.8 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and  $23.3 \text{ fb}^{-1}$  delivered integrated luminosity by the end of that year [7]. The increment of the instantaneous luminosity leads to a no more negligible number of simultaneous interactions per bunch crossing, the so-called *pileup* (PU) events.

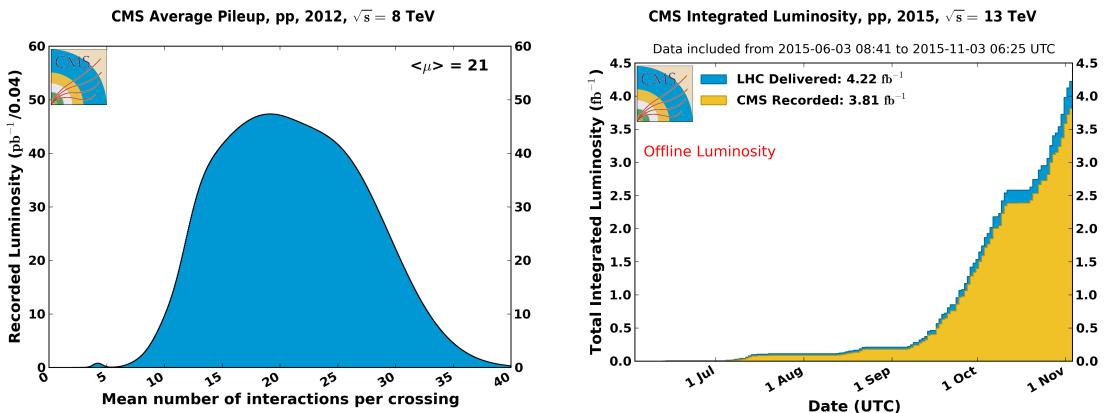


**Figure 3.1:** The LHC scheme together with its injection chain and the locations of the four main experiments ATLAS, CMS, LHCb and ALICE [6].

It depends on the cross section of inelastic collisions (75 mb at  $\sqrt{s} = 8$  TeV [8]) and it is directly linked to the instantaneous luminosity. The average PU of the data collected in 2012 is equal to 21 (Fig. 3.2) while it has been around 15 in 2011 [7].

A shut-down period for the LHC (LS1) occurred in the whole 2013 and 2014, where upgrades and technical improvements have been performed in order to reach the designed instantaneous luminosity and center-of-mass energy. On March, 21st 2015 the first pp collisions at  $\sqrt{s} = 13$  TeV has been obtained, a new record-breaking energy. For the first three months the machine operated with 50 ns bunch spacing while, from August 2015, it has been reduced to the designed 25 ns and the number of bunches per beam has been increased. The first part of this Run 2 phase ended on November 2015 with a total delivered integrated luminosity of  $4.2 \text{ fb}^{-1}$  and a peak luminosity of  $\sim 0.5 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with an average pileup of 12 [7].

The LHC Run 2 has been restarted in April 2016, after an end-of-the-year technical stop, reaching a peak luminosity of  $\sim 1.2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  (change this number at some point). The machine has remained in operation at  $\sqrt{s} = 13$  TeV for the whole year with a total delivered integrated luminosity of  $30 \text{ fb}^{-1}$  (change this number at some point). Accordingly to the current LHC schedule, the Run 2 will proceed up to the end of 2018 with a total expected integrated luminosity of  $100 \text{ fb}^{-1}$ . The data collected in 2016 are not considered in this work.



**Figure 3.2:** Mean number of interactions per bunch crossing in data collected in 2012 by the CMS experiment at LHC (left). Cumulative luminosity versus day (right) delivered by LHC (blue) in 2015; the off-line luminosity recorded by the CMS experiment is also reported (orange). [7]

### 3.2 The CMS Detector

The CMS detector is a general purpose detector installed 100 m underground at the LHC interaction point 5 (P5) near the village of Cessy in France. It has been designed to exploit the different properties of the wide range of particles and phenomena produced in high-energy collisions in the LHC. The design of the CMS detector is driven by the challenges of a physics experiment in the LHC environment. Many of the physics benchmark channels have a small cross section and the background from QCD jet production is overwhelmingly dominant. In order to achieve high rejection power with an optimal efficiency for rare channels, the detector has to be able to reconstruct the primary interaction entirely and to reduce the influence of overlapping events. Therefore, one needs to collect all possible information on the particles passing through the detector. Since these have different properties, a mixture of sub-detectors is required for a complete event reconstruction. The reconstruction of lepton signatures is essential for the extraction of rare processes and an excellent muon and electron identification and momentum resolution is desired. A precise measurement of secondary vertices and impact parameters is necessary for an efficient identification of heavy flavor quarks and  $\tau$ -leptons. Moreover, a large hermetic geometric coverage is preferred, which allows for a precise estimate of the transverse momentum carried away by invisible particles through the sum of all visible particles.

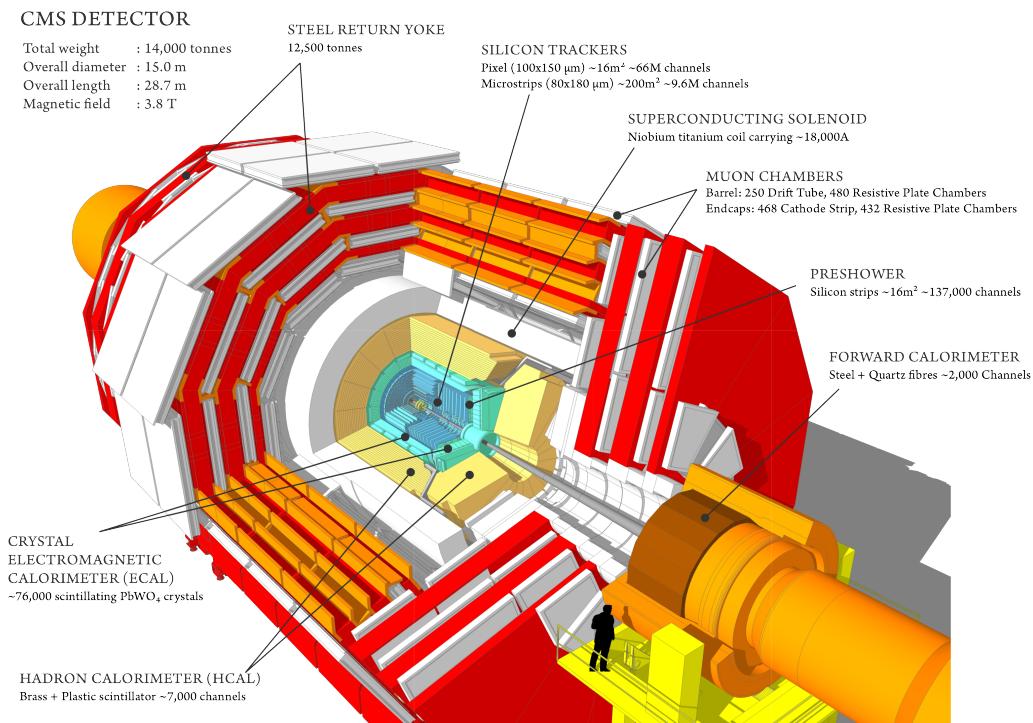
The high peak luminosities of LHC lead to large pileup imposing further challenges to the design. As a consequence of pileup, the products of an interaction under study may be confused with those from other interactions in the same bunch crossing. This effect can be reduced by using high-granularity detectors resulting in low occupancy. In addition, the short bunch crossing requires fast response time and good time resolution of each detector element in order to discriminate the interaction under study from the interactions occurring in neighbouring bunch crossings. Hence, a large number of detector channels and an excellent synchronization among them are required. Another challenge arises from the large flux of particles near the interaction point which leads to high radiation levels and the need of radiation hard detectors and front-end electronics.

Figure 3.3 shows the layout of the CMS detector. The detector is built in a cylindrical structure composed of a barrel in the center and endcaps at both sides. This structure is

21.6-m-long, 14.6-m in circumference and 12500-tons-heavy. The detector design and layout was driven by the choice of the magnetic field configuration. Large bending power is needed for a precise measurement of the momentum of high-energy charged particles. Within the CMS detector this is achieved by a superconducting solenoid with a length of 12.9 m and an inner diameter of 5.9 m generating a magnetic field of 4 T. The bore of the magnet coil is large enough to accommodate the inner tracker and the calorimetry inside. The inner tracker consists of a pixel and a strip detector both made out of Silicon, and it is the key component of CMS to measure the momenta of charged particles and identify primary and secondary vertices. The calorimetry system comprises a crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadronic calorimeter (HCAL), which provide information on the energies and directions of all charged and neutral particles. Outside the magnet are the large muon detectors, which, integrated inside the return yokes of the magnet, provide identification of muons and measurement of their momenta.

For the description of the CMS detector the following coordinate system is used. The origin is centered at the nominal collision point inside the experiment with the  $y$ -axis pointing vertically upward, the  $x$ -axis pointing radially inward toward the center of the LHC, and the  $z$ -axis points along the beam direction. The azimuthal angle  $\phi$  is measured from the  $x$ -axis in the  $x$ - $y$  plane. The polar angle  $\theta$  is measured from the  $z$ -axis. Pseudorapidity is defined as  $\eta = -\ln \tan(\theta/2)$ . Thus, the momentum and energy measured transverse to the beam direction, denoted by  $p_T$  and  $E_T$ , respectively, are computed from the  $x$  and  $y$  components.

In the following sections the three main components of the CMS detector will be described together with a section on the triggering system.

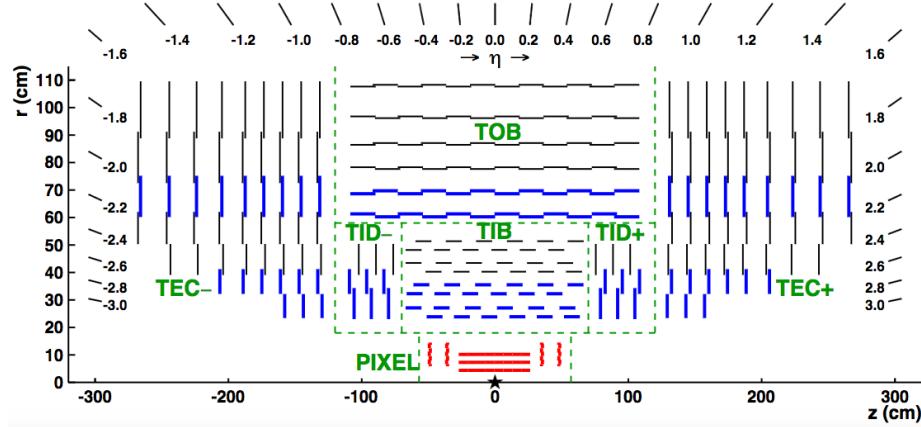


**Figure 3.3:** Layout of the CMS experiment and its sub-detectors.

### 3.2.1 Tracking detectors

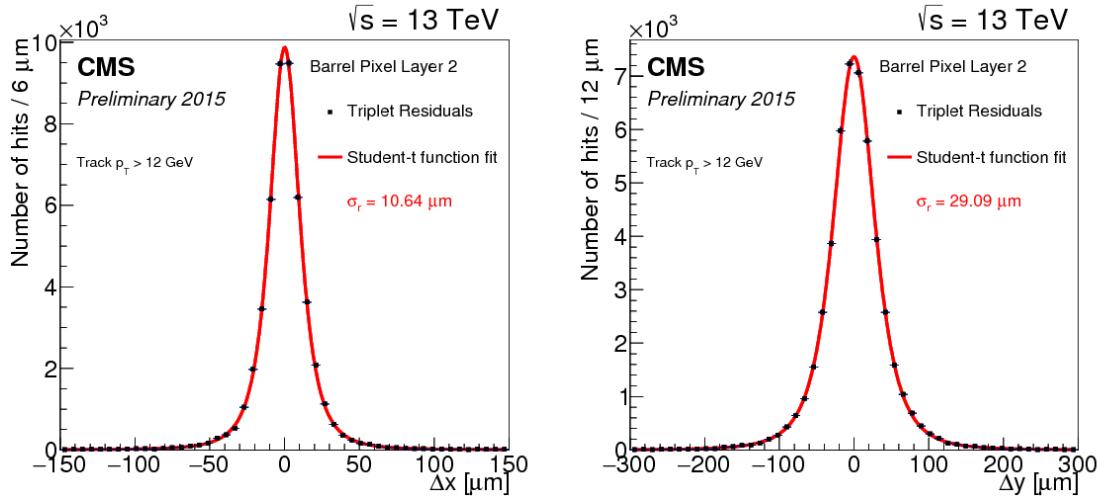
The tracking system of CMS (Fig. 3.4) is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions, as well as a precise reconstruction of secondary vertices [9]. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m providing coverage up to  $|\eta| < 2.5$ . In order to achieve high tracking efficiency at the high luminosities of LHC, a detector technology featuring granularity, speed and radiation hardness is required. Furthermore, the material budget of the tracking system has to be as low as possible in order to avoid a worsening of the tracking efficiency and resolution due to material interaction effects of the charged particle, such as multiple scattering, bremsstrahlung, photon conversion or nuclear interactions. These requirements lead to a tracker design entirely based on silicon detector technology. With about  $200 \text{ m}^2$  of active silicon area the CMS tracker is the largest silicon tracker ever built. It is divided into a pixel detector close to the interaction region and a strip detector in the outer region. At LHC design luminosity more than 1000 particles are hitting the tracking volume in each bunch crossing. This leads to a hit rate density of  $1 \text{ MHz/mm}^2$  at a radius of 4 cm which imposes severe challenges to the design of the tracking detectors. With a pixel size of  $100 \times 150 \mu\text{m}^2$  in  $r\text{-}\phi$  and  $z$ , respectively, an occupancy of the order of  $10^{-4}$  per pixel and LHC bunch crossing can be achieved. The hit rate density falls with the distance from the interaction point to  $60 \text{ kHz/mm}^2$  at a radius of 22 cm and to  $3 \text{ kHz/mm}^2$  at a radius of 115 cm. Therefore, at intermediate radii (20–55 cm), silicon micro-strip detectors are used, with a typical cell length of 10 cm and a pitch of  $80 \mu\text{m}$ . At the outermost radii (55–110 cm) the strip size can be further increased to  $25 \text{ cm} \times 180 \mu\text{m}$ . With this choice an occupancy of less than 3% is maintained in the strip detector. However, the strip capacitance scales with its length and therefore the electronics noise is a linear function of the strip length as well, becoming not negligible in the outermost region where the strip size is the largest. In order to maintain a good signal to noise ratio of well above 10, CMS uses thicker silicon sensors for the outer tracker region ( $500 \mu\text{m}$  thickness as opposed to the  $320 \mu\text{m}$  in the inner tracker) with correspondingly higher signal. To mitigate the radiation damage effects and prolong the lifetime of the detector modules, the tracking detectors are designed to run at subzero temperatures. The cooling is established using a mono-phase liquid cooling system with  $\text{C}_6\text{F}_{14}$  as cooling fluid. The whole tracker system operated at  $+4^\circ\text{C}$  during Run 1. After this phase, several improvements have been implemented and an operative temperature of  $-15^\circ\text{C}$  is currently maintained for Run 2.

The pixel detector is built from 3 barrel layers at radii of 4.4, 7.3 and 10.2 cm (BPix) and two end disks (FPix) on each side at a distance of  $z = \pm 34.5, \pm 46.5 \text{ cm}$  from the interaction point. It consists of 1440 segmented silicon sensor modules with a total of 66 million readout channels covering an area of about  $1 \text{ m}^2$ . The pixel detector is essential for the reconstruction of secondary vertices from bottom quarks and  $\tau$  leptons decays. It provides precise tracking points in  $r\text{-}\phi$  and  $z$  and therefore is responsible for a small impact parameter resolution that is important for good secondary vertex reconstruction. This is achieved thanks to the read out of the analog pulse height information. The sensor surface in the barrel layers is parallel to the magnetic fields, hence the charge carriers produced by a particle traversing experience a Lorentz drift, which leads to charge spreading over more than one pixel. The analog pulse height information can be used to calculate a center of gravity of the charge distribution improving the hit information. The forward detectors are tilted at  $20^\circ$  in a turbine-like geometry to induce charge-sharing. As shown in Fig. 3.5, a spatial resolution of  $10 \mu\text{m}$  in the transverse plane and  $30 \mu\text{m}$  in the longitudinal plane can be achieved for BPix. For FPix a spatial resolution of  $20 \mu\text{m}$  is obtained. A detailed description of the design and



**Figure 3.4:** Longitudinal section of half of the CMS Tracker system; the different detector types are indicated.

the functioning of the CMS pixel barrel detector is given in Chapter 16.



**Figure 3.5:** Distribution of hit residuals on pixel barrel layer 2 in the transverse (left) and longitudinal (right) direction with respect to the beam. The distributions are fitted with a Student's t-function for which sigma is shown on the plot [10].

The strip detector occupies the radial region between 20 cm and 1.16 m. As illustrated in Fig. 3.4, it is composed of four subsystems: the four-layer Tracker Inner Barrel (TIB), the six-layer tracker outer barrel (TOB) and on each side three-disk Tracker Inner Disks (TID) and nine-disk Tracker Endcaps (TEC). The silicon micro-strip sensors have strips parallel to the beam axis in the barrel and radial on the disks. The modules in the first two layers and rings, respectively, of TIB, TID, and TOB as well as rings 1, 2, and 5 of the TECs carry a second micro-strip detector module which is mounted back-to-back with a stereo angle of 100 mrad in order to provide a measurement of the second co-ordinate ( $z$  in the barrel and  $r$  on the disks). This tracker layout ensures at least 9 hits in the silicon strip tracker in the full range of  $|\eta| < 2.4$  with at least 4 of them being two-dimensional measurements. The total

number of silicon sensors in the strip tracker is 24244, making up a total active area of 198 m<sup>2</sup>, with about 9.3 million of strips.

### 3.2.2 Calorimetry

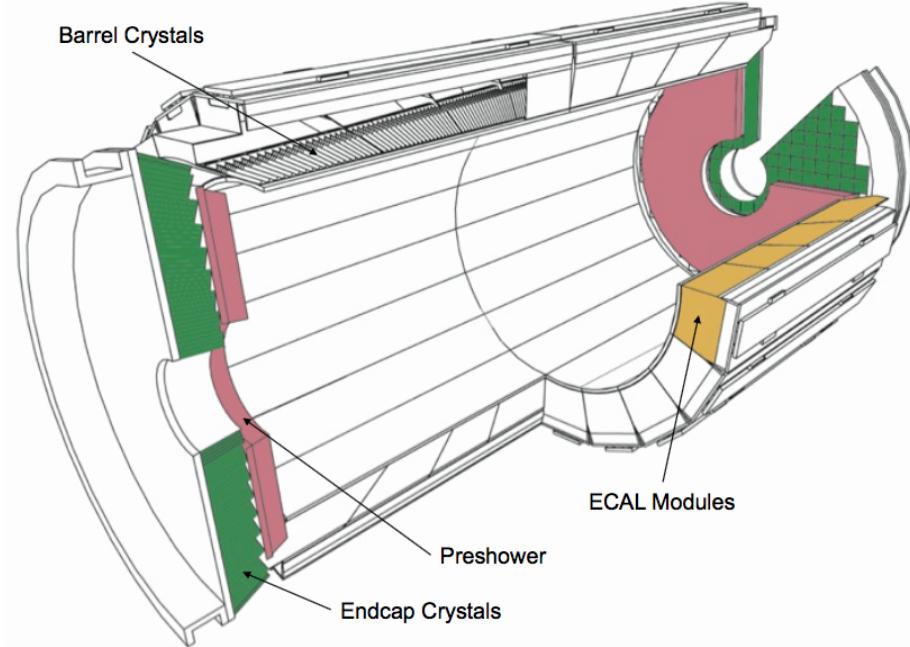
The calorimeter measures the energies and directions of all neutral and charged particles traversing the detector, with the exception of muons and neutrinos. It consists of two parts, the electromagnetic calorimeter (ECAL) [11] and the hadronic calorimeter (HCAL) [12].

The goal of ECAL is to measure precisely the energy of electrons and photons which generate electromagnetic showers inside it. It is a hermetic and homogeneous calorimeter with a large pseudorapidity coverage up to  $|\eta| < 3$ . As illustrated in Fig. 3.6, ECAL is divided into barrel and endcap detectors consisting of scintillation crystals made from lead tungstate (PbWO<sub>4</sub>). The choice of this material is motivated by its high density (8.28 g/cm<sup>3</sup>), short radiation length ( $X_0 = 0.89$  cm) and small Molière radius (2.2 cm), resulting in a high stopping power, fine granularity and therefore a compact calorimeter able to fit inside the solenoid. The ECAL comprises 61200 crystals in the barrel and 7324 crystals in each of the 2 endcaps, for a total volume of 8.14 m<sup>3</sup> and 2.9 m<sup>3</sup>, respectively. The crystals have a tapered shape and are mounted in a quasi-projective geometry. The barrel extends radially between 1.29 and 1.75 cm covering the region  $|z| < 3.05$  m and  $|\eta| < 1.479$ . The crystals have a front face cross-section of 22×22 mm<sup>2</sup> and a length of 2.3 cm (25.8  $X_0$ ). They are organised in 36 identical supermodules each covering 20° in  $\phi$ . The crystals are contained in a thin-walled glass-fibre alveola structures (“submodules”) with 2( $\phi$ )×5( $\eta$ ) crystals per each resulting in a granularity 360-fold in  $\phi$  and 2×85-fold in  $\eta$ . The endcaps are placed at a distance of 3.14 m from the interaction point and they extend radially between 3.16 and 17.11 cm, covering the region 1.479 <  $|\eta| < 3.0$ . The crystals have a front face cross section of 28.6×28.6 mm<sup>2</sup> and a length of 2.2 cm (24.7  $X_0$ ). A preshower detector with a thickness of 3  $X_0$  is placed in front of the endcaps (1.653 <  $|\eta| < 2.6$ ) to guarantee a reliable discrimination of single photons and photons produced in pairs in neutral pion decays. The relatively low light yield of the crystals (30 γ/MeV) requires use of photodetectors with intrinsic gain that can operate in a magnetic field. Silicon avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps. The light output and the amplification have a strong temperature dependence. The response to an incident electron changes by (3.8±0.4)%/°C which in turn means that the temperature has to be closely monitored and kept stable to a precision of ±0.05°C. The nominal operating temperature of the ECAL is 18°C and is provided by a water cooling system.

The energy resolution of the electromagnetic calorimeter can be parametrized by the following expression:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E(\text{GeV})}} \oplus \frac{N}{E(\text{GeV})} \oplus C. \quad (3.4)$$

The first term is stochastic including contributions from the shower containment, the number of photoelectrons and the fluctuations in the gain process. The second contribution corresponds to the noise term, which includes noise in the readout electronics and fluctuations in pile-up. The third term is a constant and dominates the energy resolution for high-energy electron and photon showers, depends on non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter, single-channel response uniformity and stability. The value of the three coefficients were determined by a electron test beam measurement in a matrix of 33 crystals to be S = 2.8%, N = 12% and C = 0.3% [13].



**Figure 3.6:** Schematic view of the CMS Electromagnetic Calorimeter [3].

The energy measurement of the ECAL is complemented by the measurement of the hadronic calorimeter. The HCAL is designed to be as near to hermetic around the interaction region as possible to allow events with missing energy to be identified. It is a sampling calorimeter composed of layers of brass absorber interlaced with tiles of plastic scintillators as active material to detect the showers generated by the hadrons in the brass. The energy released in the scintillator tiles causes them to emit blue-violet light, a fraction of which is absorbed and re-emitted by embedded wavelength-shifting fibres in the green region of the spectrum. The green light is then carried by special fibre-optic waveguides to the readout system. The photodetection readout is based on multi-channel hybrid photodiodes (HPDs), photodetectors configured especially for CMS that can provide gain and operate in a high magnetic field. Figure 3.7 shows a schematic cross section of the HCAL detectors. The hadron barrel (HB) and endcaps (HE) calorimeters sit behind the tracker and the electromagnetic calorimeter as seen from the interaction point. The HB is radially restricted between the outer extent of the electromagnetic calorimeter ( $r = 1.77$  m) and the inner extent of the magnet coil ( $r = 2.95$  m). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an Outer Hadron (HO) calorimeter is placed outside the solenoid complementing the barrel calorimeter. The HO uses the solenoid as additional absorbing material and provides sufficient containment for hadronic showers with a thickness of 11.8 interaction lengths ( $\lambda_l$ ). The first scintillators are placed in front of the first absorber plate in order to sample showers developing in the material between the ECAL and the HCAL, while the last scintillators are installed after the last absorber plate to correct for late developing showers leaking out. 70000 and 20916 scintillator tiles are installed in the HB and the HE, respectively. The HB and HE cover the region  $|\eta| < 1.3$  and  $1.3 < |\eta| < 3.0$ , respectively. Beyond  $|\eta| = 3$ , the Hadron Forward (HF) calorimeter placed at 11.2 m from the interaction point extends the pseudorapidity coverage down to  $|\eta| = 5.2$ . The barrel depth goes from  $5.46 \lambda_l$  at  $\eta = 0$  to  $10.8$  at  $\eta = 1.3$ , while the endcaps coincide with an

average of  $11 \lambda_l$ . The calorimeter is segmented and arranged in towers as summarized in Table 3.1. The HF is positioned at a longitudinal distance of 11.2 m from the interaction point. It will experience unprecedented particle fluxes with an energy of 760 GeV deposited on average in a proton-proton interaction at  $\text{ps} = 14 \text{ TeV}$ . This energy has to be compared to the average of 100 GeV deposited in the rest of the detector. The situation is even more severe as the energy is not spread equally among the HF, but has a pronounced peak at the highest rapidity. The HF is a sampling calorimeter made from steel absorber plates composed of 5mm thick grooved plates with quartz fibers inserted as active medium. The signal is generated when charged shower particles above the threshold generate Cherenkov light in the quartz fibres, thereby rendering the calorimeter mostly sensitive to the electromagnetic component of showers.

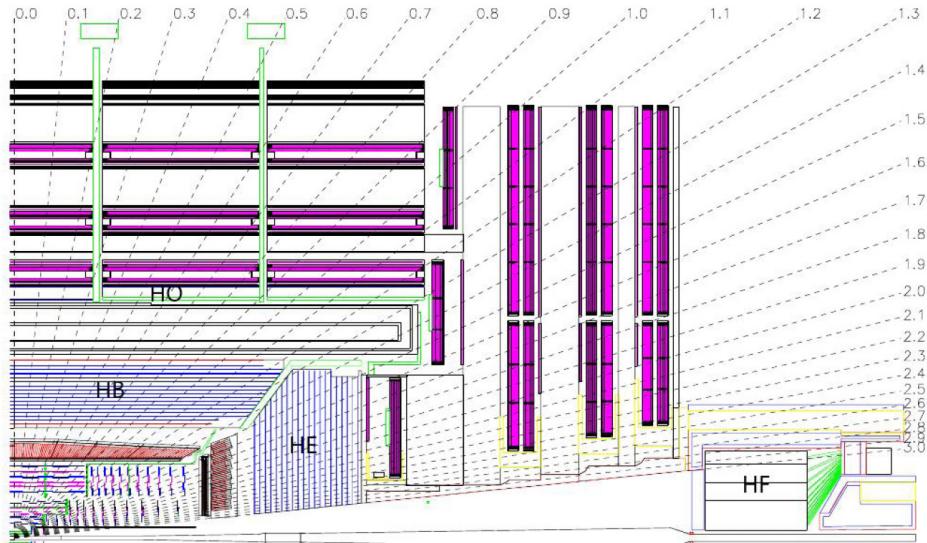
The HCAL energy resolution is

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E(\text{GeV})}} \oplus 5\% \quad (3.5)$$

where  $a$  is 65% in the barrel, 85% in the endcaps and 100% in the forward calorimeter.

**Table 3.1:** Tower segmentation in azimuthal and polar angle for the hadronic barrel, endcap and forward calorimeter.

	HB/HO	HE ( $ \eta  \leq 2.5$ )	HE ( $ \eta  > 2.5$ )	HF ( $ \eta  \leq 4.7$ )	HF ( $ \eta  > 4.7$ )
$\Delta\phi \times \Delta\eta$	$0.087 \times 0.087$	$0.087 \times 0.087$	$0.175 \times 0.175$	$0.175 \times 0.175$	$0.175 \times 0.35$



**Figure 3.7:** Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [3].

### 3.2.3 Muon detectors

The muon system is the outermost part of the CMS detector. It is located in the steel return yoke of the solenoid, covering the pseudorapidity region  $|\eta| < 2.4$ . This is possible because muons are hardly affected by this large material budget. The coil acts as a shield from electromagnetic and hadronic particles escaping the calorimeters and the yoke provides

a magnetic field between consecutive muon stations, allowing a momentum measurement independent from the inner tracker. The muon system is designed for three major functions: robust and fast identification of muons, good resolution of momentum measurement, and integration to a fast and reliable trigger system. The gaseous detectors has been chosen as muon detectors since they are robust and with a relative fast response. Furthermore, the area to be covered is extremely wide and a gaseous detector system allows to reduce the cost and the amount of readout channels. The muon system is thus composed of three types of gaseous detectors arranged in barrel and endcap sections, as shown in Fig. 3.8: Drift Tubes (DTs), Resistive Plate Chambers (RPCs) and Cathode Strip Chambers (CSCs). The choice of different detector topologies lies essentially in the different expected particle rates.

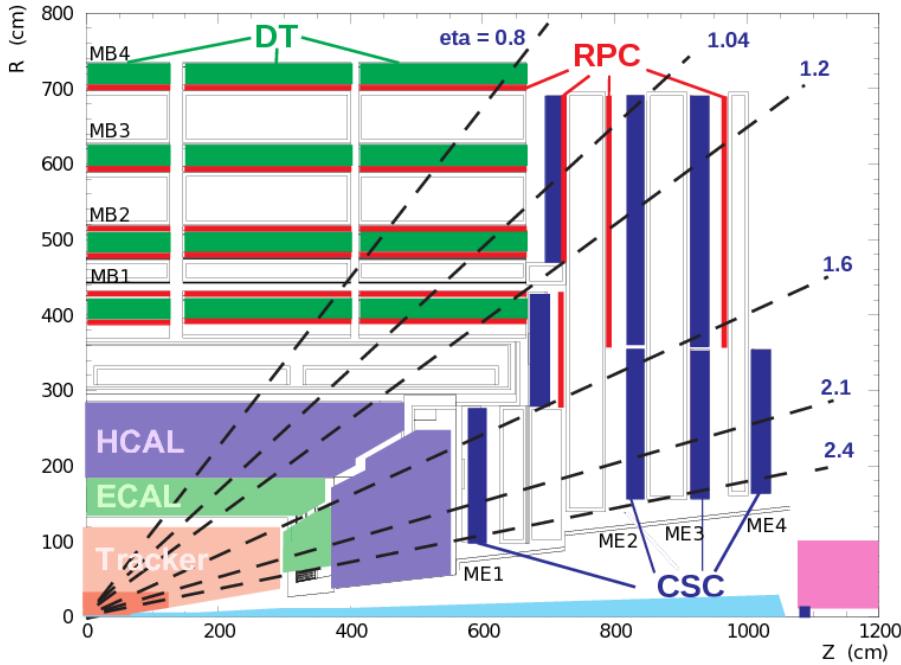
In the barrel region, where the neutron-induced background is small, the muon rate is low, and the 4-T magnetic field is uniform, DTs with standard rectangular drift cells are used covering the pseudorapidity region  $|\eta| < 1.2$ . A DT cell is a 4 cm wide gas tube with a positively charged stretched wire inside. The barrel DT chambers are organized in five separate wheels. Each wheel is divided into 12 sectors, each covering a  $30^\circ$  azimuthal angle. In each of the 12 sectors there are 4 chambers per wheel which are concentric around the beam line and separated by the iron return yoke. Each DT chamber, on average  $2\text{ m} \times 2.5\text{ m}$  in size, consists of 12 layers of DT cells, arranged in three groups of four. For the first 3 stations in each wheel, the middle group measures the  $z$  coordinate while the two outside groups measure the  $r\phi$  coordinate. The fourth and outermost station does not contain the  $z$ -measuring planes. Each one of the 250 DT chambers has a resolution of  $\sim 100\text{ }\mu\text{m}$  in  $r\phi$  and up to  $150\text{ }\mu\text{m}$  in  $z$ , and can measure the particle direction with 1 mrad accuracy.

In the two endcap regions of CMS, where the muon rates and background levels are high and the magnetic field is large and non-uniform, CSCs are used with their fast response time, fine segmentation, and radiation resistance, covering the pseudorapidity region between 0.9 and 2.4. Each CSC is trapezoidal shaped multiwire proportional chambers which consists of 6 gas gaps, each gap having a plane of radial cathode strips and a plane of anode wires running almost perpendicularly to the strips. The gas ionization and subsequent electron avalanche caused by a charged particle traversing each plane of a chamber produces a charge on the anode wire and an image charge on a group of cathode strips. Thus, each CSC provides a two-dimensional position measurement, where the  $r$  and  $\phi$  coordinates are determined by the cathode strips and the anode wires, respectively. 540 CSC are arranged in 4 disks per endcaps divided in concentric rings (3 rings in the innermost station, 2 in the others). Each chamber has a spatial resolution of about 200 mm in  $r$ , and  $75 \times 150\text{ }\mu\text{m}$  in the  $r\phi$  coordinate.

In addition, there is a total of 610 RPCs added in both the barrel and endcap regions to provide a fast, independent, and highly-segmented trigger over a large portion of the rapidity range ( $|\eta| < 1.6$ ) of the muon system. They produce a fast response, with good time resolution ( $\sim 2\text{ ns}$ ) but coarser position resolution than the DTs or CSCs. RPCs are made from two high resistive plastic plates with a voltage applied and separated by a gas volume. The signal generated by the muon when passing through the gas volume is detected by readout strips mounted on top of one of the plastic plates. Six layers of RPCs are installed in the barrel muon system, two layers in each of the first two stations and one layer in each of the last two stations. One layer of RPCs is built into each of the first three stations of the endcap.

### 3.2.4 The trigger system

The LHC provides proton-proton and heavy-ion collisions at unprecedented high luminosity and interaction rates. Given the high segmentation of the CMS detector, about 100 million readout channels are present and this corresponds to an enormous volume of data at the



**Figure 3.8:** A longitudinal view of one quarter of the CMS experiment; the three muon detectors detector types are highlighted.

detector front-ends. At the design luminosity and collision frequency, each crossing produces approximately 1 MB of zero-suppressed data resulting in a raw data rate of about 40 TB per second. These figures are many orders of magnitude larger than the archival storage capability of  $\sim 1$  kHz at data rates of  $\mathcal{O}(10^2)$  MB/s. Technical difficulties in handling, storing and processing such extremely large amounts of data impose a reduction factor on the rate of events that can be written to permanent storage. This task is performed by the trigger system, which is the baseline of the physics event selection process. The key point of the trigger system is a fast time rejection of all the “non-interesting” events. This can be done by exploiting event topologies common to group of physics processes, such as the presence of one or more leptons in the event. The trigger system needs to be as inclusive as possible, in order to collect data for all the physics searches that can be performed looking at pp collision, but it has also to operate within the CMS time restriction and to not saturate the storage capability. The required rejection power of  $\mathcal{O}(10^5)$  is too large to be achieved in a single processing step, if a high efficiency is to be maintained for the physics phenomena CMS plans to study. For this reason, the full selection task is split into two steps. The first step (Level-1 Trigger) is designed to reduce the rate of events accepted for further processing to less than 100 kHz. The second step (High-Level Trigger or “HLT”) is designed to reduce this maximum L1 accept rate of 100 kHz to a final output rate of 1 kHz.

The L1 Trigger is built from custom designed, programmable electronics and is housed partly on the detectors, partly in the underground control room located at a distance of approximately 90 m from the experimental cavern. It is designed to take a fast accept/reject decision every bunch crossing, on the basis of a rough reconstruction of the event. The detector information used at L1 are coarsely segmented data from the calorimeters and the muon system only. Within a time budget of  $3.2 \mu\text{s}$ , it has to decide if an event is discarded or kept, and transfer this decision back to the sub-detectors, which in the meantime keep the high resolution data in the front-end electronics. Figure 3.9 shows the L1 Trigger architecture:

it has local, regional and global trigger components.

Trigger primitives are generated by calculating the transverse energy of a trigger tower and assigning it to the correct bunch crossing. A regional calorimeter trigger then determines regional electron, photon and jet candidates and information relevant for muon and tau identification. The global calorimeter trigger provides information about the jets, the total transverse energy and the missing energy in the event and identifies the highest-ranking trigger candidates.

In the muon system all three types of detectors take part in the trigger decision. The DT chambers provide track segments in the projection and hit pattern in  $\eta$ , while the CSC determine three-dimensional track segments. The track finders in the DT chambers and the CSCs calculate the transverse momentum of a track segment and its location and quality. The RPCs deliver an independent measurement derived from regional hit patterns. The global muon trigger receives up to four candidates from each subsystem (DT, barrel RPC, CSC and endcap RPC) together with the isolation information from the global calorimeter trigger. The aim is to improve the efficiency and to reduce the rate by making use of the complementarity and the redundancy of the subsystems. In the end, the global muon trigger selects a maximum of four muon trigger candidates and determines their momentum, charge, position and quality.

The trigger objects extracted by the global calorimeter trigger and the global muon trigger are sent to the global trigger where the decision to accept or reject an event is taken and distributed to the sub-detectors. The simplest triggers are in general those based on the presence of one object with an  $E_T$  or  $p_T$  above a predefined threshold (single-object triggers) and those based on the presence of two objects of the same type (di-object triggers) with either symmetric or asymmetric thresholds. Other requirements are those for multiple objects of the same or different types (“mixed” and multiple-object triggers). Up to 128 algorithms can be executed in parallel. The decision is also based on the readiness of the sub-detectors and the data acquisition system (DAQ), which is supervised by the Trigger Control System (TCS). The Level-1 Accept (L1A) decision is communicated to the sub-detectors through the Timing, Trigger and Control (TTC) system.

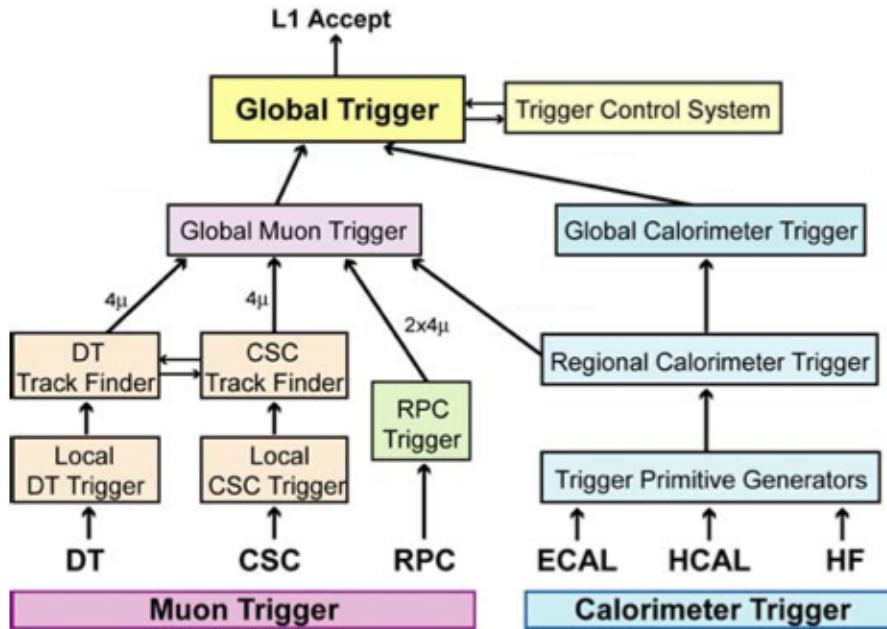
If an event is accepted by the L1 trigger, the full detector information (1 MB) is read out by the DAQ system and passed to the HLT system for further analysis. The HLT is a special part of the CMS software which runs on a farm of several thousand processors performing high-level object reconstruction and analysis. Each processor works on the reconstruction of one event at a time, to get to a trigger decision within on average 100 ms. Since the time budget for one event is much larger than at the L1 trigger, more complicated algorithms, including tracking, can be executed at the HLT. Once an event is accepted, it is stored on disk and fully reconstructed offline at a later time.

The full detector readout is available at HLT, but in order to meet the timing requirements given by the input rate from L1, events are discarded before being fully reconstructed, as soon there is enough reconstructed information to take the decision. Therefore the selection is organized in a sequence of logical steps. The Level-2 uses the full information from calorimeters and muon detectors to reconstruct the physical objects and to reduce the event rate by roughly one order of magnitude. The data from the silicon tracker represent almost 80% of the event size and require complex and time consuming algorithms for the reconstruction. For this reason this information is used only during the Level-3 selection.

The HLT consists of approximately 400 trigger paths. Each trigger path starts from the seed provided by the L1 trigger and it is built from reconstruction modules and filter modules. After some parts of the data are reconstructed, a filter module decides if the reconstructed objects pass the thresholds and the next step in reconstruction is started, or if the event is

not accepted by the path. In the later case, the execution of the path is stopped and the following reconstruction steps and filter steps are not performed to save computation time. If an event is not accepted by a path, it can still be accepted by a different path.

If, for some paths with low thresholds, the acceptance rate is too high, they can be prescaled to lower the rate. A prescale value of ten means, for example, that the path is executed only for every tenth event that was accepted by the L1 trigger, and, consequently, the trigger rate for that path is ten times smaller. The prescale value for one trigger path has several predefined levels, depending on the instantaneous luminosity of the LHC machine. During an LHC fill, the instantaneous luminosity decreases, and the prescale values can be changed during a CMS run to keep the global trigger rate at an optimal level.



**Figure 3.9:** Architecture of the Level-1 Trigger [3].

## **Part I**

# **Search for diboson resonances with CMS**

CHAPTER 4

# Diboson resonances as signature for new physics

---

# Event simulation

---

## 5.1 Monte Carlo event generators

## 5.2 Simulation of physics processes

### 5.2.1 Simulation of signal processes

### 5.2.2 Simulation of background processes

# Object and event reconstruction

---

In the proton-proton collisions at the LHC a large number of particles are produced which must be properly reconstructed and identified. These particles travel through the CMS detector and they are classified in objects depending on their specific signature in each subdetector. This chapter covers the reconstruction of physics objects that are needed for the identification of signal events in the lepton plus jet event topology as described in Chapter 4.

The measurement of a track in the tracker detector for charged particles and the reconstruction of the primary vertices represent key aspects of the reconstruction of the various objects and are detailed in Sec. 6.1. In this analysis  $\tau$ -leptons are reconstructed as electrons (Sec. 6.2) or muons (Sec. 6.3) and accounted to the respective channel if they decay leptonically, or as jets (Sec. 6.4) if they decay hadronically. Only the leptonic decay mode contributes to the analysis since at least one muon or electron has to be reconstructed in the event. However, the resulting gain in sensitivity is limited by the small branching ratios. In addition to leptons and jets, the last type of particle present in the final state is the neutrino, whose presence can be inferred from an imbalance of the transverse energy (Sec. 6.5). The Identified lepton and the missing transverse energy in the event are associated with the  $W \rightarrow \ell\nu$  candidate which is entirely reconstructed through the algorithm described in Sec. 6.6.

**FIXME:** Particle Flow

## 6.1 Tracks and vertices

The reconstruction of tracks of charged particles allows for their momentum measurement and aids in particle identification as described later. The reconstruction of the tracks' vertices is important to distinguish the primary interaction, i.e. the hard interaction, from additional interactions that might take place in the event and also for the identification of secondary vertices of jets that contain c or b quarks called c-/b-tagging (see Sec. 6.4.3).

### 6.1.1 Track reconstruction

The track reconstruction at CMS [14] is based essentially on information coming from the silicon tracker system. A charged particle passing through a tracker layer can in general induce a signal in more than one pixel or more than one strip. The first step of the tracking procedure is the clusterization which assembles nearby tracker channels into one hit cluster. The particle position and its uncertainty is then inferred from the relative signal amplitudes in each channel.

Due to the magnetic field charged particles travel through the tracking detectors on a helix trajectory which is described by 5 parameters: the curvature  $k$ , the track azimuthal angle  $\phi$  and polar angle  $\theta$ , the signed transverse impact parameter  $d_0$  and the longitudinal impact parameter  $z_0$ . The transverse (longitudinal) impact parameter of a track is defined as the transverse (longitudinal) distance of closest approach of the track to the primary vertex.

The trajectories of charged particles are reconstructed through a iterative procedure which consists in multiple iterations of the *Combinatorial Track Finder algorithm* (CTF) [15], which uses the reconstructed hits in the silicon detectors to determine the track parameters. In

the first iterations the algorithm searches for tracks of relative large  $p_T$  and produced near the interaction region. Then, hits associated to high quality tracks are iteratively removed from the input list to reduce the combinatorial complexity of the next iterations and to allow the more difficult reconstruction of low  $p_T$  or displaced tracks. Each iteration of the CTF algorithm is made of four steps: track seeding, track finding and track fitting.

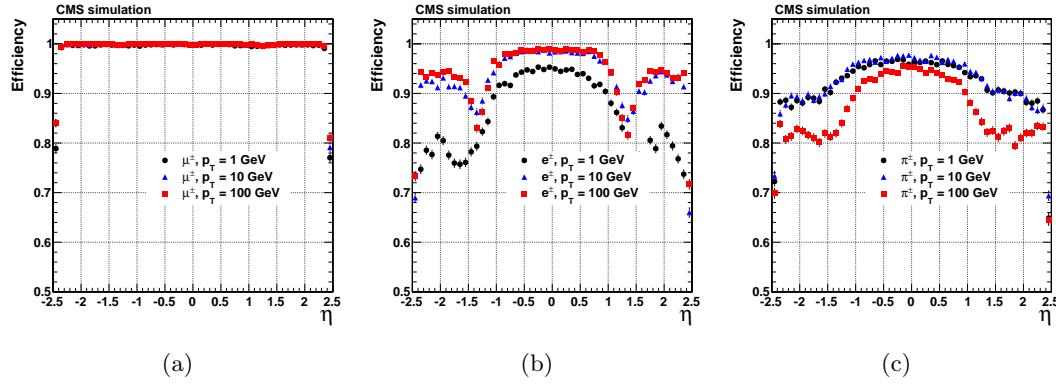
In the first step, a first estimate of the helix parameters and of its covariance matrix is provided using only pairs or triplets of hits compatible with the hypothesis of a track coming from the p-p interaction region. Track candidates are best seeded from hits in the pixel detector because of the low occupancy, high efficiency and unambiguous 3-dimensional position information.

The track finding stage associates new hits in the next tracker layers to the trajectory obtained from seeds using a standard Kalman Filter (KF) pattern recognition approach [16,17] which takes into account the effect of multiple scattering in the tracker layers. The current trajectory is extrapolated to the next tracker layer and compatible hits are assigned to the track on the basis of the  $\chi^2$  between the predicted and measured positions. In case multiple compatible hits are found when extrapolating the helix to a single layer, the algorithm creates one trajectory candidate for each hit and they are propagated independently. Furthermore, in order to take into account possible inefficiencies one additional candidate is created without including any hit information. The tracks are assigned a quality based on the  $\chi^2$ , the number of missing hits and how compatible they are with originating from a primary interaction vertex. Only the best quality tracks are kept for further propagation and ambiguities are resolved between tracks during and after track finding. In case two tracks share more than 50% of their hits, the lower quality track is discarded. The fake rate, defined as the fraction of reconstructed tracks not associated with a charged particle, is substantially reduced through these quality requirements.

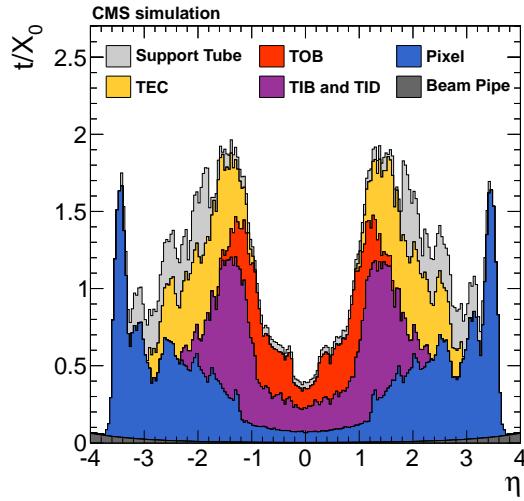
For each trajectory the finding stage results in an estimate of the track parameters. However, since the full information is only available at the last hit and constraints applied during trajectory building can bias the estimate of the track parameters, all valid tracks are refitted using the KF to determine the most accurate estimate of the helix parameters. The usual fit starting from the interaction point to the end of the tracker is complemented with a second fit run backward from the outermost tracker layer to the interaction point. This approach is found to improve the accuracy of the  $p_T$  and impact parameter measurement by 0.5% and 1% respectively.

The performance of the track reconstruction is shown in Fig. 6.1 for simulated muons, electrons and pions. For isolated muons with  $1 < p_T < 100$  GeV, the track reconstruction efficiency is  $> 99\%$  over the full  $\eta$ -range of tracker acceptance, and does not depend on  $p_T$  (Fig. 6.1(a)). The fake rate is completely negligible. For pions and electrons the efficiency is in general lower along with a higher fake rate because of interactions with the material in the tracker. The material budget of the CMS tracker in units of radiation length is presented in Fig. 6.2.

In Fig. 6.3(a) the transverse momentum resolution for muon tracks with  $p_T = 1, 10$  and 100 GeV is shown. At high transverse momentum (100 GeV), the resolution is 2–3% up to  $|\eta| = 1.6$ . The material of the tracker accounts for 20–30% of the transverse momentum resolution. At lower momenta, the resolution is dominated by multiple scattering and its distribution reflects the amount of material traversed by the track. The resolution of the track impact parameter in the transverse and longitudinal plane are also shown in Fig. 6.3. At high momentum the transverse impact parameter resolution is fairly constant and is dominated by the hit resolution in the first pixel layer. It is progressively degraded by multiple scattering at lower momenta. The same applies to the longitudinal impact parameter resolution. The



**Figure 6.1:** Track reconstruction efficiency for simulated muons (a), electrons (b), and pions (c) passing the high-purity quality requirements as a function of  $\eta$  and for  $p_T = 1, 10$ , and  $100$  GeV [14].



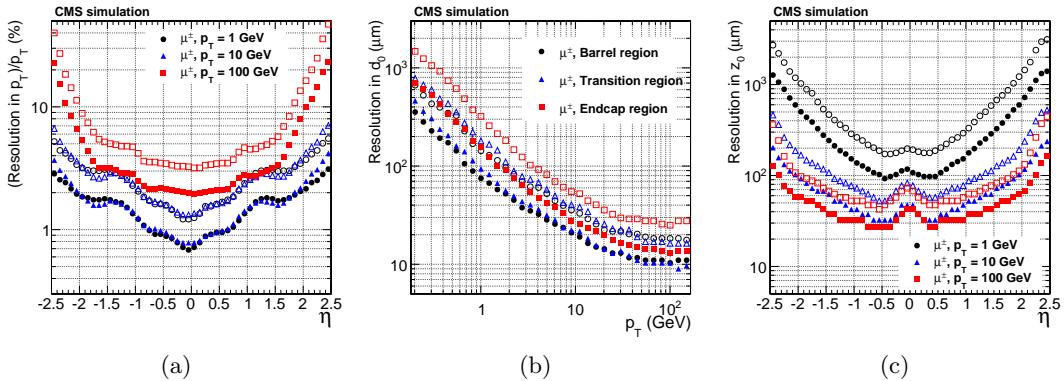
**Figure 6.2:** Material budget of the CMS tracker in units of radiation length  $X_0$  as a function of pseudorapidity divided into the contributions of the different subdetectors [14].

improvement of the  $z_0$  resolution up to  $|\eta| = 0.5$  is due to the charge sharing effects among neighboring pixels.

### 6.1.2 Vertex reconstruction

The identification of vertices is essential to distinguish the primary vertex associated with the hard interaction from additional pileup vertices that might be present in the event. This became even more important with the higher LHC luminosity reached at the end of 2016 where on average up to **FIXME: quote number** pp interactions took place simultaneously.

In the primary-vertex reconstruction [18], the measurements of the location and uncertainty of an interaction vertex are computed from a given set of reconstructed tracks. The prompt tracks originating from the primary interaction region are selected based on the transverse impact parameter significance with respect to the beam line, number of strip and pixel hits, and the normalized track  $\chi^2$  from a fit to the trajectory. The selected tracks are then clustered on the basis of their  $z$ -coordinates at their point of closest approach to the centre of the beam spot using a *deterministic annealing* (DA) algorithm [19]. This clustering allows



**Figure 6.3:** Resolution of track transverse momentum (a), transverse (b) and longitudinal (c) impact parameter for simulated muons passing the high-purity quality requirements as a function of  $\eta$  and for  $p_T = 1, 10, \text{ and } 100 \text{ GeV}$  [14].

for the reconstruction of any number of proton-proton interactions in the same LHC bunch crossing. Vertices are resolved with separations about 1 mm, appropriate for a multiplicity of interactions per bunch crossing up to 20, as the longitudinal RMS spread of the luminous region is about 6 cm.

After identifying candidate vertices based on the DA clustering in  $z$ , those candidates containing at least two tracks are then fitted using an *adaptive vertex fitter* [20] to compute the best estimate of vertex parameters, including its  $x$ ,  $y$ , and  $z$  position and covariance matrix. This algorithm addresses the issue of secondaries and fake tracks in the cluster by iteratively down-weighting the tracks which are not compatible with the common vertex being fitted. The primary vertex originating the hard scattering is chosen as the vertex with highest sum of  $p_T^2$  of the clustered tracks. The first vertex is usually the one corresponding to the collision of interest, where the hard process takes place.

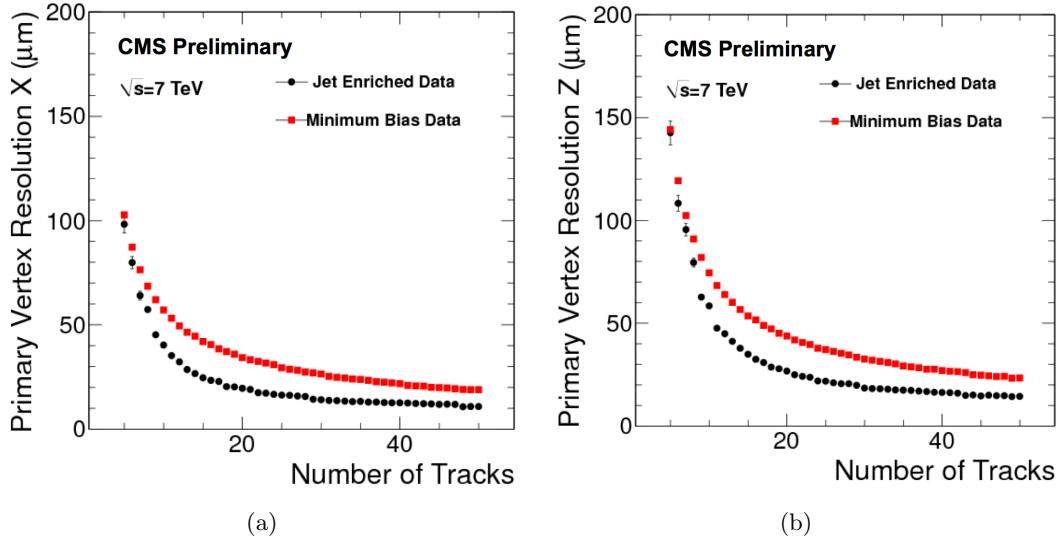
The primary vertex spatial resolution depends on the event topology and on the number of tracks related to the vertex, as shown in Fig. 6.4. For minimum-bias events, the resolutions in  $x$  and  $z$  are, respectively, less than  $20\mu\text{m}$  and  $25\mu\text{m}$ , for primary vertices reconstructed using at least 50 tracks. The resolution is better for the jet-enriched sample where tracks have significantly higher mean  $p_T$  resulting in higher resolution in the track impact parameter and consequently better vertex resolution. For these events, the resolutions approach  $10\mu\text{m}$  in  $x$  and  $12\mu\text{m}$  in  $z$  for primary vertices using at least 50 tracks.

## 6.2 Electrons

### 6.2.1 Electron reconstruction

The electron reconstruction in CMS [22] is based on the association of an energy deposit in the ECAL with a track reconstructed in the tracker detector. Electrons lose energy primarily through bremsstrahlung when interacting with the tracker layers and large energy losses are common. Given the non-Gaussian properties of the energy loss distributions, the standard track reconstruction algorithm based on the Kalman filter is not appropriate and lead in general to a reduced hit-collection efficiency, as well as to a poor estimation of track parameters. Better performance for electron reconstruction are achieved by using dedicated techniques that make use of information, not only from the tracker, but also from the ECAL.

The electron reconstruction starts by searching for clusters of energy in the ECAL. As the



**Figure 6.4:** Primary-vertex resolution in  $x$  (a) and  $z$  (b) as a function of the number of tracks at the fitted vertex, for two kinds of events with different average track  $p_T$  values; the results in  $y$  is almost identical to the one in  $x$  [21].

electrons are degraded in energy, the effect of the magnetic field is to enhance the bending of their trajectories, which ultimately results in a spread of irradiated photons along the  $\phi$  coordinate. To recover this radiated energy, ECAL superclusters are formed, by merging clusters of similar  $\eta$  over some range of  $\phi$ . Because of the different geometry of the detector in barrel and endcap, different clustering algorithms are used in different regions.

For the electron track reconstruction two approaches are used. In the first one, referred to as “ECAL driven”, the supercluster energy and position, and the assumption that the electron originated near the centre of the beam spot, are used to extrapolate the electron trajectory in the tracker. Tracker seeds compatible with the predicted trajectory is sought in the first or second layer of the pixel detector (and also in the TEC to improve efficiency in the forward region). This method is designed for isolated electrons with  $p_T > 5 \text{ GeV}$ .

A second approach, referred to as “tracker driven” is developed as part of the PF-reconstruction algorithm, and complements the electron track reconstruction, especially for low- $p_T$  or non isolated electrons, as well as for electrons in the barrel-endcap transition region. This method takes the standard track collection reconstructed with the KF algorithm and attempts to identify a subset of these tracks that are compatible with being electrons. Electrons that suffer only little bremsstrahlung loss can be identified by searching for tracks extrapolated to the ECAL that pass close to an ECAL PF cluster. Electrons that suffer large bremsstrahlung loss can be identified by the fact that the fitted track will often have poor  $\chi^2$  or few associated hits. The track seeds originally used to generate these electron-like tracks are retained.

The seed collections obtained by using these two methods are merged, and used to initiate electron track finding. This procedure is similar to that used in standard tracking, except that the  $\chi^2$  threshold, used by the KF to decide whether a hit is compatible with a trajectory, is weakened. This is to accommodate tracks that deviate from their expected trajectory because of bremsstrahlung.

To obtain the best parameter estimates, the final track fit is performed using a modified version of the KF, called the Gaussian Sum Filter (GSF) [23]. The fractional energy loss of an electron, as it traverses a layer of material, follows a Bethe–Heitler distribution. This

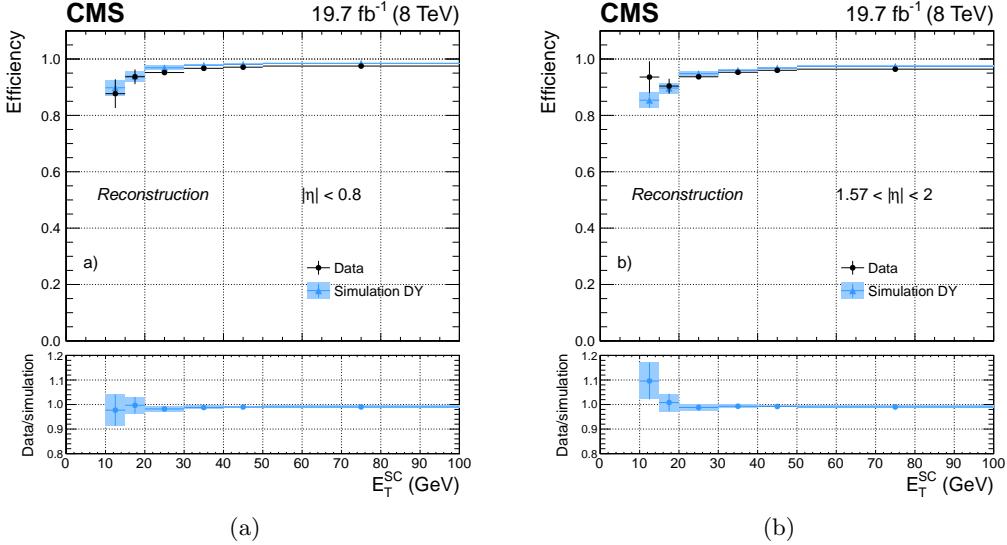
distribution is non-Gaussian, making it unsuitable for use in a conventional KF algorithm. The GSF technique solves this by approximating the Bethe–Heitler energy-loss distribution as the sum of several Gaussian functions. This method is then a generalization of the KF where the trajectory in each tracker layer is described by a weighted sum of KF components for which the energy loss follows a Gaussian law with a given width. The propagation of each component is done separately from one layer to another and the weights are then updated given the measurement in the new site. The allowed window to search for a hit in the next tracker layer is larger than for the usual KF track. This procedure is iterated until the last tracker layer, unless no hit is found in two subsequent layers. A minimum of five hits is finally required to create a track.

A GSF electron candidate is built by associating an ECAL supercluster with a GSF track with compatible  $\eta$  and  $\phi$  positions. The electron transverse energy  $E_T$  is equal to the transverse energy of the correspondent ECAL energy deposit (or supercluster) defined as  $E_T^{SC} = E \sin \theta$ , where  $\theta$  is the polar angle of the supercluster relative to the beam axis, and  $E$  the energy measured in the supercluster. The performance of the GSF electron reconstruction are studied using the “tag-and-probe” (T&P) method [24]. The method uses a known SM resonance mass and decay (e.g.  $Z \rightarrow e^+ e^-$ ) to select particles of the desired type, and probe the efficiency of a particular selection criterion on those particles. In general the “tag” is an object that passes a set of very tight selection criteria designed to isolate the required particle type (in this case an electron, though the method is not strictly limited to this case). A generic set of the desired particle type (i.e. with potentially very loose selection criteria) known as “probes” is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the resonance. Combinatorial backgrounds are usually eliminated through a variety of background subtraction methods. The definition of the probe object depends on the specifics of the selection criterion being examined. The efficiency itself is measured by counting the number of “probe” particles that pass the desired selection criteria. It is found that the estimated efficiencies are almost insensitive to any specific definition of the tag. The GSF electron candidate reconstruction efficiency measured with this method is above 95% for electrons in the ECAL barrel with  $E_T > 35$  GeV, as shown in Fig. 6.5(a). Slightly lower efficiencies are obtained for electrons reconstructed in the ECAL endcaps (Fig. 6.5(b)). A good agreement is found between data and simulation, resulting in scale factors consistent with unity almost in the entire range. The performance are presented here for the electron reconstruction in Run 1 but similar results are obtained in CMS for Run 2.

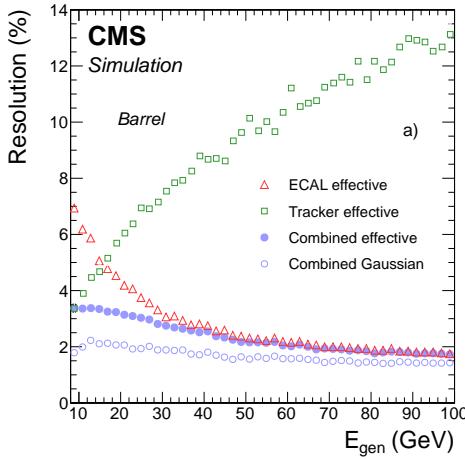
Once a GSF electron candidate is reconstructed, the energy measurement provided by electromagnetic calorimeter can be combined with the tracker momentum measurement to improve the estimate of electrons with energies below 35 GeV as shown in Fig. 6.6. At energies above 35 GeV however, the momentum measurement is completely driven by the supercluster.

### 6.2.2 Electron trigger

As explained in Sec. 3.2.4, the trigger system has two layers: the Level 1 (L1) and the High Level Trigger (HLT). At the L1, where the tracker information is not available, electrons and photons are indistinguishable and based on calorimeter trigger towers, consisting, in the barrel, of a  $5 \times 5$  matrix of ECAL crystals and the corresponding HCAL tower (a more complicated definition of the trigger tower is needed in the endcaps due to the different geometry). At this stage, the trigger choice is based only on the transverse energy of the calorimetric deposit, coarse shower shape requirements and the fraction of the total L1 seed energy belonging to the hadronic calorimeter, required to be smaller than 5%. Events passing



**Figure 6.5:** Electron reconstruction efficiency measured in dielectron events in data (dots) and Drell-Yan simulation (triangles), as a function of the  $E_{\text{T}}$  of the electron supercluster for electrons reconstructed in the ECAL barrel (a) and endcaps (b). The bottom panels show the corresponding data-to-simulation scale factors [25].



**Figure 6.6:** Expected resolution in  $p_{\text{T}}$  for isolated electrons in the ECAL barrel as a function of the electron generated energy, obtained from the ECAL, the tracker and the combined estimates [25].

L1 are then filtered by the HLT. Here, the pixel tracker information is used to separate electrons from photons. The starting point of any electron HLT selection consists of building a supercluster and a trajectory as described Section 6.2.1. Many different triggers involving electrons are designed at the HLT level and various additional requirements on the electrons are made for each of them. They consist of conditions on:

- the longitudinal and transverse shower shape of the supercluster energy deposits;
- the ECAL and HCAL activity around the candidate;
- the existence of a KF track (or, rarely, a GSF track when the rate of passing events is

low enough to allow the GSF algorithm be run) matching the supercluster position;

- the track activity around the candidate.

The conditions used and their severity depends on the number of electrons requested by the trigger and their transverse energy threshold, each trigger being designed to have a rate of accepting events of 50 Hz or less. Practically, all the HLT steps and criteria involving only calorimeters information are done first, the time consuming track reconstruction being only used at last for events passing the previous criteria. The L1 and HLT trigger used to collect the data analyzed in this thesis are listed in Tables 6.1 and Tables 6.2 for 8 and 13 TeV, respectively. The tables also detail the conditions imposed on several variables described in Section 6.2.3. Both the L1 and HLT triggers require one electron (or  $\gamma$ ) candidate. Figure 6.7 shows the L1 trigger efficiencies for different  $E_T$  thresholds as a function of the electron  $E_T$ . The curves exhibit the typical turn on behaviour in correspondence of the imposed  $E_T$  threshold.

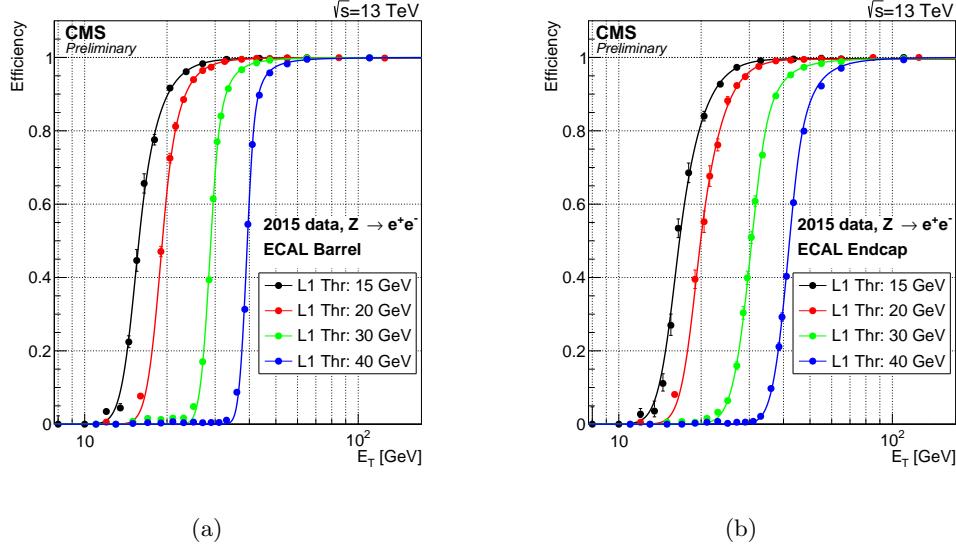
**Table 6.1:** The L1 and HLT single-electron triggers used to collect the 8 TeV data analyzed in this thesis together with the imposed requirements on the electron candidate.

Trigger	Name	Selections
Level 1	L1_SingleEG20	1 e/ $\gamma$ candidate $E_T > 20$ GeV
HLT	HLT_Ele80_CaloIdVT_GsfTrkIdT OR HLT_Ele80_CaloIdVT_GsfTrkIdT	1 electron: $E_T > 80$ GeV $ \Delta\eta_{in}  < 0.008$ $ \Delta\phi_{in}  < 0.07$ (barrel) or 0.05 (endcaps) $H/E < 0.05$ $\sigma_{i\eta i\eta} < 0.011$ (barrel) or 0.031 (endcaps)

**Table 6.2:** The L1 and HLT single-electron triggers used to collect the 13 TeV data analyzed in this thesis together with the imposed requirements on the electron candidate.

Trigger	Name	Selections
Level 1	L1_SingleEG35 OR L1_SingleEG40	1 e/ $\gamma$ candidate $E_T > 35$ GeV OR $E_T > 40$ GeV
HLT	HLT_Ele105_CaloIdVT_GsfTrkIdT OR HLT_Ele115_CaloIdVT_GsfTrkIdT	1 electron: $E_T > 105$ GeV OR $> 115$ GeV $ \Delta\eta_{in}  < 0.008$ $ \Delta\phi_{in}  < 0.07$ (barrel) or 0.05 (endcaps) $H/E < 0.05$ $\sigma_{i\eta i\eta} < 0.011$ (barrel) or 0.031 (endcaps)

The  $E_T$  thresholds required for the data collected in pp collisions at 13 TeV are higher compared to the one used in Run 1, in order to keep low trigger rates at the higher luminosities reached in Run 2. The chosen HLT triggers require a reconstructed GSF track. Furthermore, there are no requirements imposed on the electron candidate isolation. In general, this results in high fake rates of jets misreconstructed as electrons from multijet background, and, as a consequence, in high trigger rates which would require a prescale. However, the high  $E_T$  threshold allows for an unprescaled trigger, as jets from multijet background are characterized by low momentum. Furthermore, the kinematic region of the analyses presented in this thesis is located at very high lepton  $p_T$  and the signal efficiency is mainly affected at very low



**Figure 6.7:** L1 electron triggering efficiency in ECAL barrel (a) and endcaps (b) as a function of the offline reconstructed electron  $E_T$ . The efficiency is shown for the 15, 20, 30, 40 GeV EG trigger thresholds [26].

resonance masses ( $< 1 \text{ TeV}$ ) with a loss in efficiency of 20–25%.

The offline reconstructed  $p_T$  of the electrons must be greater than 90 (120) GeV for the 8 (13) TeV analysis, where the trigger reaches the plateau. In fact, this choice is made to simplify the analysis avoiding the need for modelling the trigger turn-on curve and, as a consequence, reducing the associated systematic uncertainties.

The efficiency for an electron passing the high- $E_T$  selections described in Sec. 6.2.3 to fire the HLT triggers of Tables 6.1 and 6.2 have been measured in data with T&P method and are found to be 98–99% for electrons with  $E_T$  in the trigger plateau, with data-to-simulation scale factors close to unity.

### 6.2.3 Electron identification

All the physics analyses in CMS involving one or two electrons in the final state start with the general electron reconstruction algorithm presented in Section 6.2.1. A high efficiency in any kinematical conditions is therefore needed and, as a consequence, the probability for other particles to be reconstructed as electrons is sizeable. For instance, a charged pion can mimic the signature of an electron if it interacts early and leaves most of its energy in the ECAL. Moreover electrons can emerge in a jet through the weakly decay of a hadron containing a c or b quark. Finally, in addition to jets, photons can also lead to GSF electron candidates. This happens if the photon converts into a dielectron pair in one of the first layers of the tracker detector. If one of the electron takes most of the photon momentum, a GSF electron candidate is likely to be reconstructed. An analysis dependent selection, which takes into account the specific kinematics and background level needs therefore to be applied on top of the electron reconstruction. This thesis focuses on the search for very heavy resonances decaying into dibosons where one of the bosons is a W decaying leptonically with a highly energetic electron or muon in the final state. A high and stable selection efficiency for  $E_T$  above 100 GeV is therefore the main requirement. Since this a common feature of most of the searches for new physics involving highly energetic electrons, a specific cut based selection has

been developed in CMS [27]. It consists of a series of cuts on several variables that exploit the specificity of high- $E_T$  electrons. Only GSF electron candidates with  $E_T > 35 \text{ GeV}$  and well reconstructed in the tracker and ECAL sensitive regions are selected. Candidates in the ECAL transition region ( $1.442 < |\eta_{SC}| < 1.56$ ) and beyond the  $\eta$  coverage ( $|\eta_{SC}| > 2.5$ ) of the tracker are therefore discarded. A different selection is applied for candidates reconstructed in the ECAL barrel ( $|\eta_{SC}| < 1.442$ ) and endcaps ( $1.56 < |\eta_{SC}| < 2.5$ ). For Run 2 the values of  $\eta_{SC}$  have been slightly adjusted to match the acceptance of the detector more accurately. The selections are summarized in Tables 6.3 and 6.4, for the 8 and 13 TeV analysis, respectively, and discussed in the following.

**Table 6.3:** List of the variables used in the high- $E_T$  selections for the 8 TeV analysis and the associated cut values for electrons reconstructed in the ECAL barrel and endcaps.

Variable	ECAL barrel	ECAL endcaps
ECAL-driven	yes	yes
$E_T$	$> 35 \text{ GeV}$	$> 35 \text{ GeV}$
$ \eta_{SC} $	$< 1.442$	$1.56\text{--}2.5$
$ \Delta\eta_{in} $	$< 0.005$	$< 0.007$
$ \Delta\phi_{in} $	$< 0.06$	$< 0.06$
Relative track isolation	5%	5%
Calorimeter isolation	$< 2 + 0.03E_T + 0.28\rho$	$< 2.5 + 0.28\rho \text{ if } E_T < 50$ $< 2.5 + 0.03(E_T - 50) + 0.28\rho \text{ if } E_T \geq 50$
Transverse shower shape	$E_{2\times 5}/E_{5\times 5} > 0.94$ OR $E_{1\times 5}/E_{5\times 5} > 0.83$	$\sigma_{i\eta i\eta} < 0.03$
$H/E$	$< 0.05$	$< 0.05$
$ d_{xy} $	$< 0.02$	$< 0.05$
Inner layer lost hits	$\leq 1$	$\leq 1$

**Table 6.4:** List of the variables used in the high- $E_T$  selections for the 13 TeV analysis and the associated cut values for electrons reconstructed in the ECAL barrel and endcaps.

Variable	ECAL barrel	ECAL endcaps
ECAL-driven	yes	yes
$E_T$	$> 35 \text{ GeV}$	$> 35 \text{ GeV}$
$ \eta_{SC} $	$< 1.4442$	$1.566\text{--}2.5$
$ \Delta\eta_{in} $	$< 0.004$	$< 0.006$
$ \Delta\phi_{in} $	$< 0.06$	$< 0.06$
Relative track isolation	5%	5%
Calorimeter isolation	$< 2 + 0.03E_T + 0.28\rho$	$< 2.5 + 0.28\rho \text{ if } E_T < 50$ $< 2.5 + 0.03(E_T - 50) + 0.28\rho \text{ if } E_T \geq 50$
Transverse shower shape	$E_{2\times 5}/E_{5\times 5} > 0.94$ OR $E_{1\times 5}/E_{5\times 5} > 0.83$	$\sigma_{i\eta i\eta} < 0.03$
$H/E$	$< 1/E + 0.05$	$< 5/E + 0.05$
$ d_{xy} $	$< 0.02$	$< 0.05$
Inner layer lost hits	$\leq 1$	$\leq 1$

As a starting point, electrons are selected if the reconstruction was seeded in the ECAL (ECAL driven, see Section 6.2.1). In fact, while useful for low energy non isolated electrons, particle flow is less suitable for high energy electrons.

The difference in  $\eta$ ,  $\Delta\eta_{in}$ , and in  $\phi$ ,  $\Delta\phi_{in}$ , between the track position as measured in the inner layers, extrapolated to the interaction vertex and then extrapolated to the calorimeter

and the position of the supercluster are required to be  $< 0.005$  and  $< 0.06$ , respectively. In fact, for jets, the position of the center of the ECAL deposit can be far from the track position, as all of the constituents can leave an energy deposit in the ECAL. The  $\Delta\phi_{in}$  distribution is however much broader than  $\Delta\eta_{in}$ , because of the wider spread of the energy in  $\phi$  due to photons from bremsstrahlung, resulting in a looser requirement. The distributions of  $\Delta\phi_{in}$  and  $\Delta\eta_{in}$  become narrower with increasing  $E_T$  and a higher background signal discrimination can be achieved with a tighter requirement at high  $E_T$  compared to the usual selections for low or intermediate energetic electrons. The reason of this behaviour comes from the fact that bremsstrahlung photons are more collinear to the electron at higher  $E_T$ . The definition of  $\Delta\eta_{in}$  has been changed for Run 2 to use instead the  $\eta$  of the seed cluster of the supercluster which is found to provide a more accurate indication of the  $\eta$  of the original electron before bremsstrahlung.

To suppress the misidentification of jets as electrons, the sum of the  $p_T$  of all other tracks in a cone of  $\Delta R < 0.3$  around the track of the electron candidate is required to be less than 5 GeV, which imposes an isolation condition on the track. To be used in the calculation of the isolation of the candidate track, the tracks have to be within 0.2 cm, in the  $z$  direction, of the primary vertex with which the electron candidates are associated. This requirement reduces the impact of pileup and it does not show a dependency with the electron  $E_T$  for high values. For electrons with transverse energies above 100 GeV, a negligible change in the selection efficiency is observed as the number of pileup events increases from 0 to 40. For electrons with  $E_T$  much lower than 100 GeV, the efficiency falls by between 5 and 10% depending on the region of the detector in which the electrons are detected.

The calorimeter isolation variable used in these selection is defined as the sum of:

- ECAL isolation: sum of the  $E_T$  of the energy deposits in the ECAL calorimeter in a cone of  $\Delta R < 0.3$  around the track of the electron candidate excluding those associated with the candidate;
- HCAL1 isolation: sum of the  $E_T$  of the energy deposits in the first layer of the HCAL calorimeter in a cone of  $\Delta R < 0.3$  around the track of the electron candidate excluding those associated with the candidate.

The calorimeter isolation so defined, is required to be less than 3% (plus a small  $\eta$ -dependent offset) of the candidate  $E_T$ . This sum, which allows a selection on the isolation of the electron candidate, is corrected for the average energy density in the event,  $\rho$ , to minimize the dependence of the efficiency of this selection criterion on pileup. This requirement differs from the selection usually applied for electrons of low or intermediate  $E_T$ . For these cases, a PF based isolation is generally used, which merges the information of the tracker, the ECAL and the HCAL allowing to measure the contribution to the isolation from charged hadrons, neutral hadrons and photons separately. One of the main advantage of this procedure is that the energy deposit in the calorimeters associated to a charged hadron produced in another interaction, characterized by a different primary vertex, can be removed from the isolation sum. For very high energy ( $> 1$  TeV) electrons, however, the PF sometimes fails to recognize an electron from a GSF electron candidate and assigns all its energy deposit to the photon isolation. Furthermore, the combined PF isolation is generally requested to be below 15% of the electron  $E_T$ . This requirement is suitable for electrons with  $E_T$  in the low or intermediate range ( $E_T < 50\text{--}100$  GeV) but it results in a very loose selection for electrons with much higher  $E_T$  (150 GeV for  $E_T = 1$  TeV).

Further suppression of the misidentification of jets as electrons is achieved by requiring that the ratio  $H/E$  of the energy in the HCAL towers in a cone of  $\Delta R < 0.15$  centered

on the electron candidate position to the electromagnetic energy of the electron candidate supercluster is required to be less than 5%. For Run 2, the selection on this variable has been increased to match the new HCAL conditions (**FIXME: Not sure this is true**). Additionally, the profile of the energy deposition in the ECAL is required to be consistent with that expected for an electron. In particular, the transverse shower shape can be quantified by the following variables:

- $E_{1\times 5}/E_{5\times 5}$ : ratio of the energy contained in the  $1\times 5$  domino in  $\eta \times \phi$  in the barrel ( $x \times y$  in the endcaps) centered on the seed crystal of the supercluster over the energy of the  $5 \times 5$  matrix centered on the seed crystal;
- $E_{2\times 5}/E_{5\times 5}$ : ratio of the energy contained in the most energetic  $2\times 5$  domino in  $\eta \times \phi$  in the barrel ( $x \times y$  in the endcaps) centered on the seed crystal of the supercluster over the energy of the  $5 \times 5$  matrix centered on the seed crystal;
- $\sigma_{inj\eta}$  : measure of the spread in  $\eta$  in units of crystals of the electrons energy in the  $5 \times 5$  block centred on the seed crystal.

In the barrel, the best performance are obtained using a cut based on both  $E_{1\times 5}/E_{5\times 5}$  and  $E_{2\times 5}/E_{5\times 5}$ . The two variables are indeed complementary: while  $E_{1\times 5}/E_{5\times 5}$  is well designed for electrons hitting the center of a crystal,  $E_{2\times 5}/E_{5\times 5}$  allows to recover electrons that hit the crystal close to its edge. Combining the two variables instead of using just one of them allows to set strong cut values on both and thus well reject background while keeping a high efficiency on simulated electrons. The distributions of these variables are much broader for signal electrons in the endcaps and a better discrimination power is obtained applying a selection on the variable  $\sigma_{inj\eta}$ .

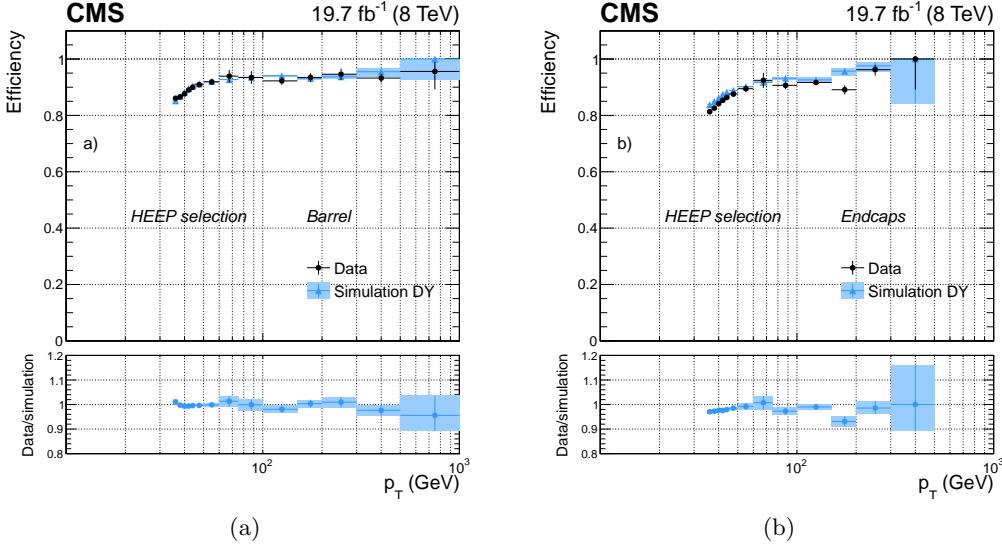
Two additional requirements are applied to reject photons that convert into a electron-positron pair in the tracker. First, the track associated with the cluster is required to have no more than one hit missing in the pixel layers. The signature left by the photon conversion process is very similar to the one from real electrons and the gain in discrimination using shower shape variables is limited. However, one of the main differences is the absence of hits in the first layers of the tracker, before the conversion happens. Furthermore, the transverse impact parameter  $d_{xy}$ , defined as the closest distance, in the transverse plane, between the primary vertex and the track of the electron candidate, is required to be  $< 0.02$  cm (barrel) or  $0.05$  cm (endcaps). The distribution of the transverse impact parameter is usually wider in the endcaps due to the poorer resolution on the track position in that region.

The efficiency of the high- $E_T$  electron selection measured with the T&P method in 8 TeV collision data and simulation as a function of the electron  $p_T$  is shown in Fig. 6.8, for electrons reconstructed in the ECAL barrel and endcaps. Similar results are obtained using 13 TeV data. The efficiencies and data-to-simulation scale factors are summarized in Tables 6.5 and 6.6 for the 8 and 13 TeV analysis, respectively. The scale factors are close to unity, indicating a good agreement between data and simulation. They are used in the analyses presented in this thesis to correct the normalization of simulations.

## 6.3 Muons

### 6.3.1 Muon reconstruction

The CMS detector is specifically designed for the optimization of muon detection, as its name clearly states. In general, muons will not be absorbed by the calorimeters, as it happens with



**Figure 6.8:** Efficiency of the high- $E_T$  electron selection as a function of electron  $p_T$  for dielectron events in 8 TeV data (dots) and DY simulation (triangles) in the ECAL barrel (a), and endcaps (b) [25].

**Table 6.5:** Efficiencies and scale factors for the high- $E_T$  electron selection in the 8 TeV analysis for electrons with  $E_T > 90$  GeV.

	ECAL barrel	ECAL endcaps
Efficiency simulation	$90.2\% \pm 0.2\%$ (stat.)	$92.2\% \pm 0.5\%$ (stat.)
Efficiency data	$88.7\% \pm 0.2\%$ (stat.)	$90.7\% \pm 0.6\%$ (stat.)
Data/simulation scale factor	$0.983 \pm 0.004$ (stat.)	$0.984 \pm 0.010$ (stat.)

**Table 6.6:** Efficiencies and scale factors for the high- $E_T$  electron selection in the 13 TeV analysis for electrons with  $E_T > 120$  GeV.

	ECAL barrel	ECAL endcaps
Efficiency simulation	$91.4\% \pm 0.10\%$ (stat.)	$84.4\% \pm 0.3\%$ (stat.)
Efficiency data	$91.6\% \pm 0.04\%$ (stat.)	$82.3\% \pm 0.1\%$ (stat.)
Data/simulation scale factor	$1.002 \pm 0.001$ (stat.)	$0.975 \pm 0.004$ (stat.)

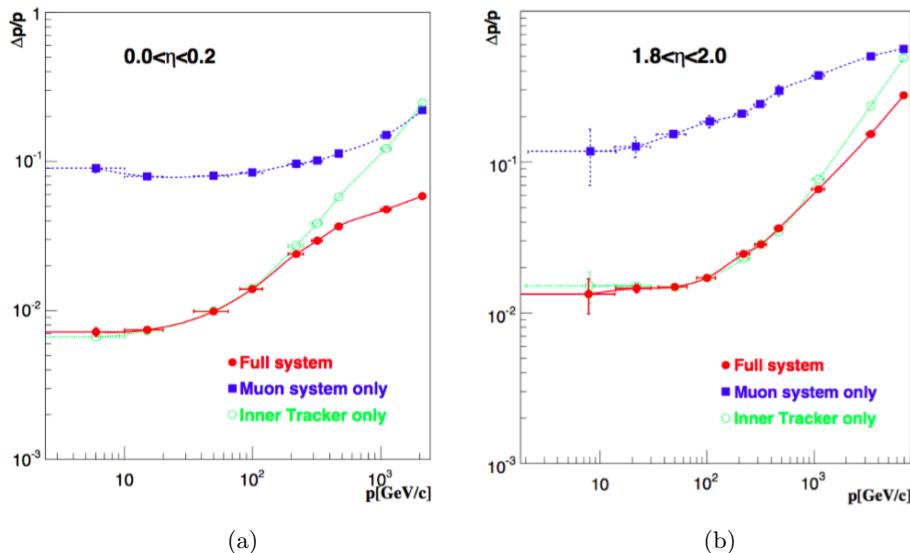
electrons, so a specific muon detection system (see Sec. 3.2.3) is needed in order to identify and correctly measure its momentum. In the standard CMS reconstruction [28], tracks are first reconstructed independently in the inner tracker (tracker track) and in the muon system (standalone-muon track). A standalone-muon track is reconstructed from pre-built track segments (i.e. a set of aligned DT or CSC hits) in the muon chambers. The state vector associated to the segments found in the innermost chambers are used to seed the muon trajectory, working from inside out, using the KF technique: the predicted state vector at the next measurement surface is compared with existing hits and updated accordingly. A suitable  $\chi^2$  cut is applied to reject bad hits and the procedure is iterated until the outermost surface of the muon system is reached. Finally the track is extrapolated to the nominal interaction point and a vertex-constrained fit is performed. The magnetic field, the multiple scattering inside the steel yoke, and the energy loses are taken into account.

Based on reconstructed standalone-muon and tracker tracks, two reconstruction approaches

are then used:

- **Global-muon reconstruction (outside-in).** Each standalone-muon track is extrapolated to the tracker and a search is performed in a cone around it to match a tracker track. A global-muon track is fitted combining hits from the tracker track and standalone-muon track, using the KF technique.
  - **Tracker-muon reconstruction (inside-out).** In this approach, all tracker tracks with  $p_T > 0.5$  GeV are considered as possible muon candidates and are extrapolated to the muon system while searching for a match with at least one muon segment.

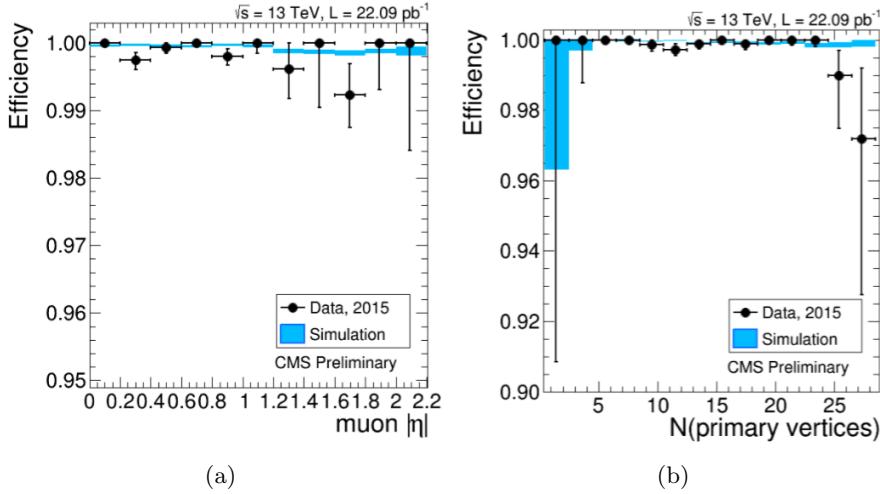
Tracker-muon reconstruction is more efficient than the global-muon reconstruction at low momenta,  $p_T \leq 5 \text{ GeV}$ , because it requires only a single muon segment in the muon system, whereas global-muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station and typically requires segments in at least two muon stations. However, given the high efficiency of both the tracker-track and muon segments reconstruction, about 99% of muons produced within the geometrical acceptance of the muon system and having sufficiently high momentum ( $p_T \geq 5 \text{ GeV}$ ) are reconstructed by both methods. As shown in Fig. 6.9 the additional information provided by the muon system is precious for the momentum reconstruction of high-energy muons ( $p_T \geq 200 \text{ GeV}$ ), for which the tracker-only momentum measurement degrades. In fact, as a particle's momentum increases and the curvature of its corresponding track decreases, the momentum resolution in the tracker becomes limited by position measurement resolution. One can then benefit from the large lever arm and 3.8 T magnetic field in the region between the tracker and the muon system by including hits in the muon chambers. For lower momenta, instead, the resolution of the tracking system is dominating.



**Figure 6.9:** Relative resolution of the muon momentum measurement for the reconstruction with the inner tracker only, the muon system only and for the combination of the inner tracker and the muon system, for simulated muons emitted in the central (a) and forward (b) regions [29].

Figure 6.10 shows the muon tracking efficiency as a function of the  $\eta$  of the probe muon and the number of primary vertices for data (2015) and simulation, evaluated using the T&P method described in Section 6.2.1. In the region  $|\eta| < 2.2$  and for events with number of

reconstructed primary vertices lower than 25 (**FIXME: point here to a plot with nPV for 2015**), the measured tracking efficiency of isolated muons is  $> 99\%$  in both data and simulation. The tracking efficiency is constant as a function of the number of vertices per event which means that does not depend on the pileup.



**Figure 6.10:** Tracking efficiency measured with a T&P technique, for muons from Z decays, as a function of the muon  $\eta$  (a) and the number of primary vertices (b), for 2015 data (black dots) and simulation (blue bands) [30].

The combination of different algorithms provides robust and efficient muon reconstruction. After the completion of both algorithms, the reconstructed stand-alone, global, and tracker muons are merged into a single software object, with the addition of further information, like isolation and energy collected in matching calorimeter towers. This information can be used for further identification, in order to achieve a balance between efficiency and purity of the muon sample as described in Section 6.3.3.

The muon reconstruction performance at high momentum are strongly affected by radiative processes and by the muon detector alignment. As the muon traverses the steel layers of the magnet return yoke, electromagnetic showers can develop, producing additional segments in the muon chambers, or resulting in large energy losses. These events can corrupt the measurement done in the muon detectors. Therefore, specialized reconstruction algorithms for high- $p_T$  muons, known as “TeV-muon” refits, have been developed in CMS as described in the following.

The *tracker-plus-first-muon-station* fit (TPFMS) only uses hits from the tracker and the innermost muon station with hits, to reduce the sensitivity to possible showering deeper in the muon system. The *Picky* fit uses all tracker hits, while a selection is applied to muon hits. Hits from chambers with a high probability of shower contamination (determined from the hit occupancy) are required to be compatible with the extrapolated trajectory by applying a  $\chi^2$  cut. The *dynamic truncation* algorithm (DYT) starts from the idea that the muon track reconstruction should be stopped after a large energy loss, as hits produced after that can only bias the momentum measurement. For every global muon trajectory the algorithm starts from the corresponding tracker track and propagates it out to the muon stations. Compatible segments (or hits) in the muon chambers are found by using an estimator which takes into account the propagation of the tracker covariance matrix through the material and the magnetic field, and the covariance matrices of the candidate muon segments (or hits).

Momentum assignment is then performed by the *Cocktail* algorithm which combines the above methods to further improve the resolution at high  $p_T$  reducing the tails of the momentum resolution distribution. In particular, the algorithm chooses, on track-by-track basis, the best muon reconstruction. For the 8 TeV CMS analyses, the Cocktail-algorithm decision is taken between the tracker-only, TPFMS, and Picky fits. This version of the algorithm is also known as the *Tune P* algorithm. It starts with the Picky fit, then switches to the tracker-only fit if the goodness of fit ( $\chi^2/\text{n.d.f.}$ ) of the latter is significantly better. Then it compares the  $\chi^2/\text{n.d.f.}$  of the chosen track with that of TPFMS; TPFMS is chosen if it is found to be better. For high- $p_T$  muons, TPFMS and Picky algorithms are selected by Tune P in most of the cases, in approximately equal amounts, while the tracker-only fit is selected only in a few percent of events.

For 13 TeV analyses, the Tune P algorithm was extended to include also the DYT fit. The selection is still made on a track-by-track basis, but using both the  $\chi^2/\text{n.d.f.}$  of the track and the relative error of the  $p_T$  measurement. The algorithm starts with the Picky fit, then switches to DYT if the DYT track has a lower relative  $p_T$  error. It then compares the  $\chi^2/\text{n.d.f.}$  of the chosen track with that of the tracker-only fit and picks tracker-only if its  $\chi^2/\text{n.d.f.}$  is significantly better. Then the  $\chi^2/\text{n.d.f.}$  of the chosen track and TPFMS are compared and the one giving the best result is kept. At the end, if the final candidate track has  $p_T$  lower than 200 GeV or the tracker-only  $p_T$  is lower than 200 GeV, the tracker-only track is selected.

The momentum resolution obtained with the Tune P algorithm for muons with  $p_T$  in the range  $350 < p_T < 2000$  GeV is found to be  $\sim 6\%$ , as measured with cosmic-ray muon data [28, 31].

### 6.3.2 Muon trigger

The Level-1 muon trigger uses signals from all three CMS muon detector systems: DT, CSC, and RPC. It has a latency of  $3.2 \mu\text{s}$  and reduces the rate of inclusive muon candidate events read-out from detector front-end electronics to a few kHz by applying selections on the estimated muon  $p_T$  and quality. In the muon HLT, first a Level-1 trigger object is used as a seed to reconstruct a standalone-muon track in the muon system, leading to an improved  $p_T$  estimate. At this point,  $p_T$  threshold filters are applied to the standalone-muon (also called Level-2 muon). Then seeds in the inner tracker are generated in the region around the extrapolated Level-2 muon, and tracker tracks are reconstructed. If a successful match is made between a tracker track and the Level-2 muon, a global fit combining tracker and muon hits is performed, yielding a Level-3 muon track on which the final  $p_T$  requirements are applied. In this way, the rate of recorded inclusive muon events is reduced to a few tens of Hz. The average processing time of the HLT reconstruction is about 50 ms.

The L1 and HLT trigger used to collect the data analyzed in this thesis are listed in Tables 6.7 and 6.8 for the 8 and 13 TeV analysis, respectively. For both analyses the HLT used to select the events is the unprescaled single-muon trigger with the lowest  $p_T$  threshold that does not include muon isolation requirements. In fact, although muons produced by the leptonic decay of a high- $p_T$  W boson tend to be isolated, their high momentum enhances the production of electromagnetic showers, that can mimic a non-isolated muon candidate. Therefore, only requirements on the muon  $p_T$  and  $\eta$  are applied at this stage. The efficiency of the L1 single-muon trigger with the 16 GeV threshold is shown in Fig. 6.11 as a function of the offline reconstructed muon  $p_T$  and  $\eta$ . In 2012 the efficiency of the L1 single-muon trigger was greater than 90%.

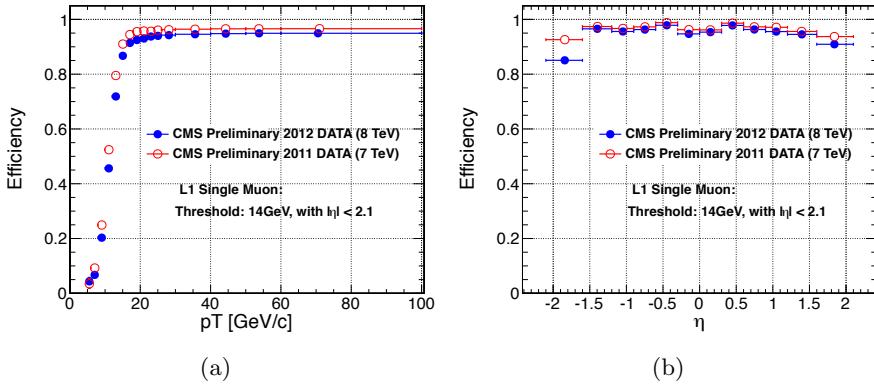
The offline reconstructed  $p_T$  of the muons must be greater than 50 (53) GeV for the 8 (13) TeV analysis, where the trigger reaches the plateau. Furthermore the selected muon must be

**Table 6.7:** The L1 and HLT single-muon triggers used to collect the 8 TeV data analyzed in this thesis together with the imposed requirements on the muon candidate.

Trigger	Name	Selections
Level 1	L1_SingleMu16_eta2p1	1 global muon with: $p_T > 16 \text{ GeV}$ $ \eta  < 2.1$
HLT	HLT_Mu40_eta2p1	1 global muon with: $p_T > 40 \text{ GeV}$ $ \eta  < 2.1$

**Table 6.8:** The L1 and HLT single-muon triggers used to collect the 13 TeV data analyzed in this thesis together with the imposed requirements on the muon candidate.

Trigger	Name	Selections
Level 1	L1_SingleMu25	1 global muon with: $p_T > 25 \text{ GeV}$
HLT	HLT_Mu45_eta2p1	1 global muon with: $p_T > 45 \text{ GeV}$ $ \eta  < 2.1$



**Figure 6.11:** Efficiency of the L1 single-muon trigger with a 14 GeV threshold as a function of the muon  $p_T$  (a) and  $\eta$  (b) [32].

reconstructed within  $|\eta| < 2.1$  as a consequence of the trigger criteria. The efficiency for a muon passing the high- $p_T$  selections described in Section 6.3.3 to fire the HLT single-muon triggers have been measured in data with T&P method and are summarized in Tables 6.9 and 6.10.

**Table 6.9:** Efficiencies and scale factors for the single-muon HLT trigger used in the 8 TeV analysis for muons with  $p_T > 50 \text{ GeV}$ ,  $|\eta| < 2.1$ , and satisfying the high- $p_T$  and isolation selections described in Section 6.3.3.

	$0 <  \eta  < 0.9$	$0.9 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
Efficiency simulation	$95.10\% \pm 0.03\%$	$87.01\% \pm 0.03\%$	$81.56\% \pm 0.03\%$
Efficiency data	$92.90\% \pm 0.02\%$	$83.14\% \pm 0.06\%$	$80.27\% \pm 0.05\%$
Data/simulation scale factor	$0.9768 \pm 0.0004$	$0.956 \pm 0.001$	$0.984 \pm 0.001$

**Table 6.10:** Efficiencies and scale factors for the single-muon HLT trigger used in the 13 TeV analysis for muons with  $p_T > 53 \text{ GeV}$ ,  $|\eta| < 2.1$ , and satisfying the high- $p_T$  and isolation selections described in Section 6.3.3.

	$0 <  \eta  < 0.9$	$0.9 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
Efficiency simulation	$97.6\% \pm 0.1\%$	$93.4\% \pm 0.4\%$	$94.8\% \pm 0.2\%$
Efficiency data	$94.6\% \pm 0.2\%$	$89.7\% \pm 0.4\%$	$91.8\% \pm 0.2\%$
Data/simulation scale factor	$0.969 \pm 0.002$	$0.961 \pm 0.006$	$0.968 \pm 0.003$

### 6.3.3 Muon identification

The standard CMS muon reconstruction provides additional information for each muon, useful for muon quality selection and identification in physics analyses [28]. In general, particles detected as muons are produced in pp collision from different sources which lead to different experimental signatures. The so-called *prompt muons* arise either from decays of W, Z, and promptly produced quarkonia states, or other sources such as Drell-Yan processes or top quark production. Real muons are also produced in the decay of heavy flavour particles, such as beauty or charmed mesons, as well as in light hadron (pions or kaons) decays. Less frequently, muons might be originated from a calorimeter shower or a product of a nuclear interaction in the detector. Furthermore, the so called “punch-through” effect, i.e. hadron shower remnants penetrating through the calorimeters and reaching the muon system, can lead to the reconstruction of a muon candidate. Most of the physics analyses in CMS searching for BSM signals use prompt muons, while all the other categories constitute the background.

All physics analysis in CMS involving muons in the final state exploit the same set of information, while applying selections on the muon variables which are slightly different depending on the interesting signature and the expected background. In this section only the specific selection developed for high- $p_T$  muons are described. One of the main difference with respect to the low- and medium- $p_T$  muon selection is that this particular identification working point does not use the PF algorithm. It is aimed to the best reconstruction of the muon track parameters for high- $p_T$  muons without relying on external information on the event. Moreover, the goodness of the global-muon track fit selection, based on the  $\chi^2$  of the track, is not requested, but an additional selection based on the relative  $p_T$  resolution for the track used for momentum determination is applied.

The high- $p_T$  muon selection criteria are described in the following and are the same for the 8 and 13 TeV analyses:

- The muon must be reconstructed both as a tracker- and a global-muon. This is effective against decays-in-flight, punch-through and accidental matching (with noisy or background tracks or segments).
- Number of pixel hits in the tracker track  $\geq 1$ . To further suppress muons from decays in flight.
- Number of tracker layers involved in the measurements  $\geq 6$ . This guarantees a good  $p_T$  measurement, for which some minimal number of measurement points in the tracker is needed. It also suppresses muons from decays in flight.
- Number of muon-chamber hits included in the global-muon track fit  $\geq 1$ . This requirement assures that the global muon is not an accidental match between the information from the muon system and the tracker. This could happen in particular for non-prompt muons or fake muons from punch-through.

- The muon track is required to have muon segments in at least 2 muon stations. To further suppress punch-through and accidental track-to-segment matches. This selection is furthermore consistent with the logic of the single-muon trigger, which requires segments in at least two muon stations to obtain a meaningful estimate of the muon  $p_T$ .
- The transverse impact parameter of the muon track has to be compatible with the interaction point hypothesis. A maximum value of 2 mm is required. This is effective against cosmic background and further suppress muons from decays in flight.
- Also the longitudinal impact parameter is used to further suppress cosmic muons, muons from decays in flight and tracks from pile-up. A maximum value of 5 mm is required.
- The relative  $p_T$  error is required to be smaller than 30%, to suppress grossly misreconstructed muons.

In addition to these identification criteria, an isolation requirement is applied to the well-identified muons. In particular, the muon must pass a relative tracker-only isolation selection: the scalar sum of the  $p_T$  of all other tracks in a cone of  $\Delta R < 0.3$  around but not including the muon tracker track must be less than 10% of the muon  $p_T$ , also as measured by the tracker. To be used in the calculation of the tracker-based isolation, tracks have to be within 2 mm, in the  $z$  direction, of the primary vertex with which the muon candidate is associated. The impact parameter criterion also reduces the effect of tracks originating from pileup on reconstructed quantities.

The efficiency and data-to-simulation scale factors of the high- $p_T$  and isolation muon selection measured with the T&P method in 8 and 13 TeV collision data are summarized, respectively, in Tables 6.11 and 6.12. The scale factors are close to unity, indicating a good agreement between data and simulation. They are used in the analyses presented in this thesis to correct the normalization of simulations.

**Table 6.11:** Efficiencies and scale factors for the high- $p_T$  and isolation muon selection used in the 8 TeV analysis for muons with  $p_T > 50$  GeV and  $|\eta| < 2.1$ .

Muon $ \eta $	$0 <  \eta  < 0.9$	$0.9 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
High- $p_T$ muon identification			
Efficiency simulation	$96.51\% \pm 0.02\%$	$96.61\% \pm 0.04\%$	$95.54\% \pm 0.03\%$
Efficiency data	$95.54\% \pm 0.02\%$	$95.87\% \pm 0.04\%$	$95.06\% \pm 0.03\%$
Data/simulation scale factor	$0.9900 \pm 0.0003$	$0.992 \pm 0.001$	$0.9949 \pm 0.0004$
Tracker-based muon isolation			
Efficiency simulation	$99.49\% \pm 0.01\%$	$99.58\% \pm 0.01\%$	$99.59\% \pm 0.01\%$
Efficiency data	$99.46\% \pm 0.01\%$	$99.51\% \pm 0.01\%$	$99.56\% \pm 0.01\%$
Data/simulation scale factor	$0.9996 \pm 0.0001$	$0.9994 \pm 0.0001$	$0.9997 \pm 0.0001$

## 6.4 Jets

Particles carrying a color charge, such as quarks, cannot exist in free form because of QCD confinement which only allows for colorless states (**FIXME: point here to theory chapter**). Quarks and gluons interact with pairs of quarks and anti-quarks produced from the vacuum until the formation of stable colourless hadrons, the process is called hadronization. The ensemble of the final colourless objects is called a jet and it is reconstructed in the detector from energy depositions and charged particle momenta. The jets point back to the primary

**Table 6.12:** Efficiencies and scale factors for the high- $p_{\text{T}}$  and isolation muon selection used in the 13 TeV analysis for muons with  $p_{\text{T}} > 53 \text{ GeV}$  and  $|\eta| < 2.1$ .

Muon $ \eta $	$0 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
High- $p_{\text{T}}$ muon identification		
Efficiency simulation	$97.6\% \pm 0.2\%$	$99.81\% \pm 0.2\%$
Efficiency data	$96.7\% \pm 0.4\%$	$1.0\% \pm 0.7\%$
Data/simulation scale factor	$0.991 \pm 0.005$	$1.002 \pm 0.007$
Tracker-based muon isolation		
Efficiency simulation	$99.8\% \pm 0.1\%$	$99.6\% \pm 0.1\%$
Efficiency data	$99.7\% \pm 0.1\%$	$99.7\% \pm 0.1\%$
Data/simulation scale factor	$0.999 \pm 0.001$	$1.001 \pm 0.001$

interaction, i.e. to the partons the jets originated from, but a correction for hadronization and detector effects is needed. Jet clustering algorithms have been developed to cluster particles (at parton, particle or detector level) into jets and reconstruct the energy and direction of the original parton. The task of a jet clustering algorithm is to allow comparisons between theoretical predictions, which are usually described by perturbative calculations, and experimental data. This is achieved reducing the complex structure of particle jets from a scattered parton to a simple four-momentum, which represents the main property of particle jets. In order to guarantee a meaningful calculation of theory predictions, jet clustering algorithms are characterized by two important properties. Clustering algorithms need to be infrared-safe, which means that the emission of infinitesimally-low-energy partons from partons inside a jet does not affect the jet properties. Furthermore, they need to be collinear-safe, which means that jet properties are not affected by the splitting of a parton inside a jet into two collinear partons. Jet algorithms for hadron colliders can be divided into two classes: cone [33] and sequential clustering [34–38] algorithms. The main algorithms used by LHC experiments are the anti- $k_t$  algorithm [38] (AK) and the Cambridge–Aachen (CA) [34, 36] algorithms, which are found to fulfill theory requirements and to exhibit good properties for experimental measurements. For this work both algorithms are used and described in the following.

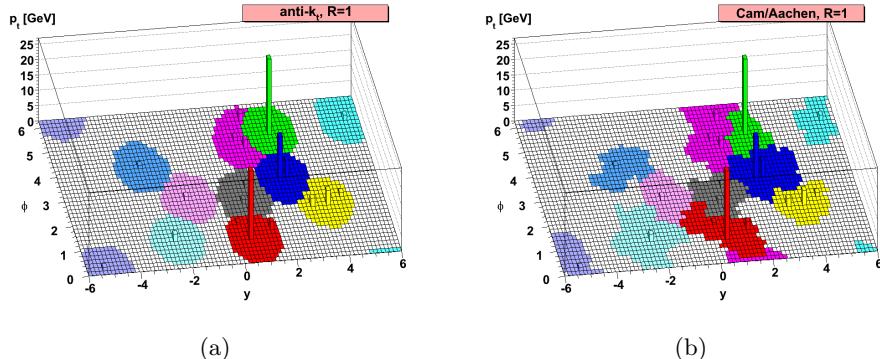
#### 6.4.1 Jet clustering algorithms

In sequential jet clustering algorithms, jets are defined through sequential, iterative procedures that combine four-vectors of input pairs of particles until certain criteria are satisfied and jets are formed. In particular, for each pair of particles  $i$  and  $j$ , a distance variable between the two particles ( $d_{ij}$ ), and the so-called “beam distance” for each particle ( $d_{iB}$ ), are computed:

$$d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \frac{\Delta R^2 ij}{R^2} \quad , \quad d_{iB} = p_{Ti}^{2n} \quad , \quad (6.1)$$

where  $p_{Ti}$  and  $p_{Tj}$  are the transverse momenta of particles  $i$  and  $j$ , respectively, “min” refers to the smaller of the two  $p_{\text{T}}$  values, the integer  $n$  depends on the specific jet algorithm,  $\Delta R^2 ij$  is the distance between  $i$  and  $j$  in  $\eta$  and  $\phi$ , and  $R$  is a free distance parameter, with all angles expressed in radians. The particle pair  $(i, j)$  with smallest  $d_{ij}$  is combined into a single object. All distances are recalculated using the new object, and the procedure is repeated until, for a given object  $i$ , all the  $d_{ij}$  are greater than  $d_{iB}$ . Object  $i$  is then classified as a jet and not considered further in the algorithm. The process is repeated until all input particles are clustered into jets.

The distance parameter  $R$  is responsible for defining the angular size of the jet. The parameter  $n$  governs the topological properties of the jets and depending on its value three different classes of clustering algorithms are distinguished. For  $n = 1$  the procedure is referred to as the  $k_t$  algorithm (KT), which clusters soft objects before harder ones are added to the final jet. The KT jets tend to have irregular shapes and are especially useful for reconstructing jets of lower momentum [38]. For this reason, they are also sensitive to the presence of low- $p_T$  pileup contributions. For  $n = 0$ , the procedure corresponds to the CA algorithm. This relies only on angular information, and, like the KT algorithm, provides irregularly-shaped jets. The CA algorithm is useful in identifying jet substructure as described in Chapter 7. For  $n = -1$ , the procedure corresponds to the AK algorithm, which compares the inverse square of the transverse momenta. The AK algorithm is used extensively in LHC experiments and by the theoretical community for finding well-separated jets. The use of inverse square of the  $p_T$  as a weight in the  $d_{ij}$  distances has the advantage that hard objects collect adjacent soft ones before these are clustered among themselves into harder objects, figuratively reproducing in reverse the parton fragmentation and gluon emission processes. This property makes the algorithm independent on soft radiation preserving infrared-safety. The AK algorithm is also collinear-safe as the clustering is driven by the angular distance between two particles. Gluons emitted at small angles are picked up by the algorithm in early steps of the iteration and therefore do not affect the jet properties. Furthermore, this algorithm tends, by construction, to form almost circular jets allowing for straight-forward calibration and understanding of the detector acceptance. The behaviours of the CA and AK jet algorithms are illustrated in Fig. 6.12.

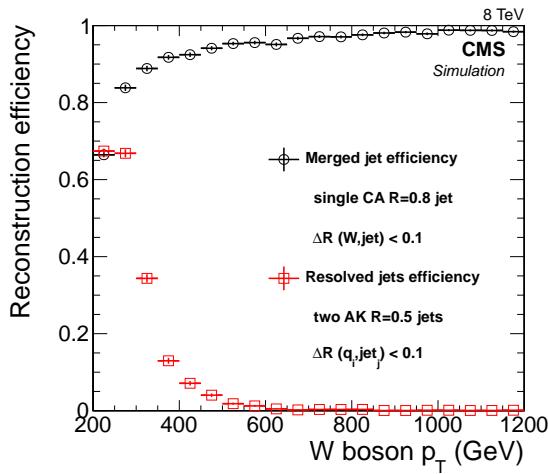


**Figure 6.12:** An example of jet clustering with the AK (a) and CA (b) algorithms. The reconstructed jets are shown as colored regions [38].

The choice of the distance parameters  $R$ , generally depends on the analysis. While large cone size jets collect all energy from the scattered parton, they also pick up a large contribution of background energy from the underlying event or pileup interactions. Small cone size jets pick up little contamination, but may not collect all energy from the scattered parton. The default choice in CMS for physics analyses at 8 and 13 TeV uses the KT algorithm with  $R = 0.5$  (AK5) and  $R = 0.4$  (AK4), respectively, since more collimated jets are expected at higher  $\sqrt{s}$ . However, a larger value of  $R$  increases the efficiency to entirely reconstruct the highly energetic products from boosted  $V \rightarrow q\bar{q}^{(\prime)}$  and  $H \rightarrow b\bar{b}$  decays. In fact, the average angular distance between the decay products is inversely proportional to the  $p_T$  of the mother particle. The default choice in CMS for physics analyses involving boosted  $V$  or  $H$  bosons decaying hadronically is  $R = 0.8$ . In particular, CA8 and AK8 jets are used in the 8 and 13 TeV analyses, respectively. The chosen value of  $R$  provides a high efficiency for  $V$  or  $H$

bosons with small boost and ensures that no efficiency is lost in the transition from classical V or H reconstruction in two small jets at low V or H  $p_T$  and reconstruction from a single large-cone jet at higher V or H  $p_T$ . Another point to consider when choosing the value of  $R$ , is the  $t\bar{t}$  data sample available for validating highly boosted V/H jets (see Section 7.2). If  $R$  is chosen too large, the b quark from the  $t \rightarrow Wb$  decay tends to merge into the W jet. The chosen value of  $R$  is the result of a compromise between high efficiency for V/H bosons with small boost and a sufficiently large sample of W jets in  $t\bar{t}$  data for validating the V or H jet identification algorithms. Figure 6.13 shows the  $p_T$  range of W bosons for which the CA8 algorithm is efficient and compares this to the efficiency for reconstructing W bosons from two AK5 jets. Above a  $p_T$  of 200 GeV, the CA8 jet algorithm, used to identify W jets, becomes more efficient than the reconstruction of a W boson from two AK5 jets.

The AK5 or AK4 algorithms are used in this analysis to put requirements on additional b jets in the event selection (see Chapter 8), along with the b tagging algorithm described in Section 6.4.3.

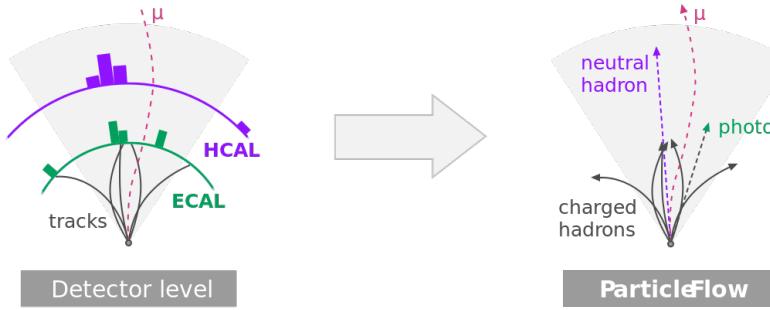


**Figure 6.13:** Efficiency to reconstruct a CA8 jet within  $\Delta R < 0.1$  of a generated W boson, and the efficiency to reconstruct two AK5 jets within  $\Delta R < 0.1$  of the generated quarks W bosons, as a function of the  $p_T$  of the W boson [39].

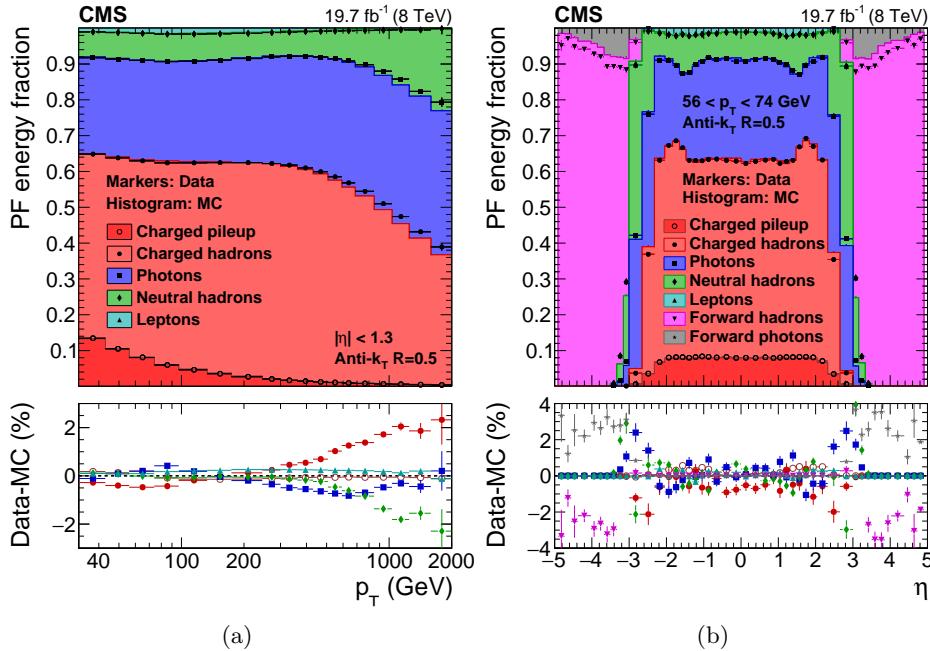
#### 6.4.2 Jet reconstruction and calibration

In CMS several standard methods for jet reconstruction are available which make use of different detector components, e.g. the tracker and the calorimeters, and give different reconstructed objects as input to the above explained jet clustering algorithms. In this work, only jet reconstructed with the particle-flow (PF) algorithm [40, 41] are used and referred to as “PFJets”. As sketched in Fig. 6.14, the PF algorithm aims at reconstructing all the stable particles produced in an event, combining the information coming from all CMS sub-detectors to optimize particle identification, direction and energy determination. These particles are classified in several types: charged hadrons, photons, neutral hadrons, electrons and muons. As shown in Fig. 6.15, jets are typically composed by 65% charged hadrons, 25% photons, 10% neutral hadrons. The PF algorithm is optimized to identify all these different components inside the jet, contrary to a calorimetric-only reconstruction. Typically, photons correspond to ECAL deposits not compatible with a tracker track. Charged hadrons correspond to HCAL and/or ECAL deposits matched to a inner track and not compatible with an electron, whereas neutral hadrons are identified as HCAL deposits not matched

to any track. While the momentum of neutral particles is measured in the calorimeters, the momentum of charged particles is measured by the tracker with a better resolution. In this way both the position and energy measurements are greatly improved with respect to calorimeter jets as this algorithm makes use of the tracking detectors and high granularity of the ECAL which is much higher than that of the HCAL. Once all the PF candidates in the event are reconstructed, they are used as input to the jet clustering algorithms and a PF jet is formed.



**Figure 6.14:** Sketch of the CMS particle-flow algorithm.

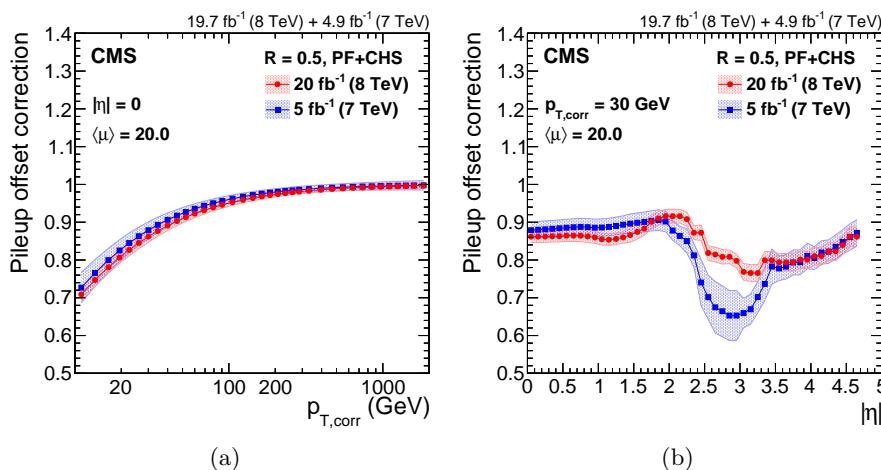


**Figure 6.15:** PF jet composition in data and simulation versus  $p_T$  for jets with  $|\eta| < 1.3$  (a), and versus  $\eta$  at  $56 < p_T < 74$  GeV (b) [42].

The jet momentum is determined as the vectorial sum of all PF candidates in the formed jet giving its “raw” estimate. At this stage, the reconstructed jet energy has a great uncertainty due to the several intrinsic limitations of the system, such as the non-linear response of the calorimeters, the detector segmentation, the presence of material in front of calorimeters, electronic noise, pileup. The raw jet energy and resolution are thus corrected for several factors in order to obtain the energy value as close as possible to the true energy of the initial

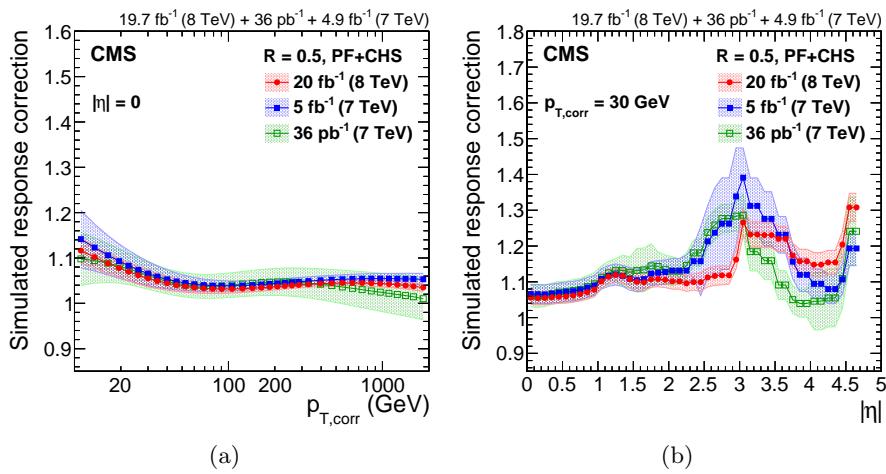
parton. CMS has adopted a factorized approach [43] to the problem of jet energy corrections, where each level of correction takes care of a different effects as described in the following.

The first step in this approach is a correction to the jet energies to mitigate pileup effects. The additional pp collisions occurring within the same bunch-crossing as the primary hard interaction produce additional tracks in the tracker and deposit energy in the calorimeters. This contribution is usually referred to as in-time pileup. Due to the finite signal decay time in the calorimeters, the pp collisions occurring in the previous and subsequent beam crossings also contribute to calorimetric energy in the same time window as the primary hard interaction. This contribution is called out-of-time pileup. The out-of-time contribution is mitigated at the level of signal processing, while the in-time one is partially removed using tracking information. This is achieved identifying which vertex the charged PF candidates originate from, and removing those unambiguously associated with pileup vertices before clustering jets and  $\vec{p}_T^{\text{miss}}$  (see Section 6.5). This method is referred to as *charged-hadron subtraction* (CHS), and represents the reference bench-mark for jet reconstruction in Run 1 and early Run 2 analyses. For the second part of Run 2, other pileup mitigation techniques in addition to CHS have been developed and tested in CMS [44–46] but they are not used in the analyses presented in this work. The CHS jets are then corrected to subtract residual contributions from neutral pileup particles, overlapping inside the jet cone. These corrections are determined from the simulation of a sample of QCD dijet events processed with and without pileup contaminations. This correction is usually parametrized as a function of the pileup energy density ( $\rho$ ) [47, 48], the jet area ( $A$ ) [49], jet  $p_T$  and  $\eta$ . The pileup offset corrections, defined as the mean value of the difference between the  $p_T$  of the reconstructed jet in events with and without pileup contamination, for the AK5 CHS jets as a function of the corrected jet  $p_T$  and  $\eta$  are shown in Fig. 6.16, estimated for typical 2012 (8 TeV) conditions with an average number of additional pileup interactions  $\langle \mu \rangle = 20$ . The typical offset correction for a AK5 jet without CHS is 0.75 for a corrected jet  $p_T$  of 30 GeV, while a correction of 0.85 is obtained for AK5 CHS jets with same  $p_T$  value. This indicates that CHS removes approximately half of this offset before jet clustering by matching tracks to pileup vertices, reducing the residual offset correction. Roughly one third of the remaining pileup is from PF charged hadrons that have not been matched to good pileup vertices, and much of the rest is from PF photons.



**Figure 6.16:** Pileup offset correction for AK5 CHS jets, with systematic uncertainty band, estimated for the typical 2012 condition of  $\langle \mu \rangle = 20$ . Corrections are shown for jets at  $|\eta| = 0$  versus corrected jet  $p_T$  (a), and for jets with  $p_T = 30 \text{ GeV}$  versus  $|\eta|$  (b) [42].

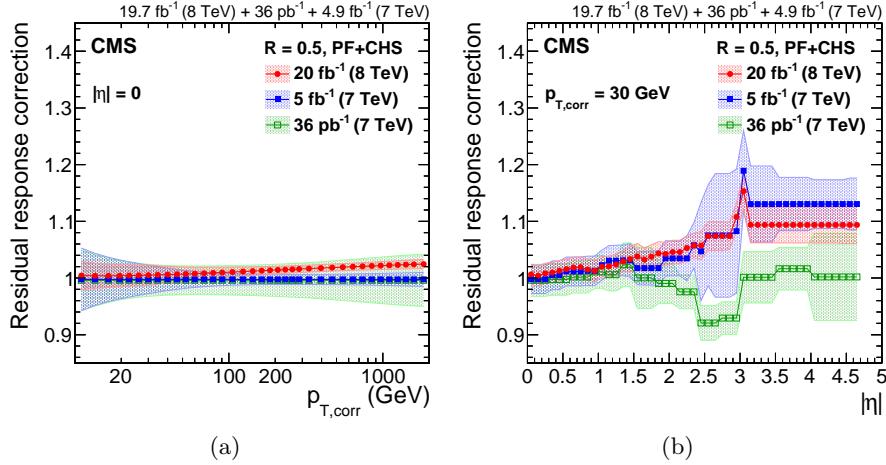
Secondly, a simulation driven jet energy response correction is applied. The detector simulation takes into account effects of particles deflected by the magnetic field, energy lost when traversing the detector material, particle conversions, and a detailed detector geometry. In this step the aim is to correct for non-uniformities in the different CMS sub-detectors, by comparing the reconstructed jet  $p_T$  to the particle-level one using simulated events only. The corrections are derived as a function of jet  $p_T$  and  $\eta$  and make the response uniform over these two variables. The simulated particle response corrections are summarized in Fig. 6.17 for data collected at 8 TeV. The response is quite flat at  $p_T > 50$  GeV, where the competing effects of increasing calorimeter response and falling tracking efficiency within the jet core compensate each other. In the barrel and endcap regions, the corrections rise with  $|\eta|$ , due to the increasing amount of material located in front of the calorimeters, which leads to effects such as an increased rate of nuclear interactions in the tracker. The corrections are higher around  $|\eta| = 1.3$  and 3.0 due to the degradation of the response in the transition regions.



**Figure 6.17:** Response correction factors with their systematic uncertainty band from simulation for the 2012 data collected at 8 TeV for AK5 CHS jets. Corrections are shown for jets at  $|\eta| = 0$  versus corrected jet  $p_T$  (a), and for jets with  $p_T = 30$  GeV versus  $|\eta|$  (b) [42].

Finally data-driven residual corrections are applied to correct for any measurable difference between the detector simulation and the jets measured in data. This correction is done in two steps. At first, an additional correction for the non homogeneous response of the detector with  $\eta$  is derived from dijet events, in which the  $p_T$  response of a probe jet, outside the barrel region, is balanced to the one in the reference tag region ( $|\eta| < 1.3$ ) as a function of the average  $p_T$  of the dijet system. Only events with back-to-back dijets and little additional activity in the event are used, to avoid any impact from unbalanced events. The jet energy is calibrated as a function of transverse momentum using a combination of  $Z(\rightarrow \ell\ell) + \text{jet}$ ,  $\gamma + \text{jet}$ , and multijet events for jets in the reference barrel region ( $|\eta| < 1.3$ ). The basic idea, in all the considered topologies, is to exploit the transverse momentum balance between the jet to be calibrated and a well reconstructed and calibrated reference object ( $Z$  or  $\gamma$ ). The jet energy response is studied using two approaches. In one method the jet response is evaluated by comparing the reconstructed jet momentum ( $p_{T,\text{jet}}$ ) directly to the momentum of the reference object ( $p_{T,\text{ref}}$ ), while the second, more advanced, method takes into account the  $E_{\text{Tmiss}}$  measured in the calorimeters to balance the reference object and jet momenta. In this method the additional event activity is taken into account by the  $E_{\text{Tmiss}}$ . Therefore, additional jets in the event have only a small impact on the measurement. The residual response corrections are summarized in Fig. 6.18 for data collected at 8 TeV. The residual response

corrections are less than 3% in the barrel, less than 10% in the endcaps, and about 10% in the forward detector.



**Figure 6.18:** Residual data/simulation response correction factors for the 2012 data collected at 8 TeV for AK5 CHS jets. Corrections are shown for jets at  $|\eta| = 0$  versus corrected jet  $p_T$  (a), and for jets with  $p_T = 30$  GeV versus  $|\eta|$  (b) [42].

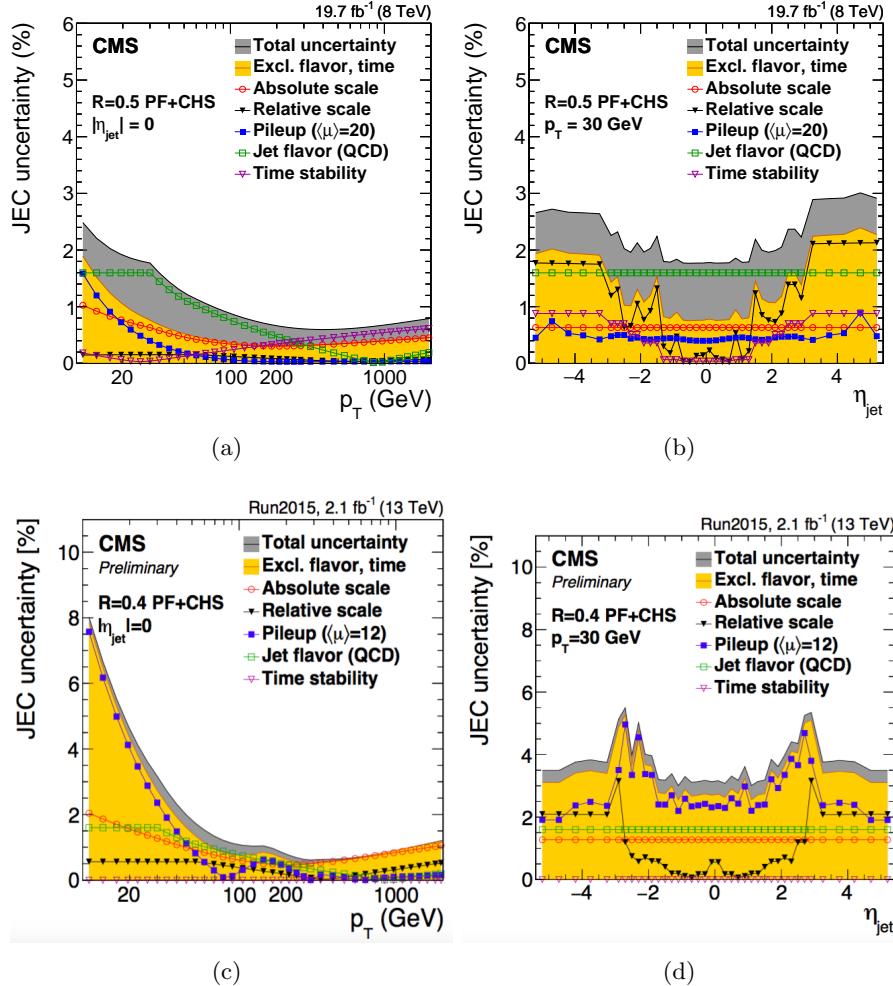
The fully calibrated PF jets are finally obtained in both data and simulation by multiplying all the above correction factors to the raw jet  $p_T$  as follows:

$$p_{T,\text{corr}} = p_{T,\text{raw}} \times C_{\text{pu}}(p_{T,\text{raw}}, \eta, \rho, A) \times C_{\text{sim}}(C_{\text{pu}} \cdot p_{T,\text{raw}}, \eta) \times C_{\text{res}}(C_{\text{pu}} \cdot C_{\text{sim}} \cdot p_{T,\text{raw}}, \eta) \quad (6.2)$$

where  $C_{\text{pu}}$  represents the pileup correction,  $C_{\text{sim}}$  is the simulated response correction and  $C_{\text{res}}$  is the global residual correction applied only on jets in data. Figure 6.19 shows the overall uncertainty on the corrections to the jet energy scale for AK5 and AK4 CHS jets for data collected at 8 and 13 TeV, respectively. In both cases, the final uncertainties are below 3% across the phase space of this analysis.

The energy resolution of jets is relatively poor compared to the resolution of other physics objects (electrons, muons, photons), and the biases caused by jet resolution smearing is important for steeply falling spectra and for resonance decays. Hence, calibrations are evaluated to correct the jet energy resolution in addition to the corrections to the jet energy scale described above. The measurements are performed with methods which are extensions of the methods used for measuring jet energy scales, but instead of looking at the mean of the response distribution, the width is in the interesting parameter. Furthermore, corrections have to compensate for effects that do not produce an overall shift in the mean, but that can widen the distribution. As shown in Fig. 6.20, the jet energy resolution in data is worse than in the simulation by 10–20% depending on  $\eta$ , and the jets in simulation need to be smeared accordingly.

Jets used in this analysis are also requested to pass loose identification criteria, in order to reject fake jets due to calorimeter noise, with more than 99% efficiency for true jets. Furthermore, the AK8 (or CA8) and AK5 (or AK4) jets are required to be separated from any well-identified muon or electron (see Sections 6.3 and 6.2) by  $\Delta R > 0.8$  and  $> 0.3$ , respectively. All AK8 (or CA8) and AK5 (or AK4) jets must have corrected  $p_T > 30$  GeV and  $> 200$  GeV, respectively, and  $|\eta| < 2.4$  to be considered in the subsequent steps of the analysis.

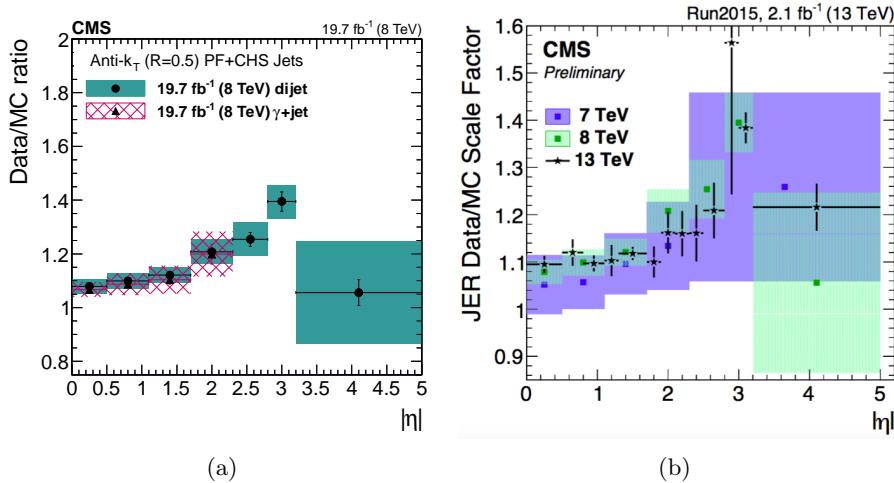


**Figure 6.19:** Summary of jet energy scale systematic uncertainties for the 2012 data collected at 8 TeV for AK5 CHS jets (upper plots) and for the 2015 data collected at 13 TeV for AK4 CHS jets (lower plots). Uncertainties are shown for jets at  $|\eta| = 0$  versus corrected jet  $p_T$  (left), and for jets with  $p_T = 30$  GeV versus  $|\eta|$  (right) [42, 50].

### 6.4.3 Identification of b jets

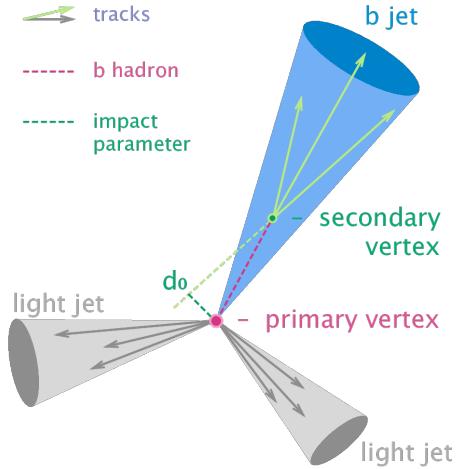
The identification of jets originating from b quarks (“b jets”) is one of the key ingredients of the analysis described in this work, which aims at isolating events of new physics with Higgs bosons decaying to  $b\bar{b}$ . The ability to identify b jets (“b tagging”) plays a crucial role in reducing background coming from processes involving jets from gluons and light-flavor quarks (u, d, s), and from c quark fragmentation.

Identifying b jets relies on the properties of the production and the weak decay of b hadrons. The most important property is the relatively long lifetime of b hadrons of about 1.5 ps ( $c\tau \equiv 450\mu m$ ) corresponding to a flight distance that is observable with high resolution tracking detectors. A b hadron with  $p_T = 50$  GeV covers, on average, almost half a centimetre ( $Lc \sim \gamma\tau$ ) before decaying. As shown in Fig. 6.21, this leads to secondary vertices displaced from the primary event vertex and charged particle tracks incompatible with the primary vertex with sizeable impact parameter. In addition, b hadrons have a large mass and large multiplicity of charged particles in the final state (about five charged particles on average per b hadron decay). Because of the hard b-fragmentation function, the b hadron in a b jet



**Figure 6.20:** Data/MC scale factors for the jet  $p_T$  resolution as a function of  $|\eta|$  determined from 2012 data collected at 8 TeV for AK5 CHS jets (a) and from 2015 data collected at 13 TeV for AK4 CHS jets (b) [42, 50].

carries a large fraction of the jet energy. Since b and c hadrons may decay semileptonically, in about 20% (per lepton species) of the cases an electron or muon is produced inside a b jet, if both direct and cascade decays are taken into account.



**Figure 6.21:** Representation of a b hadron decay and reconstructed b jet in the transverse plane.

A variety of algorithms has been developed in CMS [51] to identify b jets based on the b hadron properties described above. These algorithms use lower-level physics objects, mainly jets and charged tracks. Only the tracking detectors offer the spatial resolution needed to measure the properties of b hadron decays such as their significant flight path. Efficient track reconstruction, and in particular precise spatial reconstruction close to the interaction point, is thus the key ingredient. Some of these algorithms use just a single observable, while others combine several of these objects to achieve a higher discrimination power. Each of these algorithms yields a single discriminator value for each jet. The minimum thresholds on these discriminators define loose (“L”), medium (“M”), and tight (“T”) operating points

with a misidentification probability for light-flavor jets of 10%, 1%, and 0.1%, respectively, at an average jet  $p_T$  of about 80 GeV.

The jets used for b tagging are reconstructed with the PF algorithm and calibrated as described in 6.4.2. A sample of well-reconstructed tracks of high purity is required as input to each of the b tagging methods. Specific requirements are imposed in addition to the selection applied in the iterative tracking procedure described in Section 6.1.1:

- the fraction of misreconstructed or poorly reconstructed tracks is reduced by requiring a  $p_T > 1$  GeV;
- at least 8 tracker hits (including pixel) must be associated with the track;
- at least 2 hits are required in the pixel system since track measurements in the innermost layers provide most of the discriminating power;
- the normalised  $\chi^2$  is required to be  $< 5$  to ensure a good-quality fit;
- the absolute value of the transverse and longitudinal impact parameter of the track must be  $< 0.2$  and  $< 17$  cm, respectively, to reject charged particle tracks having their origin from sources showing much larger displacement from the primary vertex (e.g. photon conversions and nuclear interactions in the beam pipe or the first layers of the pixel detector);
- tracks are associated to jets in a cone  $\Delta R < 0.3$  around the jet axis, where the jet axis is defined by the primary vertex and the direction of the jet momentum **FIXME: check this for Run 2**;
- in order to reject tracks from pileup the distance to jet axis, defined as the distance of closest approach of the track to the axis, is required to be  $< 700\mu\text{m}$ ;
- the point of closest approach between the track trajectory and the jet axis, must be within 5 cm of the primary vertex.

Two algorithms for reconstructing secondary vertices are exploited. For the first algorithm, the tracks associated to jets and fulfilling the above selection requirements are used in the adaptive vertex reconstruction (AVR) algorithm [52] based on the adaptive vertex fitter described in Section 6.1.1. This is the secondary vertex reconstruction algorithm used for b tagging methods in CMS during LHC Run 1. A number of selection criteria are applied to remove vertices that are less likely to originate from a b hadron decay.

- at least 2 tracks must be associated to the secondary vertex;
- the fraction of tracks shared with the primary vertex is required to be  $< 65\%$ ;
- the distance between the primary vertex to the secondary vertex in the transverse plane, the 2D flight distance, must be in the range 0.1–25 mm;
- the 2D flight distance divided by its uncertainty or so-called 2D flight distance significance has to be  $> 3$ ;
- the invariant mass of charged particles associated to the vertex is required to be  $< 6.5$  GeV and not compatible with the mass of the  $K_S^0$  hadron in a window of 50 MeV;
- the angular distance  $\Delta R$  between the jet axis and the secondary vertex flight direction is required to be less than the jet distance parameter;

In contrast with the AVR algorithm, the Inclusive Vertex Finder (IVF) [53], is not seeded from tracks associated to the reconstructed jets. The IVF algorithm uses as input the collection of reconstructed tracks in the event and looser quality criteria are applied. The selected tracks are then used to identify clusters of nearby tracks based on their minimum distance and the angles between them. The clusters are fitted with the adaptive vertex fitter and a cleaning procedure is applied. At this stage, tracks can appear in multiple vertices and therefore, one of the vertices is removed based on the number of shared tracks and distance between the vertex and another one. Furthermore, tracks in the secondary vertex compatible with the primary vertex are removed. When there are at least 2 tracks associated to the secondary vertex after the track arbitration, the vertex is refitted and selection criteria similar to the case of the AVR vertices are applied.

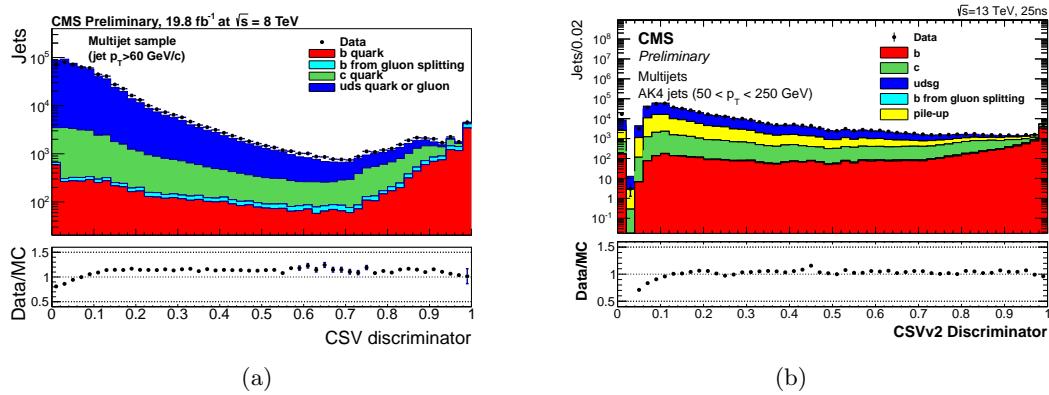
The efficiency to reconstruct a secondary vertex for b (c) jets using the IVF algorithm is about 10% (15%) higher compared to the efficiency to reconstruct a secondary vertex with the AVR algorithm. However, for light-flavour jets the probability to find a secondary vertex also increases by about 8%. Independently of the jet flavour, around 60% of the jets with an AVR vertex also have an IVF vertex.

In this analysis the *Combined Secondary Vertex* (CSV) algorithm is used, which combines the information of displaced tracks with the information of secondary vertices associated to the jet. This allows the algorithm to avoid limitations due to inefficiencies in the secondary vertex reconstruction. Jets are divided in three vertex-dependent exclusive categories: the presence of a reconstructed secondary vertex; at least two tracks with impact parameter significance larger than 2; none of the previous. The following set of variables with high discriminating power and low correlations are considered:

- the secondary vertex category;
- the 2D flight distance significance of the secondary vertex;
- the number of tracks in the jet
- the number of tracks associated to the secondary vertex;
- the secondary vertex mass;
- the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- the  $\eta$  of the tracks at the vertex with respect to the jet axis;
- the 2D impact parameter significance of the first track that raises the invariant mass above the charm threshold of 1.5 GeV when subsequently summing up tracks ordered by decreasing impact parameter significance;
- the 3D signed impact parameter significance for each track in the jet.

Two likelihood ratios are built from these variables used to discriminate between b and c jets and between b and light-flavor jets and combined with prior weights of 0.25 and 0.75, respectively. Figure 6.22(a) shows the distribution of the CSV discriminator value in a multijet sample for data collected at 8 TeV and for simulation, for jets reconstructed with the AK5 algorithm.

The CSV algorithm was further optimized for LHC Run 2 and the new version is referred to as CSV version 2 (CSVv2) [55]. The main differences with the Run 1 version of the CSV algorithm are the different vertex reconstruction algorithm used, the number of input



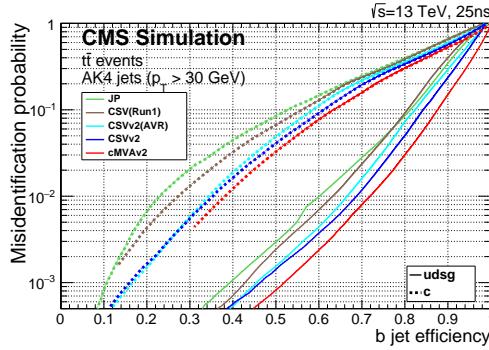
**Figure 6.22:** (a) Distribution of the CSV discriminator value in a multijet sample for data collected at 8 TeV and for simulation [54], for jets reconstructed with the AK5 algorithm. (b) Distribution of the CSVv2 discriminator value in a multijet sample for data collected at 13 TeV and for simulation, for jets reconstructed with the AK4 algorithm [55].

variables and the way those are combined. In this version the input variables are combined using a multivariate technique. In fact, the likelihood-based method used in Run 1 limited the amount of input variables since correlation between those could not be taken into account properly. In addition, the secondary vertex information is obtained with the IVF method described above. Figure 6.22(b) shows the distribution of the CSVv2 discriminator value in a multijet sample for data collected at 13 TeV and for simulation, for jets reconstructed with the AK4 algorithm.

The performance of the CSVv2 tagger is presented in Fig. 6.23 as the b jet identification efficiency versus the misidentification probability for jets in simulated  $t\bar{t}$  events requiring a jet  $p_T > 30$  GeV. A comparison is shown with the Run 1 version of the CSV algorithm, without retraining this algorithm. The CSV algorithm was trained on 7 TeV pp collisions using AK5 jets. The absolute improvement of the CSVv2 (AVR) algorithm with respect to the CSV algorithm is of the order of 2 to 4% in b jet identification efficiency when comparing at the same misidentification probability for light-flavour jets. The improvement of using IVF vertices with respect to using AVR vertices in the CSVv2 algorithm is of the order of 1 to 2%. For light-flavour jets, the absolute efficiency improves by about 4% with respect to the CSVv2 algorithm.

The value of the discriminator threshold for the b tagging algorithms used in this analysis and the corresponding efficiencies are summarized in Table 6.13. In this analysis the medium working point is used to identify and reject  $t\bar{t}$  events where a real b jet is expected in addition to the large-cone jet used to reconstruct the hadronically decaying high- $p_T$  V or H jet candidate representing instead the signal. The same b tagging algorithm but together with the loose working point is used to identify whether the CA8 jet comes from a Higgs boson decaying into bottom quarks, as described in Section 7.3.

The mismodelling of the b tagging variables in simulation is taken into account by reweighting simulation event-by-event with the ratio of the b tagging efficiency in data and simulation, determined in a sample enriched with b jets and depending on the jet  $p_T$  and  $\eta$ . The correction factors as a function of the b jet  $p_T$  are shown in Fig. 6.24(a) and 6.24(b) for the CSVM and CSVv2M operating points respectively, as measured in 8 and 13 TeV data. In a similar way, correction factors are also derived and applied to correct the misidentification probability in simulation. These factors are shown in Fig. 6.25(a) and 6.25(b) as a function

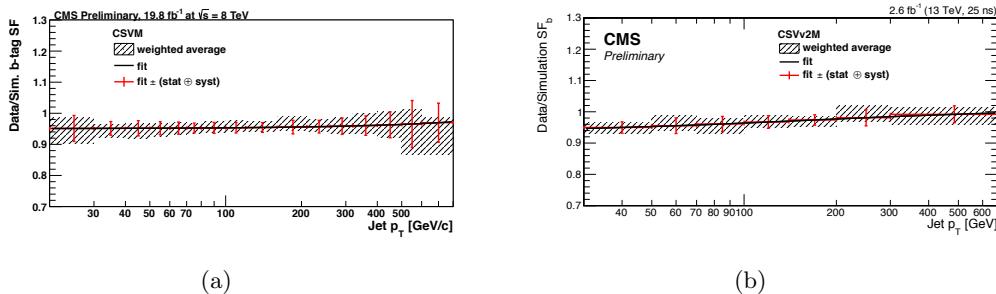


**Figure 6.23:** Performance of the CSVv2 algorithm demonstrating the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The improvement of this algorithm with respect to the Run 1 version of the algorithm is also shown [55].

**Table 6.13:** Taggers, discriminator threshold for Run 1 and Run 2 and corresponding efficiency for b jets with  $p_T > 30$  GeV in simulated  $t\bar{t}$  events.

Algorithm	operating point	discriminator value	b tagging efficiency (%)
CSV (Run 1)	CSVL	0.244	84
	CSVM	0.679	68
	CSVT	0.898	52
	CSVv2L	0.460	83
	CSVv2M	0.800	69
	CSVvsT	0.935	49

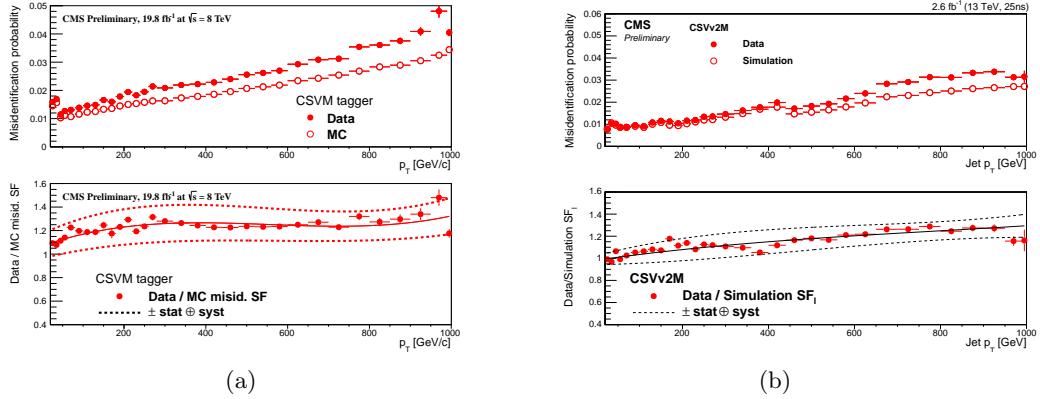
of the jet  $p_T$  for the CSVM and CSVv2M operating points, as measured in 8 and 13 TeV data.



**Figure 6.24:** Data-to-simulation correction factors for the b tagging efficiency for the CSVM (a) and CSVv2M (b) algorithms as a function of the b jet  $p_T$  as measured in 8 and 13 TeV data, respectively [54, 55].

## 6.5 Missing transverse energy

CMS is a full coverage hermetic detector which identifies and reconstructs almost all stable or long-lived particles produced in pp collisions. The only exceptions are neutrinos and hypothetical neutral weakly interacting particles. Although these particles do not leave a



**Figure 6.25:** Data-to-simulation correction factors for the misidentification probability for the CSVM (a) and CSVv2M (b) algorithms as a function of the jet  $p_T$  as measured in 8 and 13 TeV data, respectively [54, 55].

signal in the detector, their presence can be inferred from the momentum imbalance in the transverse plane, a quantity known as missing transverse momentum and denoted by  $\vec{E}_T^{\text{miss}}$ .

In CMS several standard methods are available for the reconstruction of the  $\vec{E}_T^{\text{miss}}$  as for the jet reconstruction: using only calorimeter information, including also tracker information, or the PF algorithm [56]. In this analysis, the PF  $\vec{E}_T^{\text{miss}}$  is used along with PF jets and it is calculated as the negative vector sum of the transverse momenta of all reconstructed PF candidates in a given event

$$\vec{E}_T^{\text{miss}} = - \sum_i^N \vec{p}_{T,i}. \quad (6.3)$$

Its magnitude is referred to as missing transverse energy and denoted by  $E_T^{\text{miss}}$ . The  $E_T^{\text{miss}}$  is a key variable in many searches for physics beyond the standard model such as the ones described in this thesis where a real highly energetic neutrino is expected in the final state. In addition, the precise measurement of  $E_T^{\text{miss}}$  plays a crucial role for measurements of standard model physics involving W bosons and top quarks.

The  $\vec{E}_T^{\text{miss}}$  reconstruction is sensitive to pileup, detector malfunctions and to various reconstruction effects and a precise calibration of all reconstructed physics objects is crucial for its performance. The level of mismeasurement is significantly reduced after jet energy calibration, described in Section 6.4.2. A correction to the  $\vec{E}_T^{\text{miss}}$  is derived by propagating the jet energy scale corrections as described in the following.

The raw missing transverse momentum can be written as:

$$\vec{E}_T^{\text{miss,raw}} = - \sum_i^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{raw}} - \sum_i^{N_{\text{uncl}}} \vec{p}_{T,i} \quad (6.4)$$

where the first and second sum runs over the  $p_T$  of the PF candidates clustered as jets and unclustered, respectively, and the superscript “raw” indicates the uncorrected  $p_T$ . The correction to the  $\vec{E}_T^{\text{miss}}$  is then obtained by replacing the first sum with the vector sum of the transverse momenta of the jets to which jet energy scale corrections are applied:

$$\vec{C}_T^{\text{JEC}} = \sum_i^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{raw}} - \sum_i^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{JEC}} \quad (6.5)$$

where the sum is performed over all jets with corrected  $p_T > 10 \text{ GeV}$ .

Further corrections improve the performance of the  $\vec{E}_T^{\text{miss}}$  reconstruction in events with large numbers of pileup interactions. This is achieved as explained in the following.

The raw  $\vec{E}_T^{\text{miss}}$  can be written as a sum of the two contributions due to particles produced in the primary vertex (PV) and in pileup interactions (PU)

$$\vec{E}_T^{\text{miss,raw}} = - \sum_{i \in \text{PV}} \vec{p}_{T,i} - \sum_{i \in \text{PU}} \vec{p}_{T,i}. \quad (6.6)$$

Particles produced in the pileup interactions can be further classified into neutral (PUneu) and charged (PUCch) particles so that the equation above can be expressed as

$$\vec{E}_T^{\text{miss,raw}} = - \sum_{i \in \text{PV}} \vec{p}_{T,i} - \sum_{i \in \text{PUCch}} \vec{p}_{T,i} - \sum_{i \in \text{PUneu}} \vec{p}_{T,i}. \quad (6.7)$$

The contribution to the genuine  $\vec{E}_T^{\text{miss}}$  from such interactions is close to zero, as the probability to produce neutrinos is small in inelastic pp scattering interactions (e.g. neutrinos from Kaon decays). The vectorial  $\vec{p}_T$  sum of charged particles is therefore expected to be well balanced by that of neutral particles. However, the nonlinearity and minimum energy thresholds in the calorimeters cause  $\vec{E}_T^{\text{miss}}$  to point on average in the direction of the vectorial  $\vec{p}_T$  sum of neutral particles. Nevertheless, it can be assumed that the directions of neutral pileup particles is measured with high precisions from the positions of the calorimeter cells in which we observe the energy deposits, while their energies are systematically off by the same factor. At the same time, the CMS tracker can also measure very well the charged pileup particles from their large curvature due to the low  $p_T$  characterizing this type of processes. With these assumptions, the total contribution from pileup can be estimated as

$$\vec{\Delta}_{\text{PU}} = \sum_{i \in \text{PUCch}} \vec{p}_{T,i} + \sum_{i \in \text{PUneu}} \vec{p}_{T,i} = \sum_{i \in \text{PU}} f(\vec{v}) \vec{v} \quad (6.8)$$

where  $\vec{v}$  represents the sum of the transverse momenta of charged particles for each pileup interaction. The correction  $f(\vec{v})$  is parametrized as  $f(\vec{v}) = c_1(1.0 + erfc(-c_2|\vec{v}^{\text{c3}}|))$ , where the coefficients  $c_1$ ,  $c_2$ , and  $c_3$  are extracted from simulated minimum bias events. The corrected  $\vec{E}_T^{\text{miss}}$  is then obtained removing the additional contribution  $\vec{\Delta}_{\text{PU}}$  from Eq. 6.6

$$\vec{E}_T^{\text{miss,PUCcorr}} = \vec{E}_T^{\text{miss,raw}} + \vec{\Delta}_{\text{PU}} \quad (6.9)$$

Another type of correction is derived and applied to correct for a modulation in  $\phi$  in the  $\vec{E}_T^{\text{miss}}$  present not only in data but also in simulation. The distribution of genuine  $\vec{E}_T^{\text{miss}}$  is instead independent of  $\phi$  because of the rotational symmetry of the collisions around the beam axis. The possible causes of the modulation include imperfect detector alignment, inefficiencies, a residual  $p_T$  dependence of the calibration, and a shift between the centre of the detector and the beam line. The correction for this effect can be expressed as a shift in the  $\vec{E}_T^{\text{miss}}$  components along the  $x$  and  $y$  detector coordinates, which increases approximately linearly with the number of reconstructed vertices. This correlation is used for a correction procedure as follows

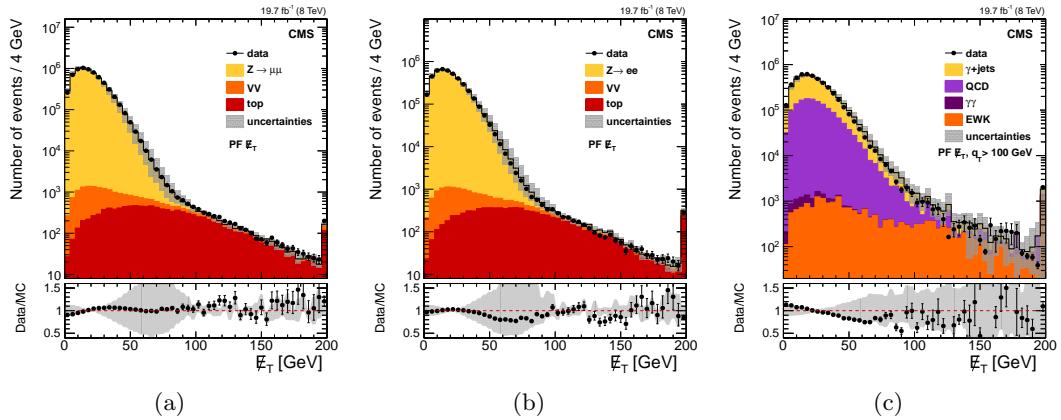
$$\vec{E}_{T,x}^{\text{miss,phi corr}} = \vec{E}_{T,x}^{\text{miss,raw}} - (c_{x_0} + c_{x_s} N_{\text{vtx}}) \vec{E}_{T,y}^{\text{miss,phi corr}} = \vec{E}_{T,y}^{\text{miss,raw}} - (c_{y_0} + c_{y_s} N_{\text{vtx}}) \quad (6.10)$$

where the coefficients are determined separately for data and simulated events.

Other more sophisticated missing energy determinations aimed at improving the resolution have been developed in CMS [57, 58] but will not be discussed in this Section since they are not used in this work.

The distributions of the PF  $E_T^{\text{miss}}$ , obtained after applying all the corrections described above, in  $Z \rightarrow \mu^+ \mu^-$ ,  $Z \rightarrow e^+ e^-$ , and prompt photon events are presented in Fig. 6.26 for simulation and for data collected at 8 TeV. Good agreement between data and simulation is observed in all distributions.

These events contain no genuine  $\vec{E}_T^{\text{miss}}$ , and thus a balance exists between the well-measured vector boson transverse momentum, denoted as  $\vec{q}_T$ , and the hadronic recoil, denoted as  $\vec{u}_T$ , which dominates the  $\vec{E}_T^{\text{miss}}$  measurement. The  $q_T$  can therefore be used as a reference to measure the scale and resolution of  $\vec{E}_T^{\text{miss}}$ . The hadronic recoil can be projected to the axis defined by  $q_T$ , yielding two signed components, parallel ( $u_{\parallel}$ ) and perpendicular ( $u_{\perp}$ ) to this axis. The parallel component is typically negative as the observed hadronic system is usually in the hemisphere opposite the boson. The scalar quantity  $-\langle u_{\parallel} \rangle / \vec{q}_T$  is referred to as the  $\vec{E}_T^{\text{miss}}$  response. The response curves, extracted from the data as a function of the vector boson boost  $\vec{q}_T$ , are shown in Fig. 6.27(a), where deviations from unity indicate a bias on the hadronic recoil energy scale which is fully recovered for  $\vec{q}_T > 40$  GeV. The resolution curves,  $\sigma(u_{\parallel})$  and  $\sigma(u_{\perp})$  as a function of  $q_T$ , are shown in Fig. 6.27(b) and 6.27(c), respectively, for each control sample. The resolution increases with increasing  $q_T$ , while the data and simulation curves are in good agreement for each control sample.

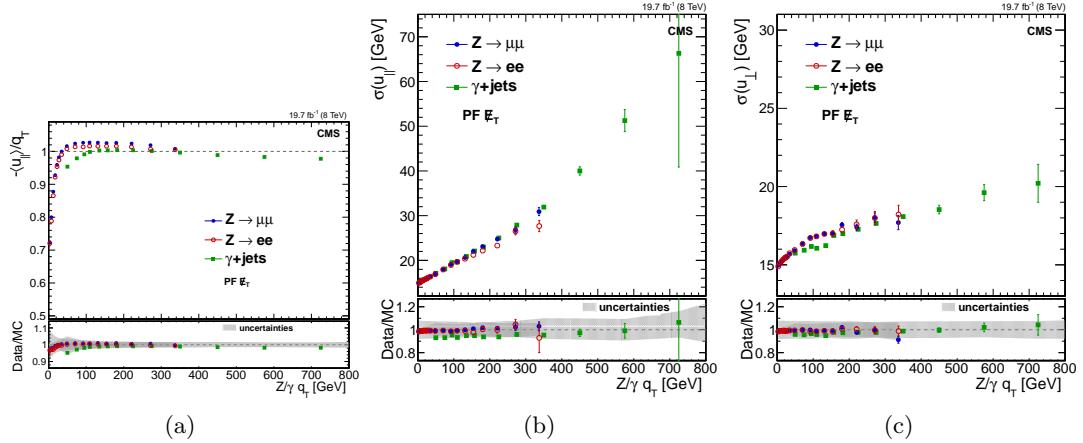


**Figure 6.26:** The PF  $E_T^{\text{miss}}$  distribution in  $Z \rightarrow \mu^+ \mu^-$  (a),  $Z \rightarrow e^+ e^-$  (b), and prompt photon (c) events for data collected at 8 TeV. The points in the lower panel of each plot show the data/MC ratio [57].

## 6.6 $W \rightarrow \ell \nu$ reconstruction

The identified muon or electron (see Section 6.2.3 and 6.3.3) is associated with the  $W \rightarrow \ell \nu$  candidate. The  $p_T$  of the undetected neutrino is assumed to be equal to the  $E_T^{\text{miss}}$ . The longitudinal momentum of the neutrino ( $p_z$ ) is obtained by solving a quadratic equation that sets the  $\ell \nu$  invariant mass to the known  $W$  boson mass [59]:

$$M_W^2 = m_\ell^2 + 2(E_\ell E_\nu - p_{x\ell} p_{x\nu} - p_{y\ell} p_{y\nu} - p_{z\ell} p_{z\nu}) = (80.4)^2 \quad (6.11)$$



**Figure 6.27:** (a) Response curves for PF  $\vec{E}_T^{\text{miss}}$  in events with a Z-boson or prompt photon. Also shown are the resolution curves of the parallel (b) and perpendicular (b) recoil components as a function of the versus  $Z/\gamma q_T$  for PF  $\vec{E}_T^{\text{miss}}$  in events with a Z-boson or prompt photon. Data correspond to pp collisions at 8 TeV. In each plot the upper frame shows the response in data, while the lower one shows the ratio between data and simulation. [57].

In the case of two real solutions, the one with the smaller absolute value is chosen; in the case of two complex solutions, only the real part is used. The four-momentum of the neutrino is used to reconstruct the four-momentum of the  $W \rightarrow \ell\nu$  candidate. The same procedure is applied for  $W \rightarrow \tau\nu$  candidates, where the  $\tau$  decays to one muon or electron and two neutrinos. In this case, the  $E_{T\text{miss}}$  represents the  $p_T$  of the three-neutrino system.

In the case of two real solutions, the one with the smaller absolute value is chosen. If the discriminant becomes negative, or equivalently  $M_T$  is larger than  $M_W$  used in the constraint, the solutions have an imaginary part. This happens because of the finite resolution of  $E_T^{\text{miss}}$ . Several schemes exist to deal with this situation. One technically simple method consists of taking the real part of the complex solutions but it leads to the wrong W boson mass. This method is used for the reconstruction of the  $W \rightarrow \ell\nu$  candidate in the 13 TeV analysis described in this work.

A second method is also studied here, which eliminates the imaginary component by modifying the components of the missing transverse energy such to give  $M_T = M_W$ , still respecting equation 6.11 [60]. This method is used in the 8 TeV analysis described in this work for the reconstruction of the  $W \rightarrow \ell\nu$  candidate and for the reconstruction of the mass of the leptonically decaying top quark in  $t\bar{t}$  events. The performance of the two methods are equivalent in terms of resolution of the reconstructed diboson mass.

# Identification of highly boosted $W/Z \rightarrow q\bar{q}(')$ and $H \rightarrow b\bar{b}$

---

**FIXME:** Assume that I already introduced jet clustering algorithms in Section 6.4. And I already introduce the signal topology in Section 4.

Large-cone jets (see Section 6.4), also referred to as “fat jets”, are used to reconstruct the  $W$  jet,  $Z$  jet, and  $H$  jet candidates resulting after the hadronization of the two quarks from the decay of highly boosted  $W$ ,  $Z$ , and Higgs boson, respectively. To discriminate against multijet backgrounds, the analyses exploit both the reconstructed jet mass, which is required to be close to the boson mass, and the jet substructure arising from the two jet cores that correspond to the two high- $p_T$  decay quarks.

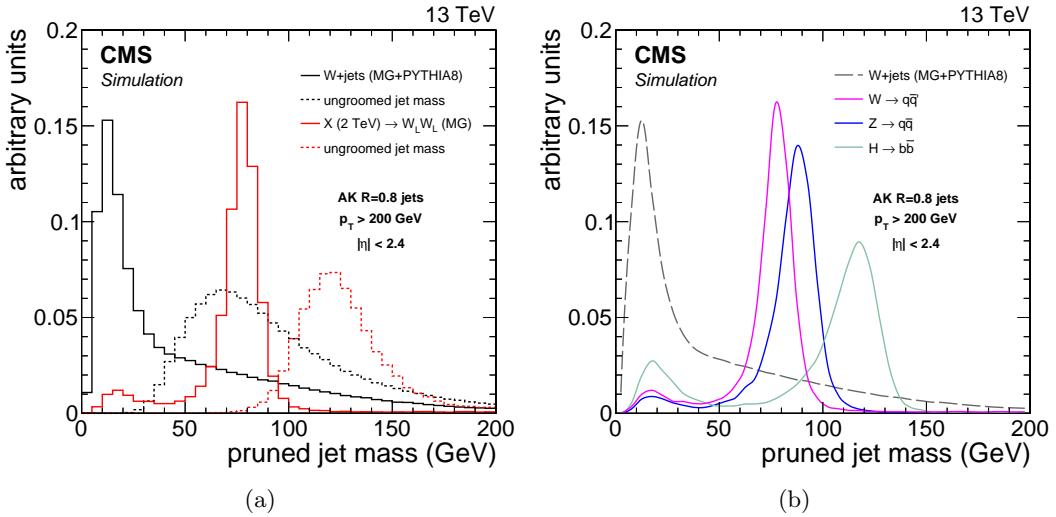
The techniques to identify jets arising from the merged decay products of a single  $V$  or Higgs boson is referred to as “ $V$  tagging” or “ $H$  tagging”, respectively. It employs novel jet substructure algorithms, which are described in Section 7.1. The features of the  $V$  tagging algorithm are described in Section 7.2 and its performance in both data and simulation are discussed. Finally, in Section 7.3, a procedure tuned to the specific properties of the Higgs boson decay into a bottom quark-antiquark pair is presented.

## 7.1 Jet substructure observables

### 7.1.1 Pruned jet mass

As the mass of the  $V$  or  $H$  boson is larger than the mass of a typical QCD jet, the jet mass is the primary observable that distinguishes a  $W$  jet from a QCD jet. The bulk of the  $W$  jet mass arises from the kinematics of the two jet cores that correspond to the two decay quarks. In contrast, the QCD jet mass arises mostly from large-angle and soft gluon radiation. As a first step in exploring potential substructure, the jet constituents are subjected to a jet grooming algorithm that improves the resolution in the jet mass and reduces the effect of pileup [39, 61]. The goal of jet grooming is to recluster the jet constituents, while applying additional requirements that eliminate soft, large-angle QCD radiation. This procedure shifts the jet mass of QCD jets to smaller values, while maintaining the mass for signal jets close to the boson mass. Furthermore, soft contributions from the underlying event and pileup, usually present in all jets, are removed. Different jet grooming algorithms have been explored at CMS and their performance on jets in multijet processes has been studied in detail [39, 61]. In this analysis, the *jet pruning* algorithm [62, 63] is used, as it provides the best discrimination against QCD background as discussed in Ref. [39, 61].

Jet pruning reclusters each large-cone jet starting from all its original constituents, through the implementation of the CA algorithm, but applying two additional conditions beyond those given in **FIXME: refer to the general jet clustering equation**. In particular, the softer of the two particles  $i$  and  $j$  to be merged is removed when the following conditions are met:



**Figure 7.1:** (a) Distribution in pruned jet mass  $m_{\text{jet}}$  for simulated events of highly boosted W bosons and inclusive QCD jets expected in the W+jet topology. The ungrouped jet mass is shown as dotted lines to illustrate the effect of pruning. MG denotes the MADGRAPH generator. (b) Comparison of the distributions in  $m_{\text{jet}}$  for simulated events of highly boosted V and Higgs bosons.

$$z_{ij} \equiv \frac{\min(p_{Ti} + p_{Tj})}{p_{Ti} + p_{Tj}} < z_{cut} , \quad \Delta R_{ij} > D_{cut} \equiv \alpha \frac{m_j}{p_T} \quad (7.1)$$

where  $m_j$  and  $p_T$  are the mass and transverse momentum of the originally-clustered jet, and  $z_{cut}$  and  $\alpha$  are parameters of the algorithm, chosen to be 0.1 and 0.5, respectively. In this particular choice of parameters, the algorithm removes the largest number of jet constituents, and can therefore be regarded as the most aggressive jet grooming technique. The resulting jet is the *pruned jet*. The pruned jet mass,  $m_{\text{jet}}$ , is computed from the sum of the four-momenta of the constituents that survive the pruning; it is then corrected by the same factor used to correct the jet  $p_T$  (see Section 6.4). Figure 7.1(a) illustrates the effect of pruning on AK8 jets: the  $m_{\text{jet}}$  spectrum of the W jet candidate from the decay of highly boosted and longitudinally polarized W bosons is shown together with the distribution in  $m_{\text{jet}}$  for the simulated background of W+jets. Dashed and solid lines correspond to the distributions before and after the application of the pruning algorithm, respectively. Fully merged jets reconstructed from the W boson decay generate a distinctive peak around the W boson mass, which is narrowed by the pruning, while background jets acquire a smaller mass on average, enhancing the discrimination. Figure 7.1(b) compares the distributions in  $m_{\text{jet}}$  for W, Z and H jet candidates from the decay of highly boosted W, Z and Higgs bosons, respectively. The distribution in  $m_{\text{jet}}$  for the W+jets background is also shown. Not-full-merged signal jets give rise to a peak at low masses.

### 7.1.2 N-subjettiness

In addition to the pruned jet mass, additional information about the jet shape is used to discriminate the signal against jets from gluon and single-quark hadronization. This information can be obtained from the quantity called *N-subjettiness* [64]. It takes advantage of the multi-body kinematics in the decay pattern of boosted hadronic objects, and it can be used to effectively “count” the number of subjets in a given jet.

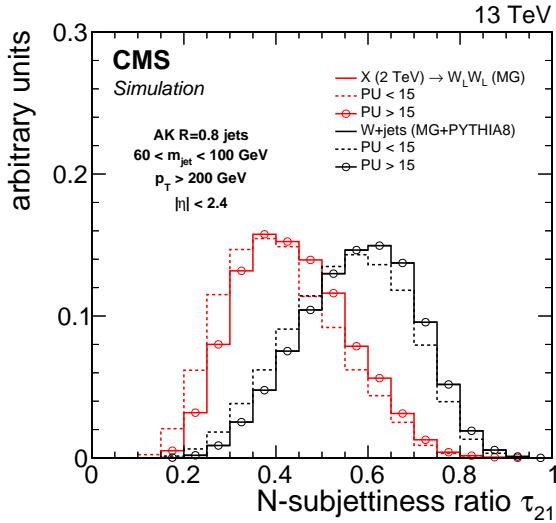
The N-subjettiness is a generalized jet shape observable which defines a measure,  $\tau_N$ , for a jet to have  $N$  subjets. The constituents of the jet before the pruning procedure are reclustered using the  $k_T$  algorithm (see Section 6.4), until  $N$  joint objects (subjets) remain in the iterative combination procedure of the algorithm. The observable  $\tau_N$  is then defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\} \quad (7.2)$$

where  $k$  runs over the constituents of the jet, and the distances  $\Delta R_{n,k}$  are calculated relative to the axis of the  $n$ th subjet. The normalization factor  $d_0$  is taken as

$$d_0 = \sum_k p_{T,k} R_0 \quad (7.3)$$

where  $R_0$  is the characteristic jet radius used in the original jet clustering algorithm. The subjet axes are obtained by running the exclusive  $k_T$  algorithm [35], and reversing the last  $N$  clustering steps. The variable  $\tau_N$  quantifies to what degree the jet can be regarded as a jet composed of  $N$  subjets. Jets with  $\tau_N \approx 0$  have all their radiation aligned with the candidate subjet directions and therefore have  $N$  (or fewer) subjets. Jets with  $\tau_N \gg 0$  have a large fraction of their energy distributed away from the candidate subjet directions and therefore have at least  $N + 1$  subjets. The ratio between 2-subjettiness and 1-subjettiness,  $\tau_{21} = \tau_2/\tau_1$ , is found to be a powerful discriminant between jets originating from hadronic V decays and from gluon and single-quark hadronization. Jets from W or Z decays in signal events are characterized by lower values of  $\tau_{21}$  relative to QCD background. Figure 7.2 shows the  $N$ -subjettiness ratio  $\tau_{21}$  distribution for W jets and QCD jets after requiring  $60 < m_{\text{jet}} < 100$  GeV, and demonstrates its discrimination power after the pruned jet mass selection.



**Figure 7.2:** Distribution in N-subjettiness ratio  $\tau_{21}$  for simulated events of highly boosted and longitudinally polarized W bosons and inclusive QCD jets expected in the W+jet topology. The distributions are shown after a selection on the pruned jet mass requiring  $60 < m_{\text{jet}} < 100$  GeV. MG denotes the MADGRAPH generator. The histograms are the expected distributions after full CMS simulation with pileup corresponding to an average number above and below 15 interactions.

## 7.2 The V tagging algorithm

The substructure techniques described in the previous section are employed for identifying, or “tagging”, W and Z jets (“V jets”). The V tagging of the jets is obtained combining selections on both the pruned jet mass  $m_{\text{jet}}$  and N-subjettiness ratio  $\tau_{21}$  observables.

The selection criteria have been optimized in the context of searches for resonances decaying into diboson in the  $\ell\nu$ +jet and dijet final states [65–67]. The optimization, based on simulation, aims at maximizing the analysis sensitivity and it leads to slightly different working points for each analysis. Typical signal efficiencies and mistagging rates of QCD jets obtained, respectively, from simulations and measurements for  $\sqrt{s} = 8$  and 13 TeV are summarized in Table 7.1, for jets with  $p_T = 500$  GeV.

**Table 7.1:** Typical selection criteria for V tagging in 8 and 13 TeV analyses. The corresponding signal efficiency and mistagging rate of QCD jets are also reported for jets with  $p_T = 500$  GeV.

$\sqrt{s}$	V tagging selections	signal efficiency	mistagging rate
8 TeV	$60 < m_{\text{jet}} < 100$ GeV $\tau_{21} < 0.5$	0.65	0.04
13 TeV	$65 < m_{\text{jet}} < 105$ GeV $\tau_{21} < 0.45$	0.55	0.03

The  $\ell\nu$ +jet analysis described in this work makes use of a looser  $\tau_{21}$  working point of 0.6 resulted from an optimization which takes into account signal efficiency and background rejection over a large jet  $p_T$  range. In fact, this channel is characterized by a low background rate and a  $\tau_{21}$  selection which provides enhanced signal efficiency over the whole jet  $p_T$  range is therefore preferred. This working point corresponds to a signal efficiency of 65% and a mistagging rate of 5%.

The V tagging performance at 8 TeV have been studied in detail in Ref. [39]. From simulation studies it is observed that the efficiency of the  $m_{\text{jet}}$  selection increases with  $p_T$  up to about 600 GeV since at higher  $p_T$  the showers from the W decay quarks are more likely to be reconstructed within a single large-cone jet. Above 600 GeV, the efficiency begins to decrease as a function of jet  $p_T$ , since at very large values the PF candidate reconstruction degrades in resolving the jet substructure and the pruning algorithm therefore removes too large a fraction of the jet mass. For Run II of the LHC, the particle flow reconstruction has been optimized by exploiting the full potential of the CMS ECAL granularity to resolve jet substructure and a constant efficiency is maintained up to at least  $p_T = 2.5$  TeV [46, 68].

The efficiency of the additional  $\tau_{21}$  selection also drops as a function of  $p_T$ . It is important to note that the same efficiency at an equivalent background rejection rate can be reached by adjusting the maximum  $\tau_{21}$  as a function of  $p_T$ . Therefore, a fixed working point will degrade the efficiency with increasing  $p_T$ . However, by shifting the working point, the same performance can be achieved. This possibility has not been explored yet in any of the searches which employ V tagging.

The efficiency of the V tagging selection as a function of the number of reconstructed primary vertices (PV) has also been studied [46]. It is observed that the efficiency of the  $m_{\text{jet}}$  selection is constant as a function of PV, whereas the additional  $\tau_{21}$  selection efficiency drops from 60% at 0 PV to 40% at 30 PV. However, the mistagging of the background also decreases with pileup for the same selection, yielding similar discrimination. Efficiency and mistagging rate are affected by pileup in the same way, since additional pileup shifts the  $\tau_{21}$  distribution towards higher values (towards background like) for both signal and

background (see Fig. 7.2). Therefore, the same signal efficiency can be reached at the same background rejection rate for up to 30 reconstructed vertices by merely adjusting the  $\tau_{21}$  selection.

An important factor that influences the V tagging performance is the polarization of the reconstructed V bosons. The pruned jet mass selection is less efficient for transversely polarized ( $V_T$ ) V bosons. This can be explained by a higher asymmetry in the  $p_T$  of the two quarks from the  $V_T$  boson decay, such that the pruning algorithm in a considerable fraction of events rejects the particles from the lower  $p_T$  quark and yields a much lower jet mass. In addition, the  $\Delta R$  separation between the partons for pure longitudinally polarized ( $V_L$ ) V bosons is smaller on average than for  $V_T$  bosons and is more likely to be accepted by a large-cone jet. In the analysis presented in this work only  $V_L$  bosons are considered.

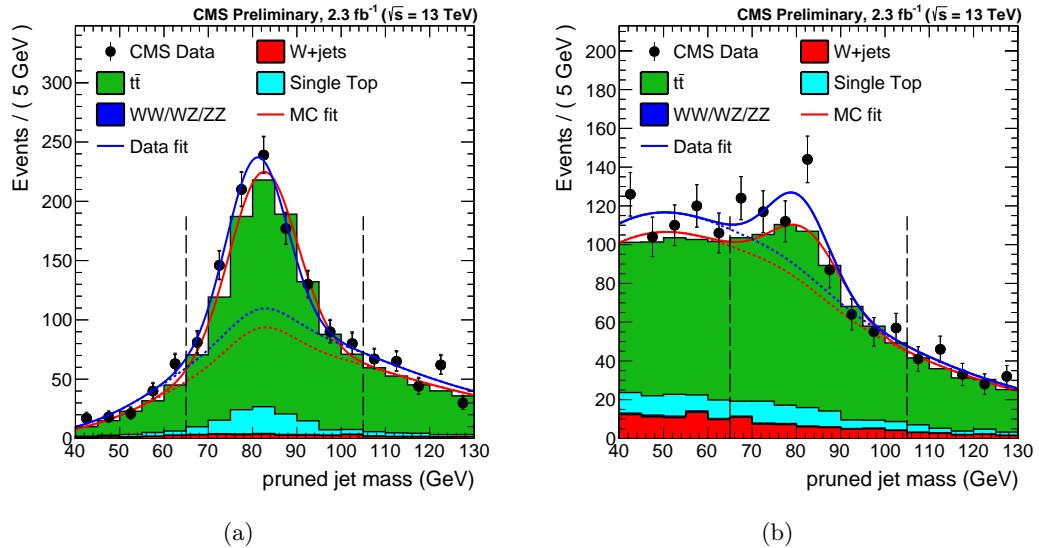
The  $\ell\nu$ +jet analysis described in this work relies on the modelling of the jet substructure variables  $m_{\text{jet}}$  and  $\tau_{21}$  in simulation. The data/simulation discrepancies in  $m_{\text{jet}}$  and  $\tau_{21}$  can bias the signal efficiency estimated from simulated samples. Therefore, the modelling of signal efficiency is cross-checked in a signal-free sample with jets having characteristics that are similar to those expected for a genuine signal [46]. A pure sample of high- $p_T$  W bosons, that decay to quarks and are reconstructed as a single large-cone jet, is obtained selecting  $t\bar{t}$  and single top quark events.

Scale factors for the  $\tau_{21}$  selection efficiency are extracted by estimating the selection efficiency on both data and simulation for the pure W jet signal. This is achieved by subtracting the background contribution. The generated W boson in the  $t\bar{t}$  simulation provides a model of the contribution from the W jet peak in the pruned jet mass. The contribution from combinatorial background is derived from  $t\bar{t}$  simulation as well. This signal plus background model is fitted directly in the distributions of data and in their simulation.

The pruned jet mass distribution of events that pass and fail the  $\tau_{21}$  selection are fitted simultaneously to extract the selection efficiency on the pure W jet component. The ratio of data and simulation efficiencies are taken as the V tagging efficiency scale factor. Figure 7.3 shows the fits for the  $\tau_{21} < 0.45$  selection and similar results are obtained for the looser  $\tau_{21} < 0.6$  selection used in the  $\ell\nu$ +jet analysis presented in this work. The extracted scale factor for this selection is  $1.01 \pm 0.03$  and it is used to correct the total signal efficiency and the VV background normalization predicted by the simulation. The quoted uncertainty includes two systematic effects. One from the modelling of the nearby jets and  $p_T$  spectrum in  $t\bar{t}$  MC events, obtained by comparing the selection efficiency estimated from LO and NLO  $t\bar{t}$  simulation. The other due to the choice of the models used to fit signal and background. The quadratic sum of these systematic uncertainties is found to be smaller than half of the statistical uncertainty on the scale factor. An additional uncertainty is calculated to account for the extrapolation of the scale factor from  $t\bar{t}$  events with an average jet  $p_T \sim 200$  GeV to higher momenta. This is estimated from the difference between PYTHIA8 and HERWIG++ [69] showering models with a resulting factor of  $4.53\% \times \ln(p_T/200\text{ GeV})$ .

The peak position in the W jet mass and its resolution are also extracted to obtain data-to-simulation corrections on the pruned jet mass listed in Table 7.2, where the quoted uncertainties are statistical. The W jet mass scale in data is  $\sim 1\%$  smaller than in simulation while its resolution is found to be larger by about 5%.

The mass peak position is slightly shifted relative to the W boson mass. The shift is found to be primarily due to extra radiation in the W jet from the nearby b quark, and additional effects are due to the presence of the extra energy deposited in the jet cone from pileup, underlying event, and initial-state radiation not completely removed in the jet pruning procedure.



**Figure 7.3:** Distribution in pruned jet mass for events that (a) pass and (b) fail the  $\tau_{21} < 0.45$  selection in the  $t\bar{t}$  control sample. The result of the fit to data and simulation are shown, respectively, by the solid and long-dashed line and the background components of the fit are shown as dashed- dotted and short-dashed line [46].  
**FIXME:** Can I put the plots for the tau21 0.6 from Thea even if they are not public?

Because the kinematic properties of W jets and Z jets are very similar, the same corrections are also used when the V jet is assumed to arise from a Z boson.

**Table 7.2:** W jet mass peak position and resolution, as extracted from top quark enriched data and from simulation. These are used to obtain corrections in the V tagging procedure.

$\tau_{21} < 0.45$	$m_{\text{jet}} [\text{GeV}]$	Standard deviation [GeV]
Data	$84.6 \pm 0.7$	$8.2 \pm 0.7$
Simulation	$85.1 \pm 0.2$	$7.8 \pm 0.3$

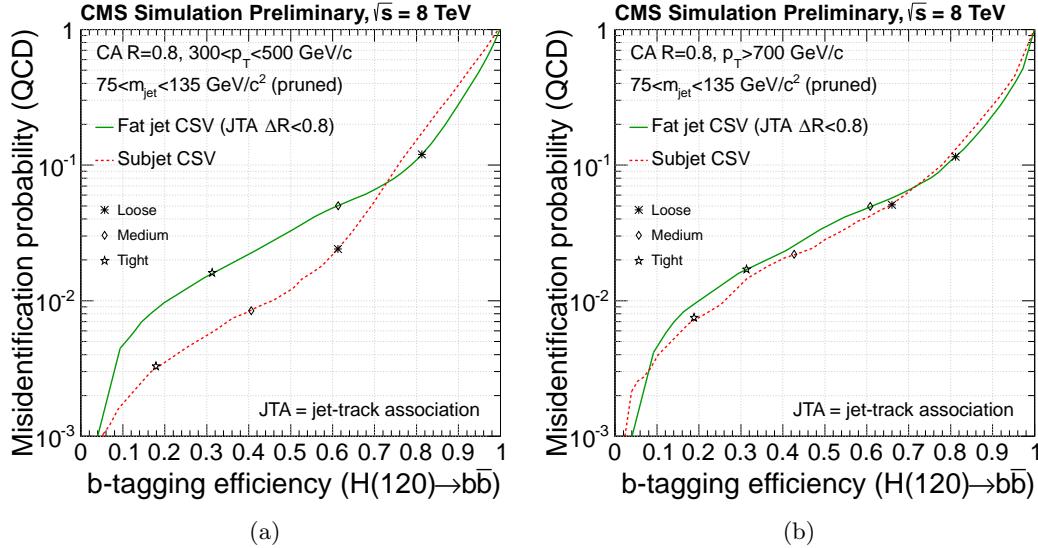
### 7.3 The H tagging algorithm

As discussed in the previous sections boosted V bosons are reconstructed using jet substructure methods through the V tagging algorithm, providing large discrimination against multijet backgrounds. However, if one or more of the decay products is a b quark, adding b jet identification (see Section 6.4) together with jet substructure information can significantly improve the sensitivity of these methods.

Two different approaches to identify boosted  $H \rightarrow b\bar{b}$  candidates have been explored and used at CMS [54]:

- application of b tagging to the fat jet (“fat jet b tagging”)
- application of b tagging to the subjets reconstructed within the fat jet (“subjets b tagging”)

Both approaches are based on the standard b tagging algorithms which take advantage of the tracking and vertexing information and are designed to identify jets from single b



**Figure 7.4:** Misidentification probability as a function of b tagging efficiency for boosted  $H \rightarrow b\bar{b}$  jets and inclusive QCD jets for the CSV algorithm applied to large-cone jets and pruned subjets for large-cone jets with (a)  $300 < p_T < 500$  GeV and (b)  $p_T > 700$  GeV. Loose, medium, and tight operating points of the CSV discriminator are indicated [54].

quarks. The CMS b tagging procedure starts with an association of tracks to jets, based on the angular distance between the tracks and the jet axis. The default b tagging algorithms use the selection  $\Delta R < 0.3$ . However, when applying this to a large-cone jet of size  $R = 0.8$ , the criteria is suboptimal. Hence, to apply b tagging to fat jets, this angular distance is enlarged to  $\Delta R < 0.8$ . For the application of b tagging to subjets, the angular distance remains at the default value of  $\Delta R < 0.3$ .

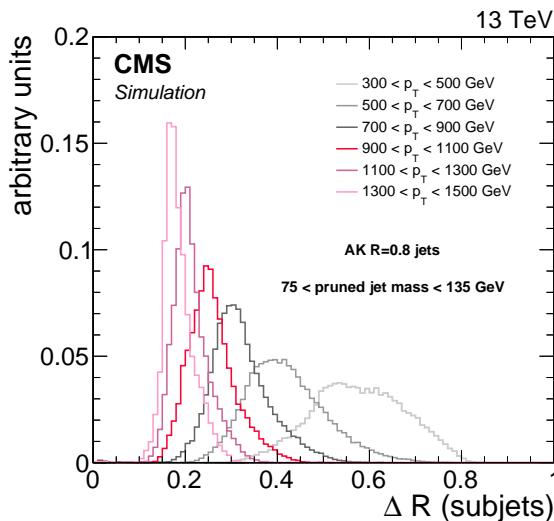
The H tagging technique starts requiring that the pruned jet mass of the H jet candidate lies in a window around the Higgs boson mass (see Figure 7.1(b)), as this requirement rejects a large fraction of QCD background as demonstrated in the previous sections. The subjets are then obtained by reversing the last step of the pruning recombination algorithm described in Section 7.1.1. In addition to the jet mass requirement, the b tagging is applied either to the whole fat jet or to the two subjets, where both subjets are required to pass the same selection on the CSV discriminator. The b tagging efficiency and misidentification probability of QCD jets after applying the selection  $75 < m_{jet} < 135$  GeV is shown in Fig. 7.4. The subjet b tagging outperforms the fat jet tagging for most of the phase space.

The H tagging efficiency obtained combining the requirement on the pruned jet mass ( $75 < m_{jet} < 135$  GeV) and the subjet b tagging at the CSVL operating point is between 40 and 50% for a H jet  $p_T$  range spanning from 300 GeV to 1 TeV, with a suppression of QCD background to about 0.4%.

The use of a fixed-size jet-track association cone inevitably leads to track sharing between the subjets of the jets once their angular separation becomes comparable or smaller than the size of the association cone. For boosted H jets the fraction of shared tracks, defined as the ratio of the number of tracks within  $\Delta R < 0.3$  from more than one subjet and the number of all tracks within  $\Delta R < 0.3$  from any of the subjets, ranges from a few percent at a jet  $p_T$  of 400 GeV and increases to 40% at a jet  $p_T$  of 700 GeV and to 80% at a jet  $p_T$  of

1 TeV. Figure 7.5 shows the distribution of the angular separation  $\Delta R$  of the two subjets reconstructed within the fat jet for different jet  $p_T$  ranges in simulated events of highly boosted Higgs bosons decaying to  $b\bar{b}$ . Because of track sharing, the b tagging probabilities for individual subjets deteriorate at large jet  $p_T$  and the subjet b tagging performance approach the fat jet b tagging ones as can be seen in Fig. 7.4. The lost in efficiency is then recovered applying the two approaches depending on the  $\Delta R$  between the two subjets. In particular, the analysis involving boosted Higgs bosons such as the one presented in this work apply subjet b tagging and fat jet b tagging if  $\Delta R > 0.3$  and  $< 0.3$ , respectively.

In the  $\ell\nu+b\bar{b}$  analysis presented in this work a requirement on the pruned jet mass of the reconstructed H jet candidate given by  $110 < m_{\text{jet}} < 135$  GeV is used. The  $m_{\text{jet}}$  window is chosen such that a contamination from possible signals with boosted V jets in the Higgs boson mass region is minimized. The b tagging is performed with the algorithm described above using the loose working point of the CSV discriminant. The total H tagging efficiency for these selections is about 35% for jet  $p_T$  of about 1 TeV with a mistagging probability below 1%.



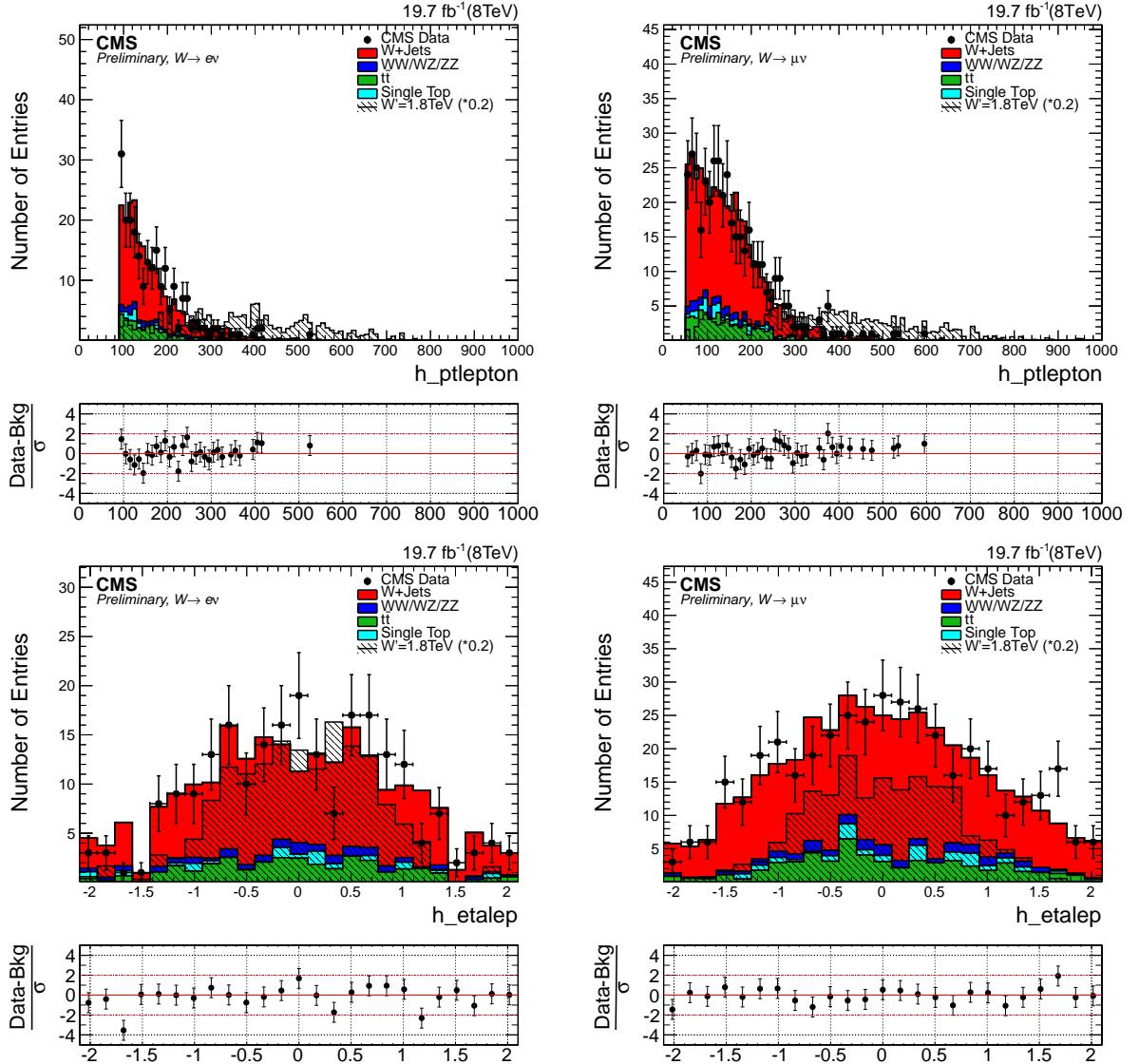
**Figure 7.5:** Distributions of the angular separation  $\Delta R$  of the two subjets reconstructed within the fat jet for simulated events of highly boosted Higgs bosons decaying to  $b\bar{b}$ . The distributions are compared for different ranges of the H jet  $p_T$ .

The validation of b tagging in boosted H jets is performed selecting events containing jets from gluon splitting to  $b\bar{b}$  ( $g \rightarrow b\bar{b}$ ) in which the b quarks hadronize inside one large-cone jet [54]. To enrich a sample of fat jets in  $g \rightarrow b\bar{b}$  component, used as an analogue of boosted  $H \rightarrow b\bar{b}$  jets, the fat jets are required to be double-muon-tagged with both subjets matched to distinct muon candidates within a cone of size  $\Delta R < 0.4$ . This sample is used to study the modelling of b tagging efficiencies in boosted  $H \rightarrow b\bar{b}$  topologies. The scale factors, given by the ratio between the efficiencies measured in data and simulation, are found to be in good agreement with those measured in the standard, non-boosted topologies, indicating that the simulation reproduces the b tagging performance in boosted and non-boosted environments equally well. These scale factors are used in the analysis to reweight the simulated events.

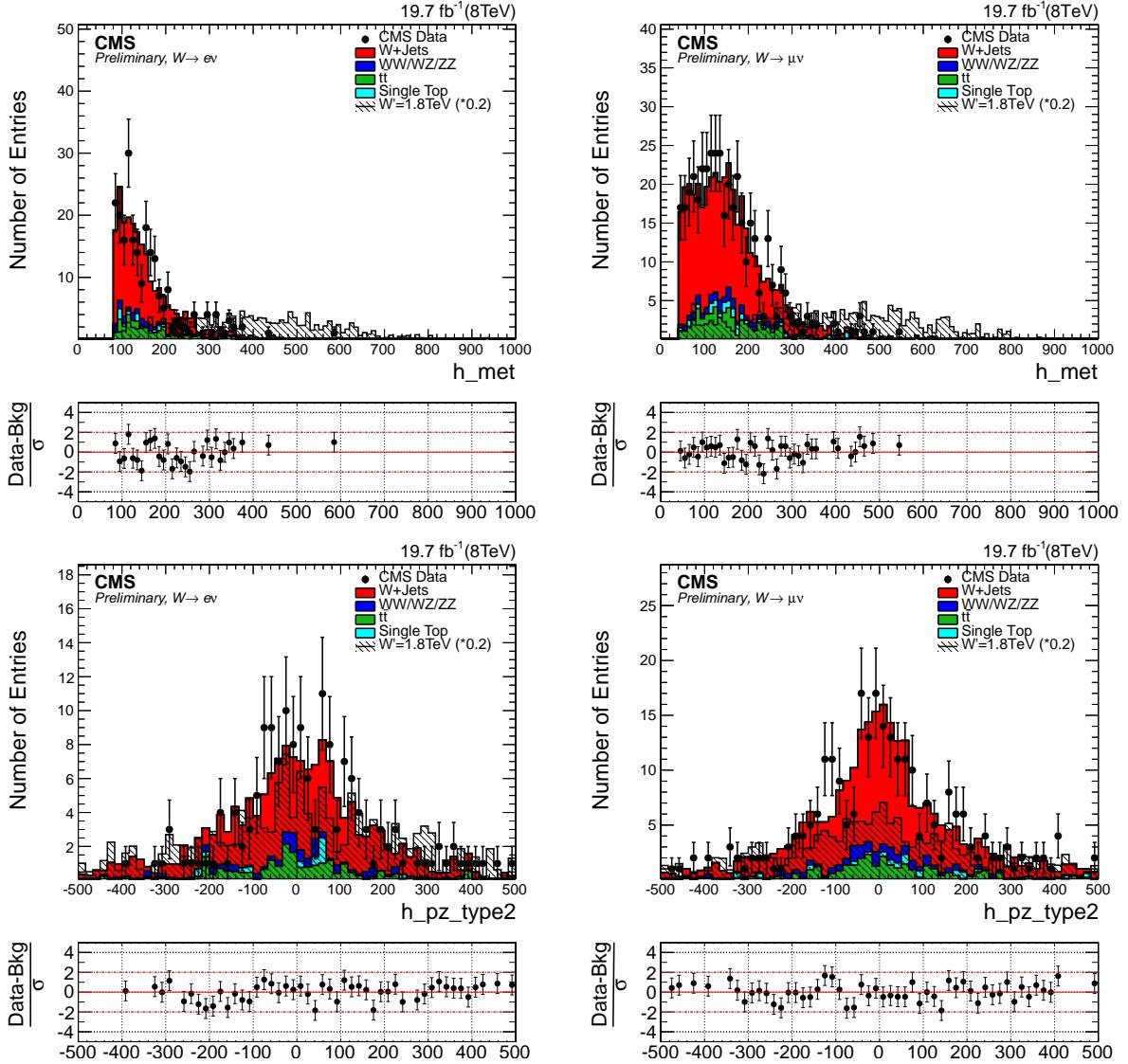
# Final event selection and categorization

---

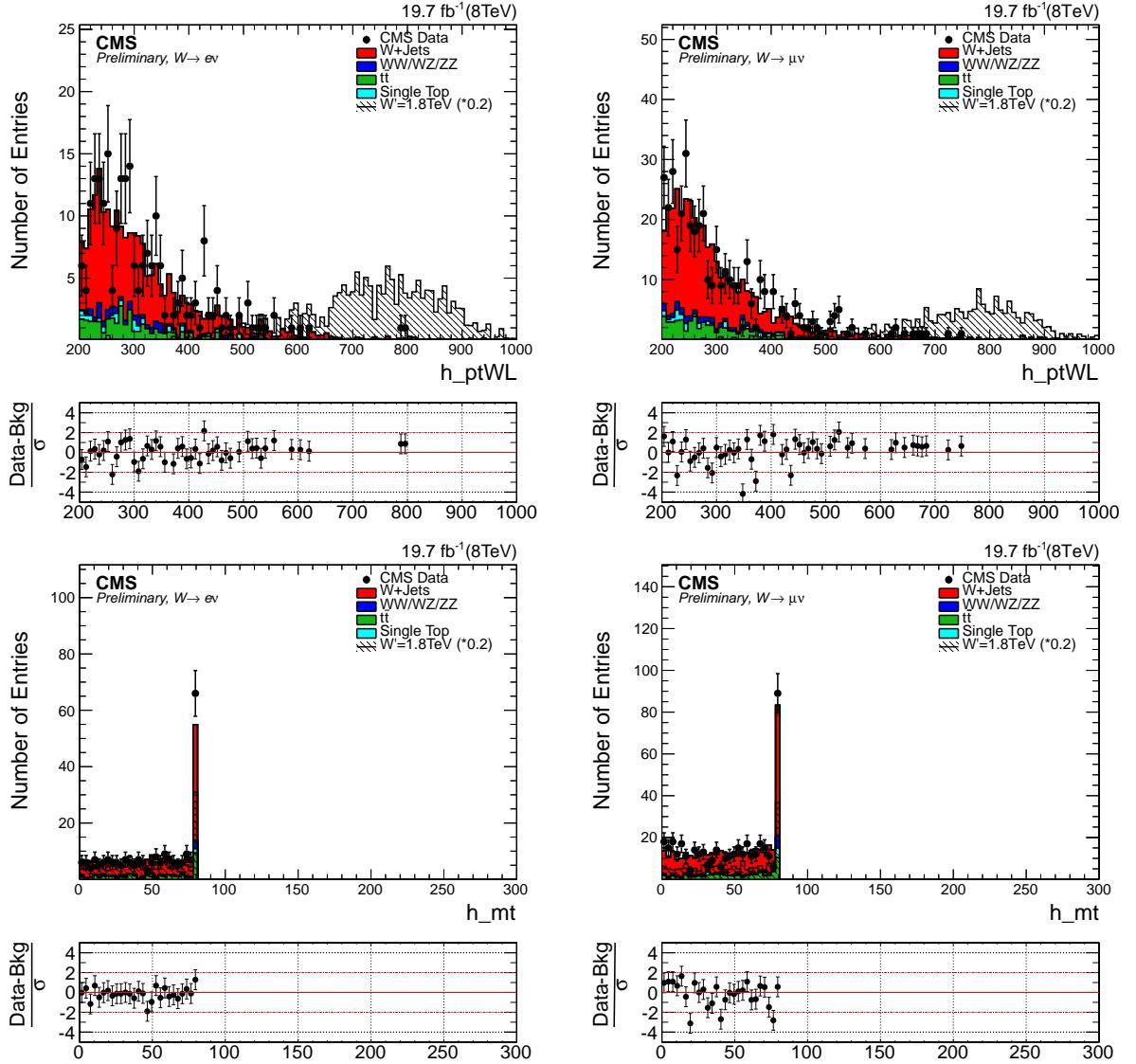
- 8.1 Search for a WH resonance in the  $\ell\nu b\bar{b}$  final state at  $\sqrt{s} = 8$  TeV
  - 8.1.1  $t\bar{t}$  background rejection
  - 8.1.2 Final selection and control plots
- 8.2 Search for WW/WZ resonances in the  $\ell\nu q\bar{q}^{(\prime)}$  final state at  $\sqrt{s} = 13$  TeV
  - 8.2.1 W/Z-jet mass categories
  - 8.2.2 Final selection and control plots



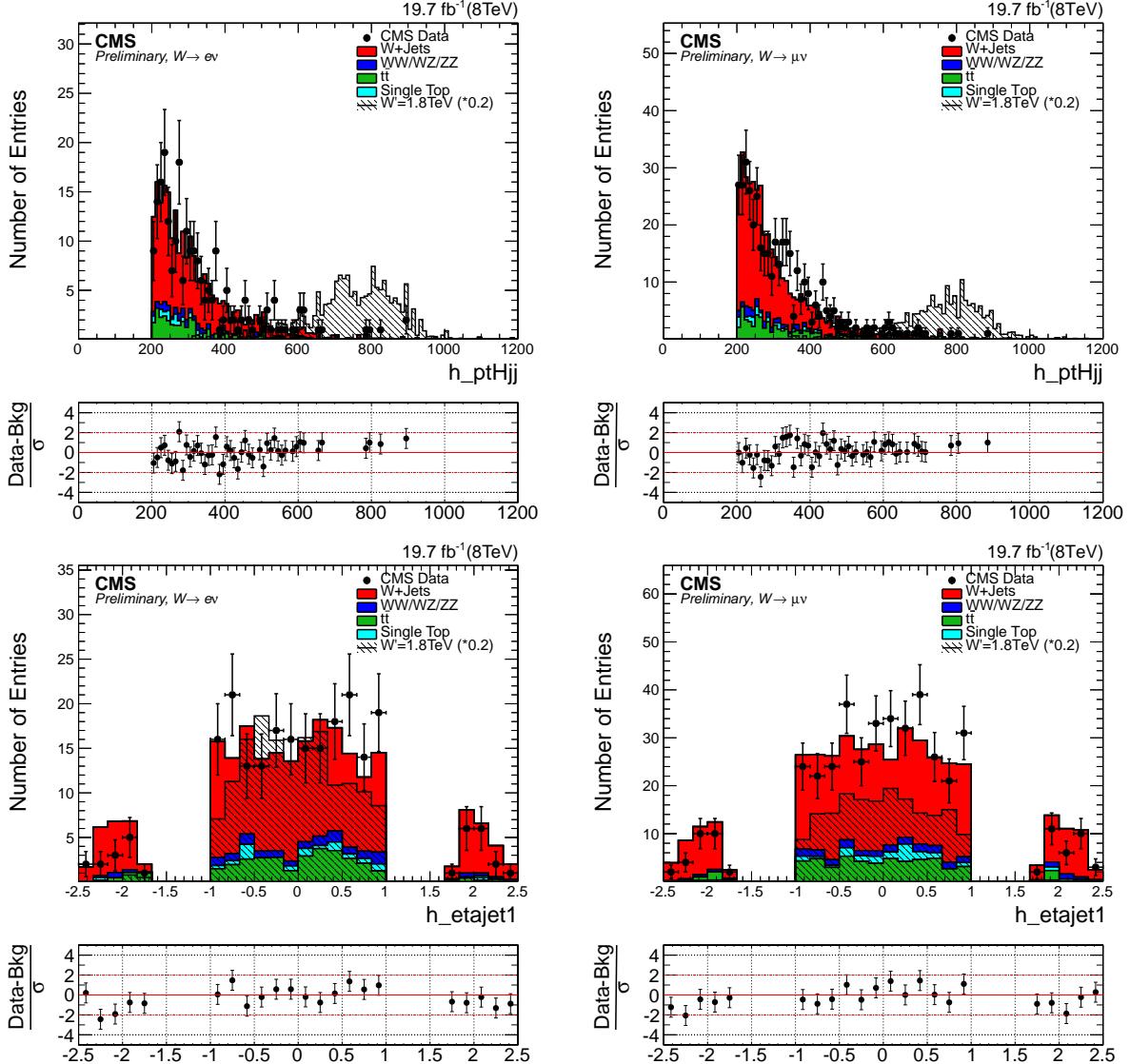
**Figure 8.1:** Lepton  $p_T$  and  $\eta$  for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 110$  GeV. The signal is normalized to 0.1 times of the selected events for the model considered ( $W'$ ,  $M = 1$  TeV).



**Figure 8.2:**  $E_T^{\text{miss}}$  and  $p_{z\nu}$  (defined in Section) for electron channel (left) and muon channel (right) for events with  $40 < m_{\text{jet}}^{\text{pruned}} < 110$  GeV. The signal is normalized to 0.1 times of the selected events for the model considered ( $W'$ ,  $M = 1$  TeV).



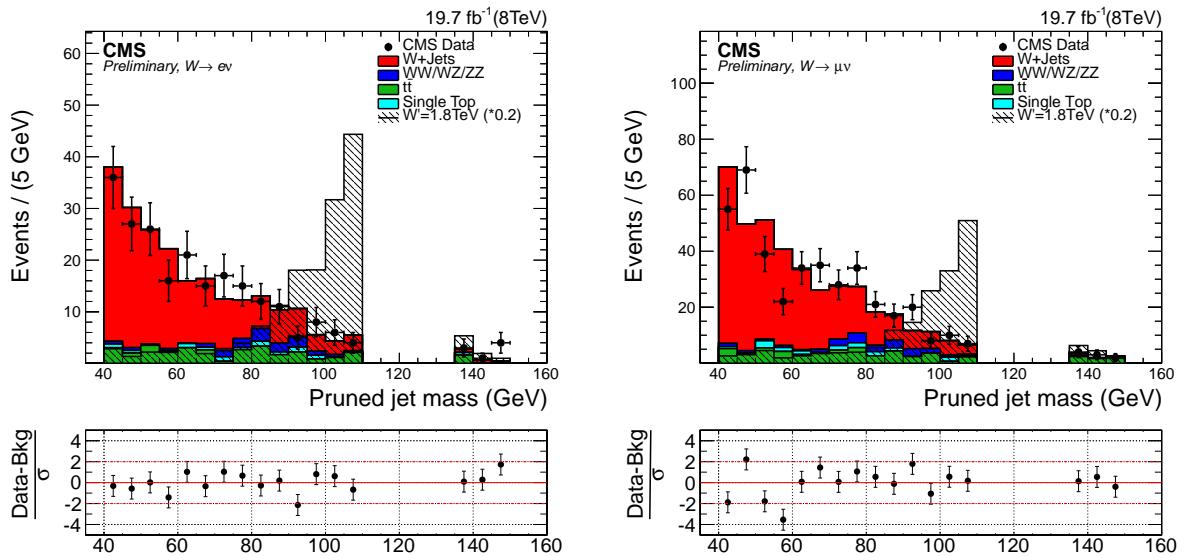
**Figure 8.3:** Leptonic  $W p_T$  and  $M_T$  for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 110$  GeV. The signal is normalized to 0.1 times of the selected events for the model considered ( $W'$ ,  $M = 1$  TeV).



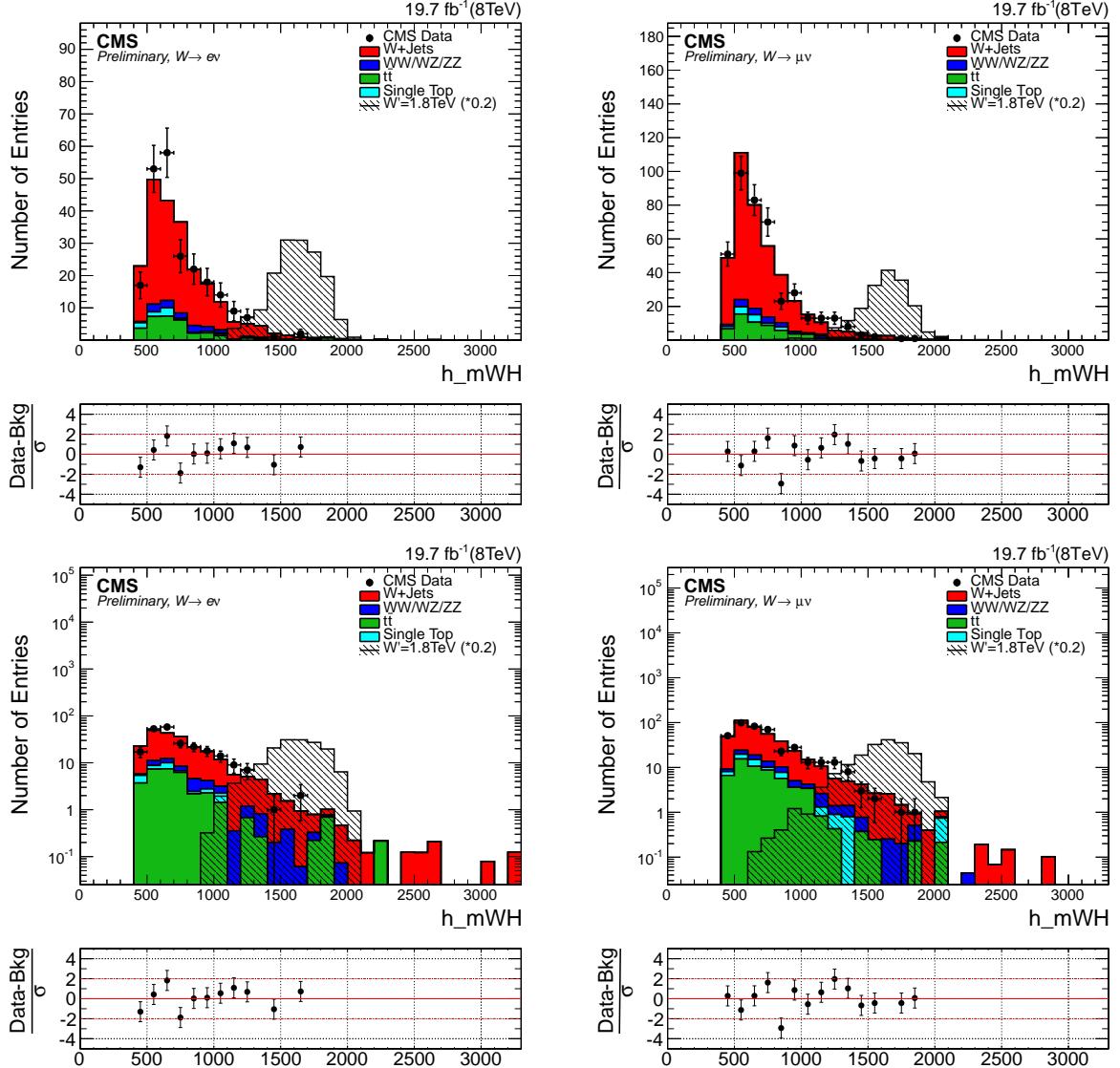
**Figure 8.4:** Hadronic  $W$   $p_T$  and  $\eta$  for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 110$  GeV. The signal is normalized to 0.1 times of the selected events for the model considered ( $W'$ ,  $M = 1\text{TeV}$ ).

**Table 8.1:** Summary of final selection.

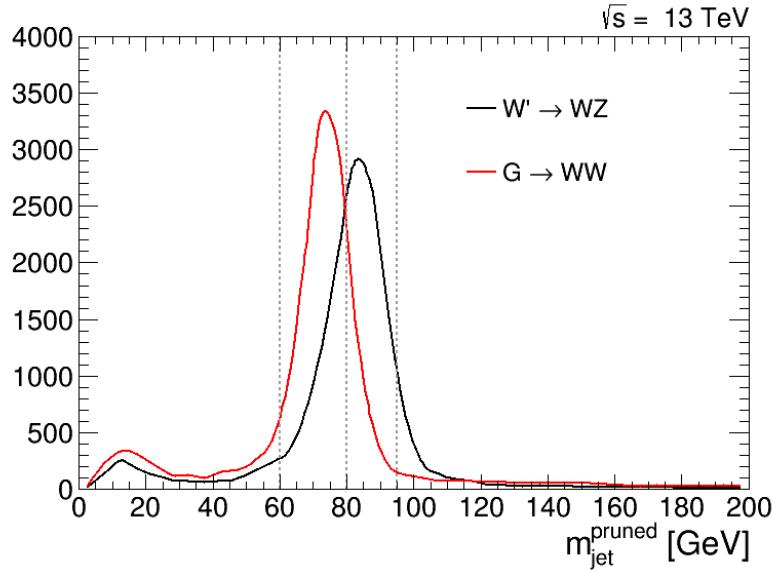
Selection	Value	Comments
<b>Tight Lepton selection</b>		
Electron $p_T$	$p_T > 90$ GeV	
Muon $p_T$	$p_T > 50$ GeV	
Electron $\eta$	$ \eta _{SC} < 2.5$ except [1.4442, 1.566] range	Avoid the ECAL gap.
Muon $\eta$	$ \eta  < 2.1$	
<b>Loose Lepton selection</b>		
Electron $p_T$	$p_T > 35$ GeV	
Muon $p_T$	$p_T > 20$ GeV	
Electron $\eta$	$ \eta _{SC} < 2.5$ except [1.4442, 1.566] range	Avoid the ECAL gap.
Muon $\eta$	$ \eta  < 2.4$	
<b>CA8 jet selections</b>		
Jet $p_T$	$p_T > 80$ GeV	Used for hadronic H reconstruction
Jet $\eta$	$ \eta  < 2.4$ except [1, 1.5] range	
<b>AK5 jet selections</b>		
Jet $p_T$	$p_T > 30$ GeV	Used for b-veto jet selection
Jet $\eta$	$ \eta  < 2.4$	
<b><math>E_T^{\text{miss}}</math> selections</b>		
$E_T^{\text{miss}}$ (electron ch.)	$E_T^{\text{miss}} > 80$ GeV	
$E_T^{\text{miss}}$ (muon ch.)	$E_T^{\text{miss}} > 40$ GeV	
<b>Boson selections</b>		
Pruned jet mass (signal)	$110 < m_{jet}^{\text{pruned}} < 135$ GeV	
Pruned jet mass (low-mass sideband)	$40 < m_{jet}^{\text{pruned}} < 110$ GeV	
Pruned jet mass (high-mass sideband)	$135 < m_{jet}^{\text{pruned}} < 150$ GeV	
Leptonic W $p_T$	$p_T > 200$ GeV	
Hadronic H $p_T$	$p_T > 200$ GeV	
Back-to-back topology	$\Delta R(\ell, H_{had}) > \pi/2$ , $\Delta\phi(H_{had}, E_T^{\text{miss}}) > 2$ $\Delta\phi(H_{had}, W_{lep}) > 2$	
B-tag veto	no CSV Medium AK5 jet within $\Delta R(\text{ca8, ak5}) = 0.8$	
Combined b-tagging cut	$\geq 2$ CSV subjets if $\Delta R > 0.3$ of subjets $\geq 1$ CSV CA8 jet if $\Delta R < 0.3$ of subjets	
Top mass veto	$m_{top}^{\text{leptonic}} < 120 \parallel m_{top}^{\text{leptonic}} > 240$ $m_{top}^{\text{hadronic}} < 160 \parallel m_{top}^{\text{hadronic}} > 280$	
<b>Veto</b>		
Number of loose electrons	0	in addition to <b>tight</b> lepton
Number of loose muons	0	in addition to <b>tight</b> lepton



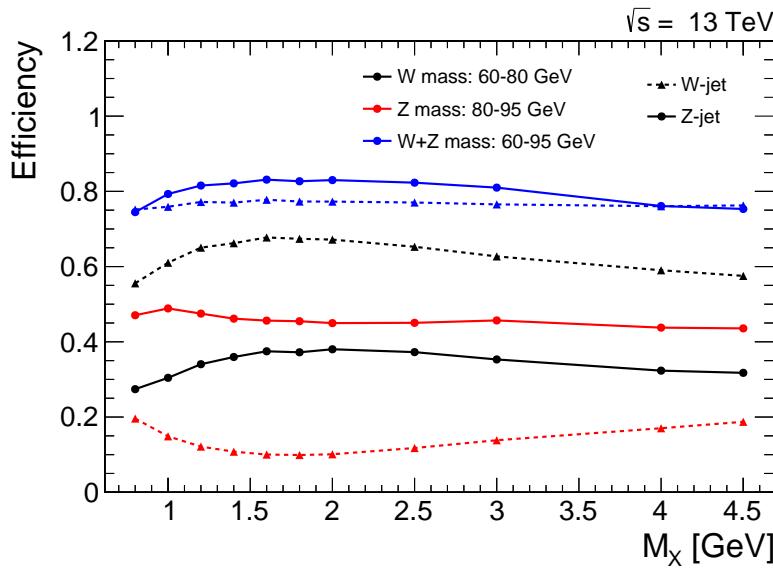
**Figure 8.5:**  $m_{\text{jet}}^{\text{pruned}}$  for electron channel (left) and muon channel (right) for events with  $40 < m_{\text{jet}}^{\text{pruned}} < 110 \text{ GeV}$ . The signal is normalized to 0.1 times of the selected events for the model considered ( $W'$ ,  $M = 1 \text{ TeV}$ ).



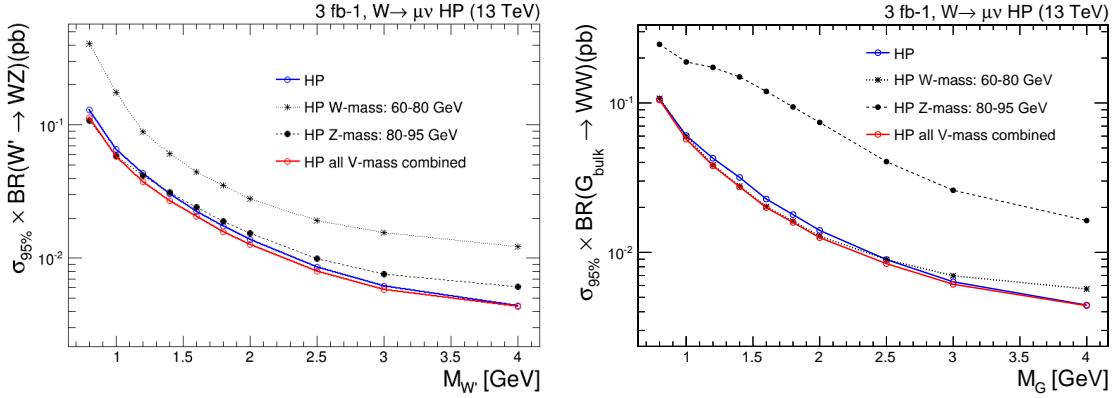
**Figure 8.6:**  $m_{WH}$  (using the  $p_{z\nu}$  defined in Section) in linear and log scale for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 110$  GeV (sideband region). The signal is normalized to 0.1 times of the selected events for the model considered ( $W'$ ,  $M = 1$  TeV).



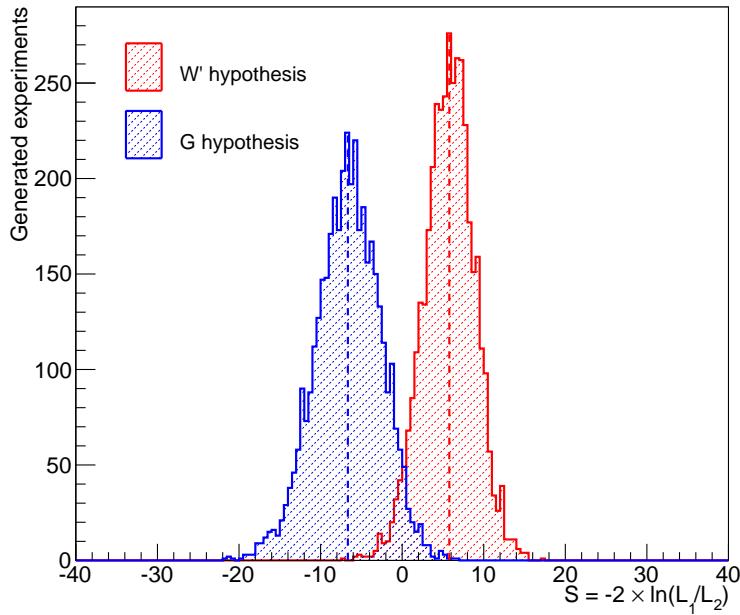
**Figure 8.7:** Pruned jet mass distributions of merged W-jets and merged Z-jets expected in  $G \rightarrow WW \rightarrow l\nu qq$  and  $W' \rightarrow WZ \rightarrow l\nu qq$  signals, respectively. The optimal separation point between the two distributions is shown together with the signal region boundaries.



**Figure 8.8:** Efficiencies of a W-jet signal ( $G \rightarrow WW$ ) (dashed lines) and of a Z-jet signal ( $W' \rightarrow WZ$ ) (solid lines) as a function of the resonance mass for different pruned jet mass windows: W-mass category (black), Z-mass category (red) and default single mass category (blue).



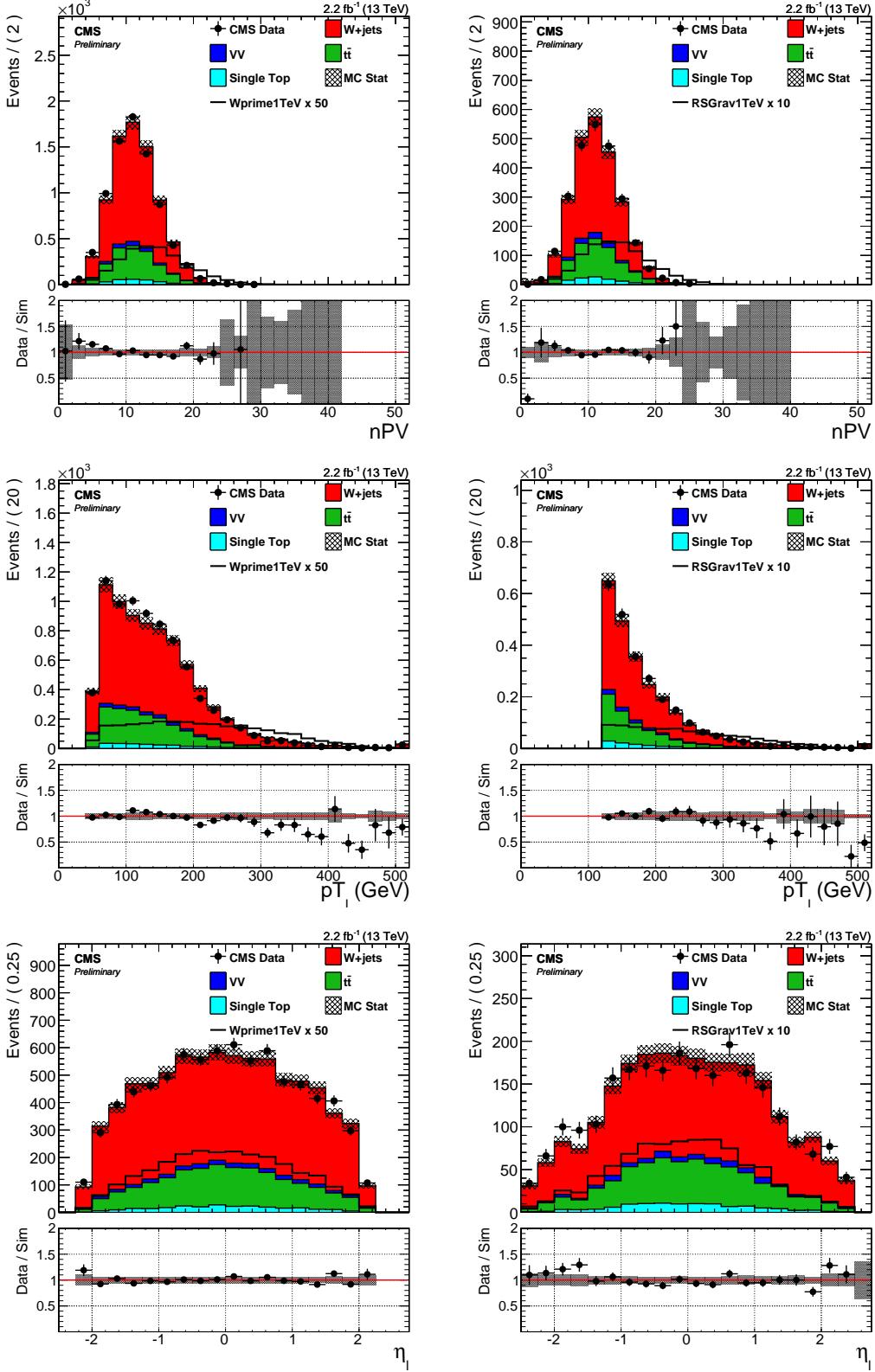
**Figure 8.9:** Expected 95% CL upper limits on the production cross section of a  $W'$  signal multiplied by the branching fraction of  $W' \rightarrow WZ$  as function of the resonance mass for the different mass categories for events passing the high-purity  $\tau_{21}$  selections. Expected 95% CL upper limits on the production cross section of a Graviton signal multiplied by the branching fraction of  $G \rightarrow WW$  as function of the resonance mass for the different mass categories for events passing the high-purity  $\tau_{21}$  selections.



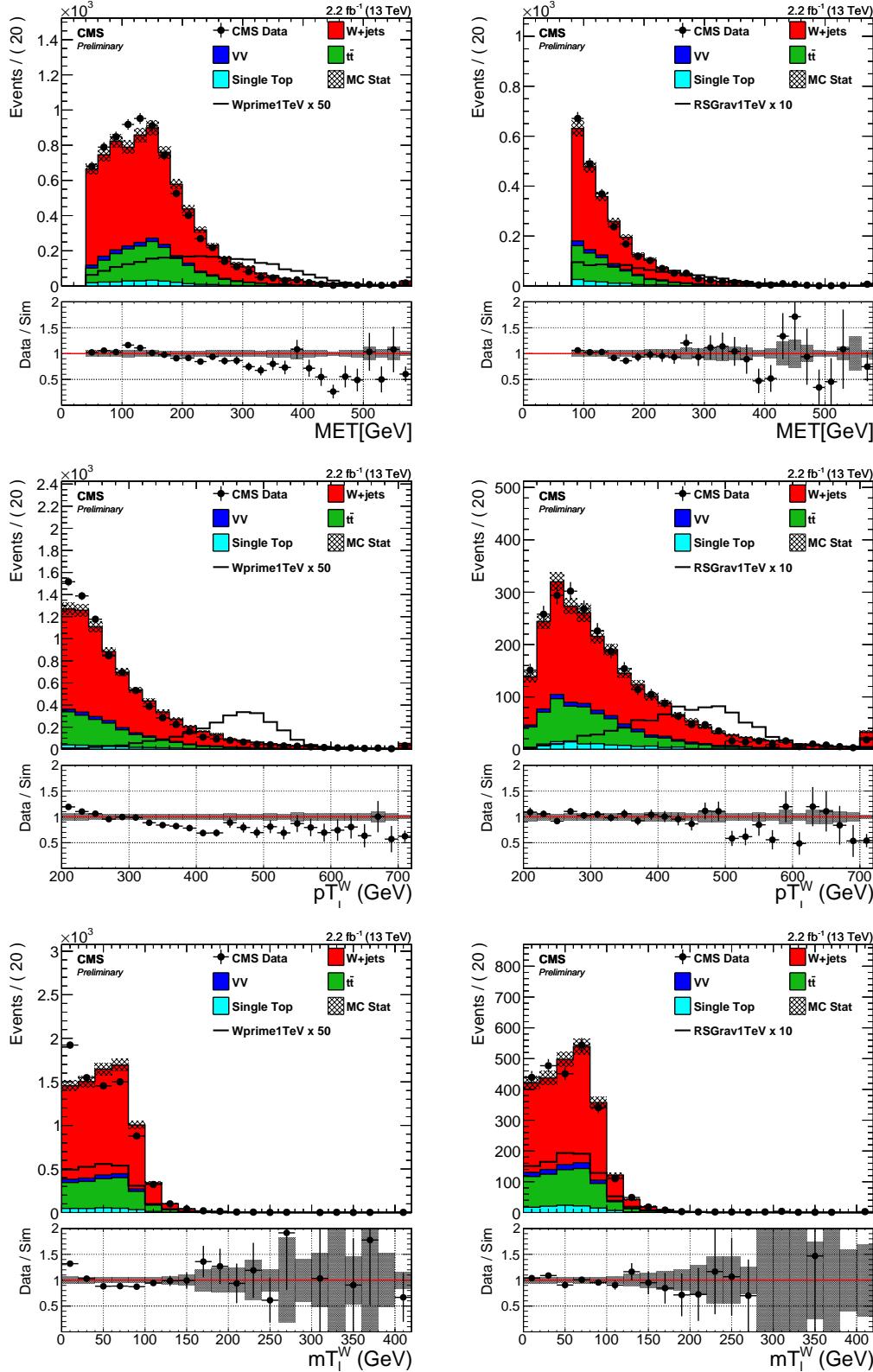
**Figure 8.10:** Distribution of the test statistic  $q = -2\ln(L_G/L_{W'})$  of the Graviton signal hypothesis (blue) tested against the  $W'$  signal hypothesis (red).

**Table 8.2:** Summary of the VW channel final selection.

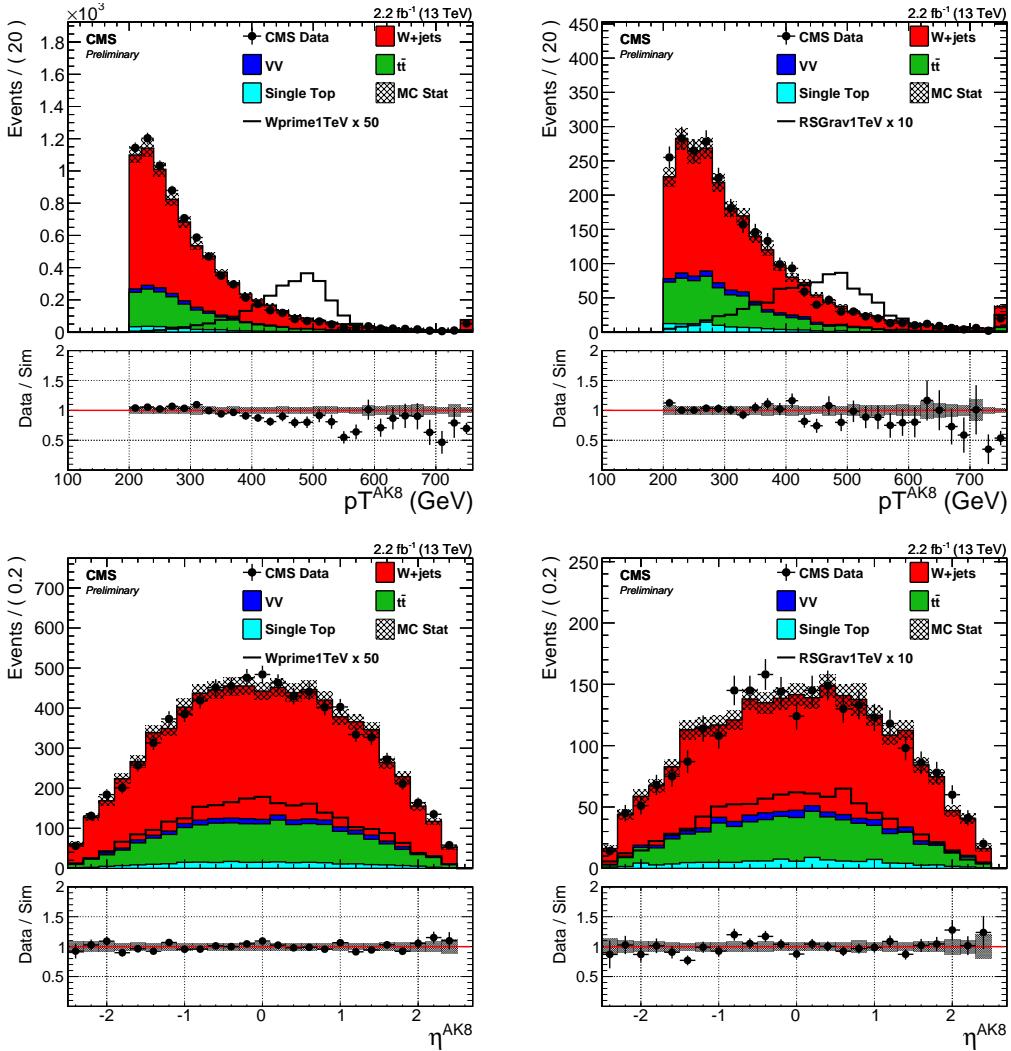
Selection	Value	Comments
<b>Tight Lepton selection</b>		
Electron $p_T$	$p_T > 120 \text{ GeV}$	
Muon $p_T$	$p_T > 53 \text{ GeV}$	
Electron $\eta$	$ \eta _{\text{SC}} < 2.5$ except $[1.4442, 1.566]$ range	Avoid the ECAL gap.
Muon $\eta$	$ \eta  < 2.1$	
<b>Loose Lepton selection</b>		
Electron $p_T$	$p_T > 35 \text{ GeV}$	
Muon $p_T$	$p_T > 20 \text{ GeV}$	
Electron $\eta$	$ \eta _{\text{SC}} < 2.5$ except $[1.4442, 1.566]$ range	Avoid the ECAL gap.
Muon $\eta$	$ \eta  < 2.4$	
<b>AK8 jet selections</b>		
Jet $p_T$	$p_T > 200 \text{ GeV}$	Used for hadronic W reconstruction
Jet $\eta$	$ \eta  < 2.4$	
<b>AK4 jet selections</b>		
Jet $p_T$	$p_T > 30 \text{ GeV}$	Used for b-tag jet selection
Jet $\eta$	$ \eta  < 2.4$	
<b><math>E_T^{\text{miss}}</math> selections</b>		
$E_T^{\text{miss}}$ (electron ch.)	$E_T^{\text{miss}} > 80 \text{ GeV}$	
$E_T^{\text{miss}}$ (muon ch.)	$E_T^{\text{miss}} > 40 \text{ GeV}$	
<b>Boson selections</b>		
Pruned jet mass (signal)	$65 < m_{\text{jet}}^{\text{pruned}} < 105 \text{ GeV}$	
Pruned jet mass (low-mass sideband)	$40 < m_{\text{jet}}^{\text{pruned}} < 65 \text{ GeV}$	
Pruned jet mass (high-mass sideband)	$105 < m_{\text{jet}}^{\text{pruned}} < 135 \text{ GeV}$	
Leptonic W $p_T$	$p_T > 200 \text{ GeV}$	
Hadronic W $p_T$	$p_T > 200 \text{ GeV}$	
Back-to-back topology	$\Delta R(\ell, W_{\text{had}}) > \pi/2$ , $\Delta\phi(W_{\text{had}}, E_T^{\text{miss}}) > 2$ $\Delta\phi(W_{\text{had}}, W_{\text{lep}}) > 2$	
<b>Veto</b>		
Number of <b>loose</b> electrons	0	in addition to <b>tight</b> lepton
Number of <b>loose</b> muons	0	in addition to <b>tight</b> lepton
Number of b-tagged jets	0	PF iCSV medium working point
<b>Diboson selections</b>		
2- to 1-subjettiness ratio (high purity)	$\tau_{21} < 0.60$	
2- to 1-subjettiness ratio (low purity)	$0.60 \leq \tau_{21} < 0.75$	



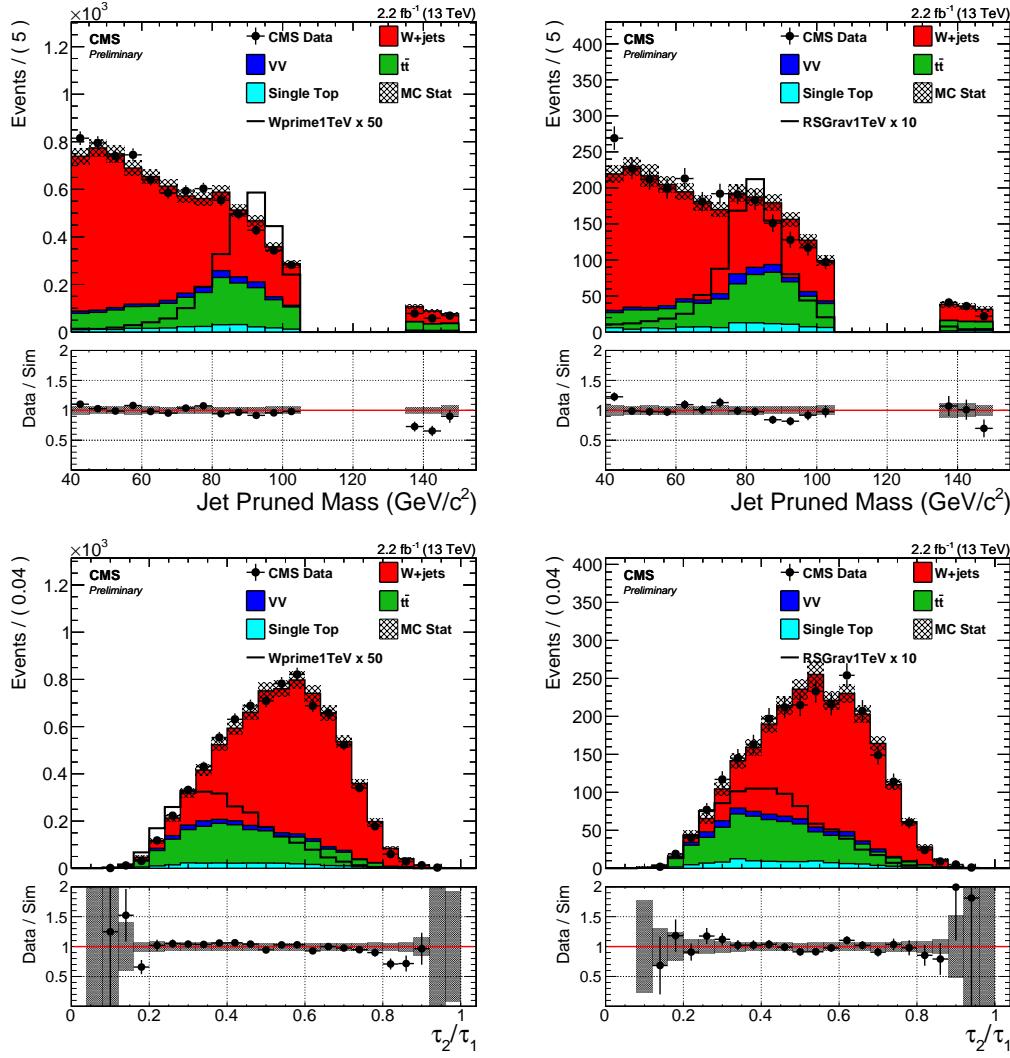
**Figure 8.11:** Comparison plots between data and MC for different observables, in the jet mass sideband. From top to bottom: number of primary vertices, lepton  $p_T$ , lepton  $\eta$ . Left: muon channel, right: electron channel.



**Figure 8.12:** Comparison plots between data and MC for different observables, in the jet mass sideband. From top to bottom:  $E_T^{\text{miss}}$ , leptonic  $W$   $p_T$ , transverse mass of the leptonic  $W$ . Left: muon channel, right: electron channel.



**Figure 8.13:** Comparison plots between data and MC for different observables, in the jet mass sideband. Top:  $p_T$  of the leading AK8 jet. Bottom:  $\eta$  of the leading AK8 jet. Left: muon channel, right: electron channel.



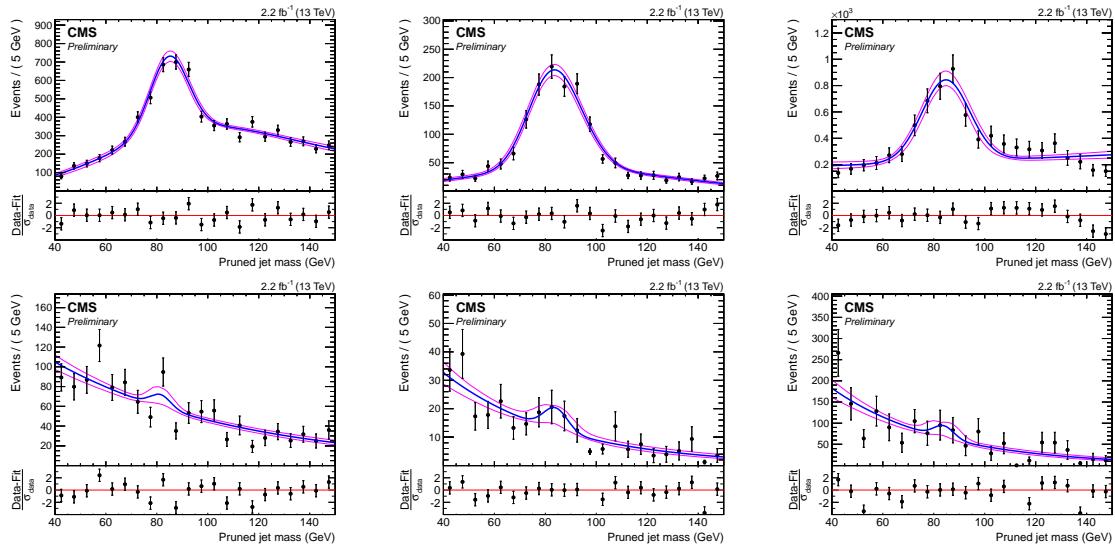
**Figure 8.14:** Comparison plots between data and MC for different observables, in the jet mass sideband. Top: pruned jet mass. Bottom: N-subjettiness. Left: muon channel, right: electron channel.

# Background modeling

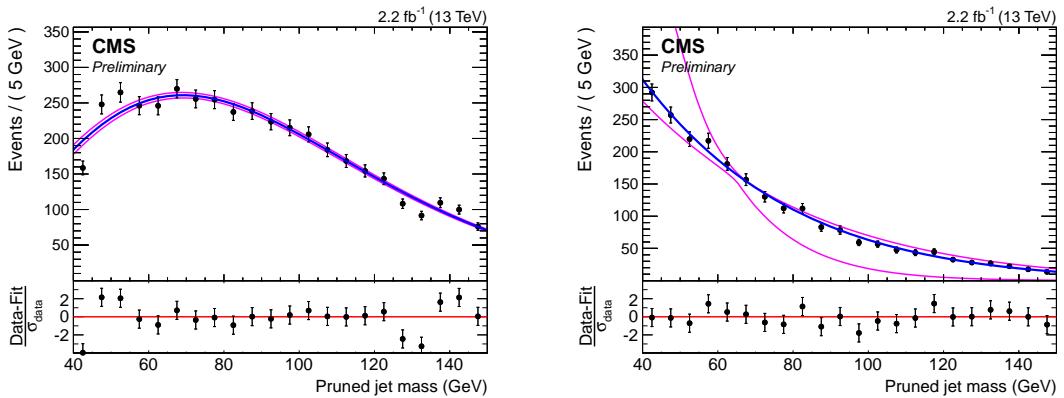
## 9.1 W+jets background estimate with alpha method

### 9.1.1 Description

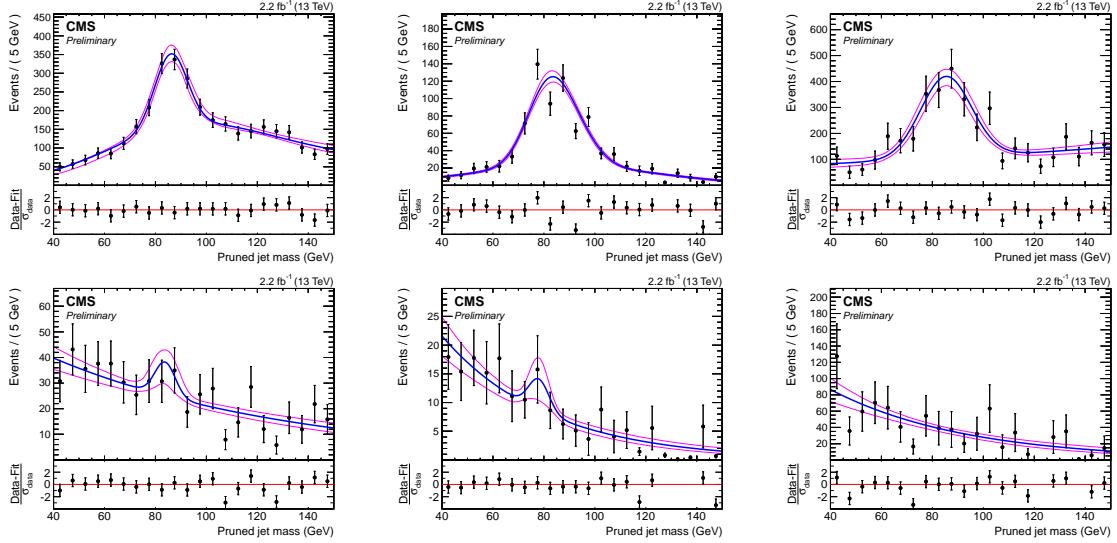
### 9.1.2 Extraction of the W+jets normalization



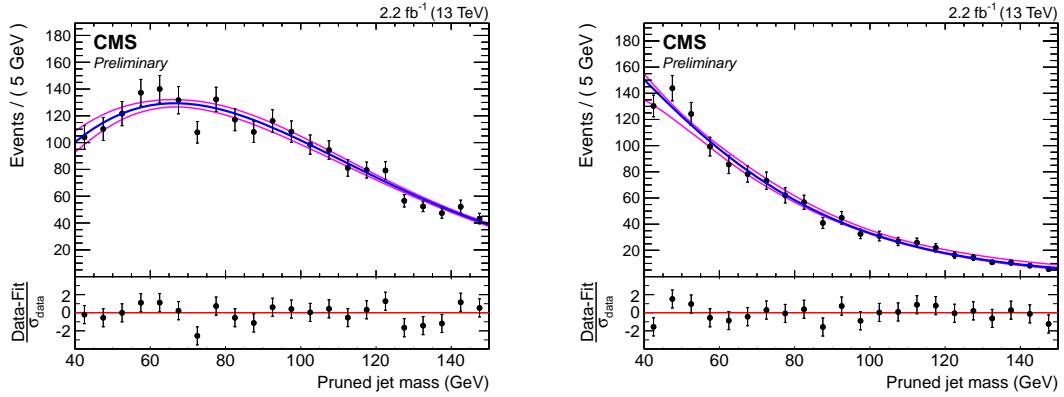
**Figure 9.1:** MC fits of non-dominant background  $m_{\text{jet}}$  spectra: on top (bottom) high purity (low purity) categories for the muon channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



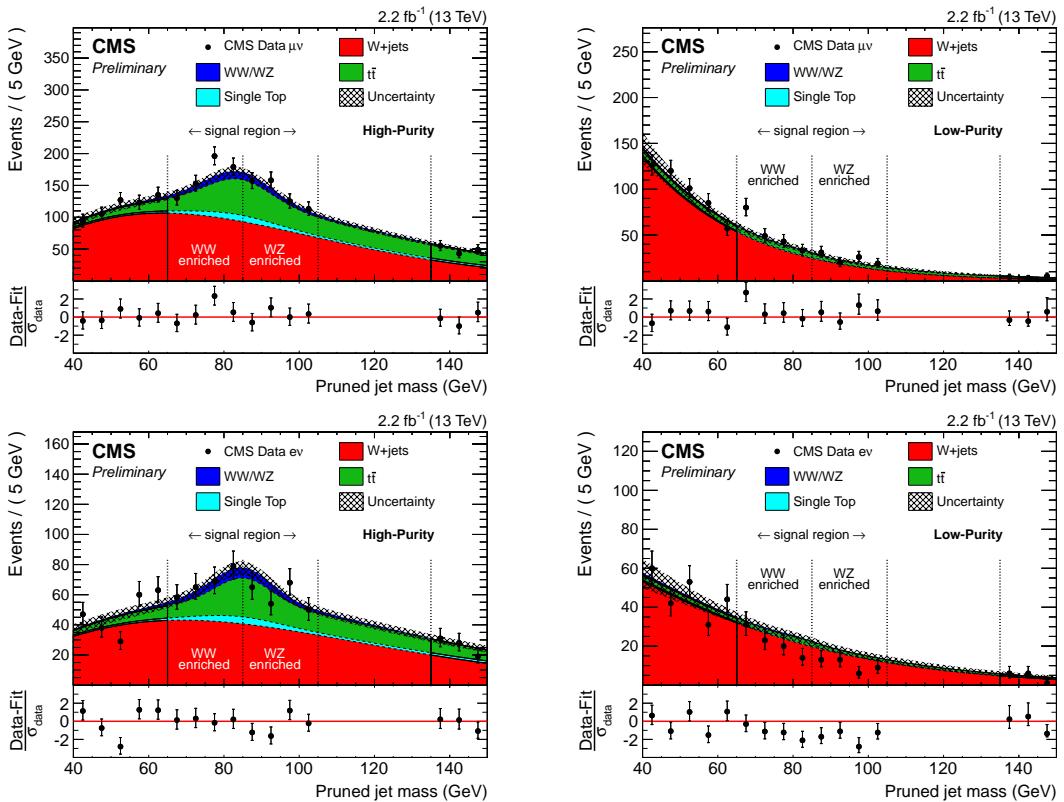
**Figure 9.2:** MC fits of dominant W+jets background  $m_{\text{jet}}$  spectra: high purity (left) and low purity (right) category for the muon channel.



**Figure 9.3:** MC fits of non-dominant background  $m_{\text{jet}}$  spectra: on top (bottom) high purity (low purity) categories for the electron channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.

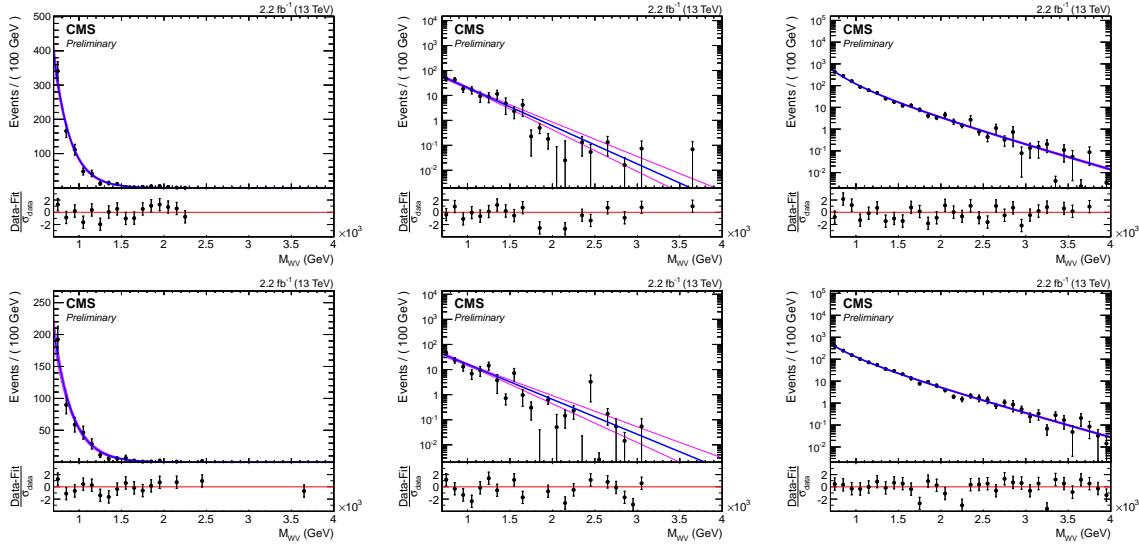


**Figure 9.4:** MC fits of dominant  $W + \text{jets}$  background  $m_{\text{jet}}$  spectra: high purity (left) and low purity (right) category for the electron channel.

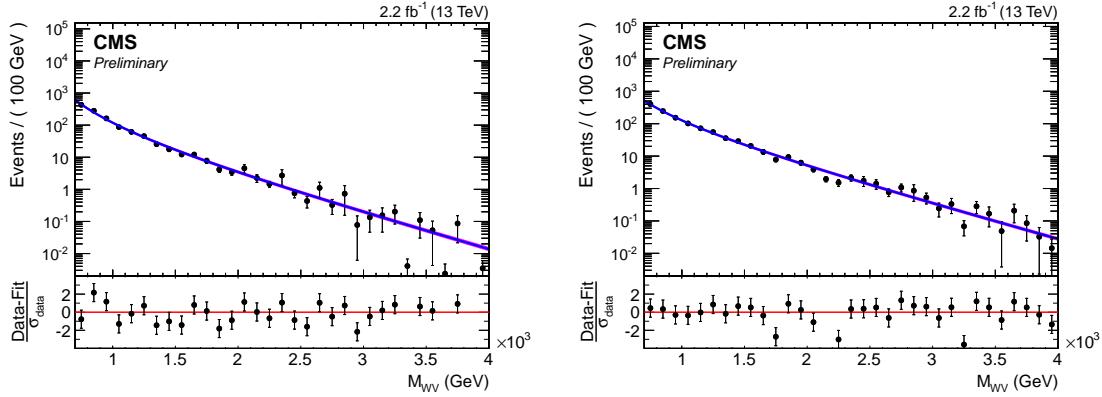


**Figure 9.5:** Fits to extract the relative shape and normalization of the W+jets contribution from the data in the jet mass distribution. Top line fits for the muon channel: high purity (left) and low purity (right) category. Bottom line fits for the electron channel: high purity (left) and low purity (right) category. The hashed area denotes the fit uncertainty, the shaded area the blinded W/Z/H signal region, and the vertical dashed lines separate the W, Z and H window from left to right.

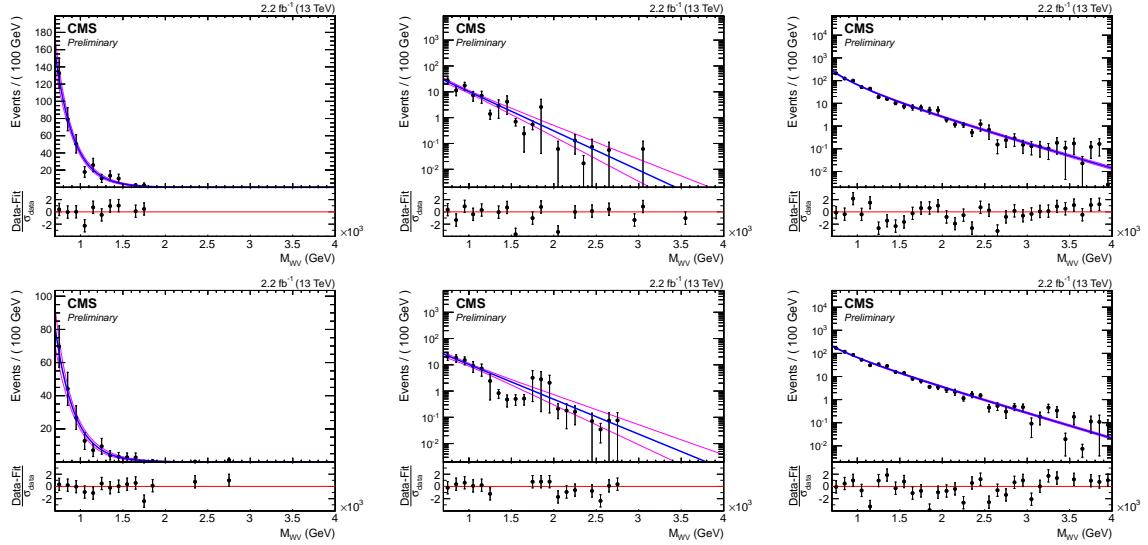
### 9.1.3 Extraction of the W+jets shape



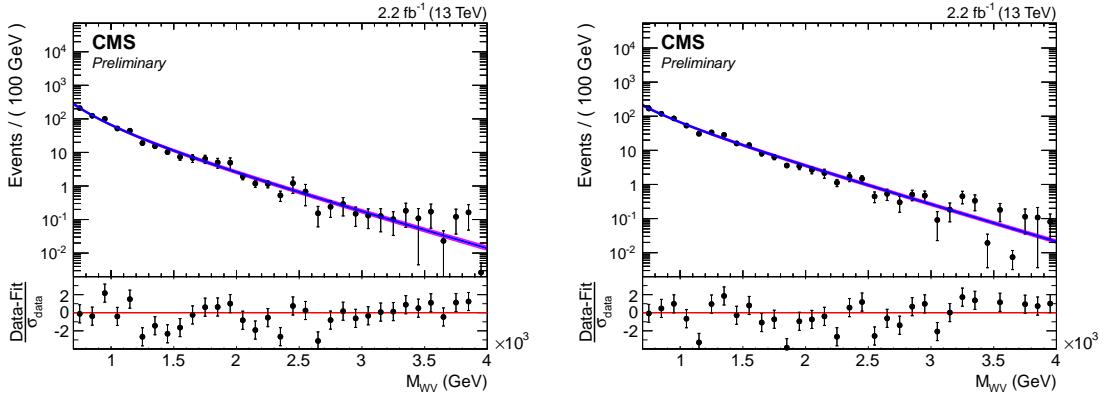
**Figure 9.6:** MC fits of non-dominant background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  sideband: on top (bottom) high purity (low purity) categories for the muon channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



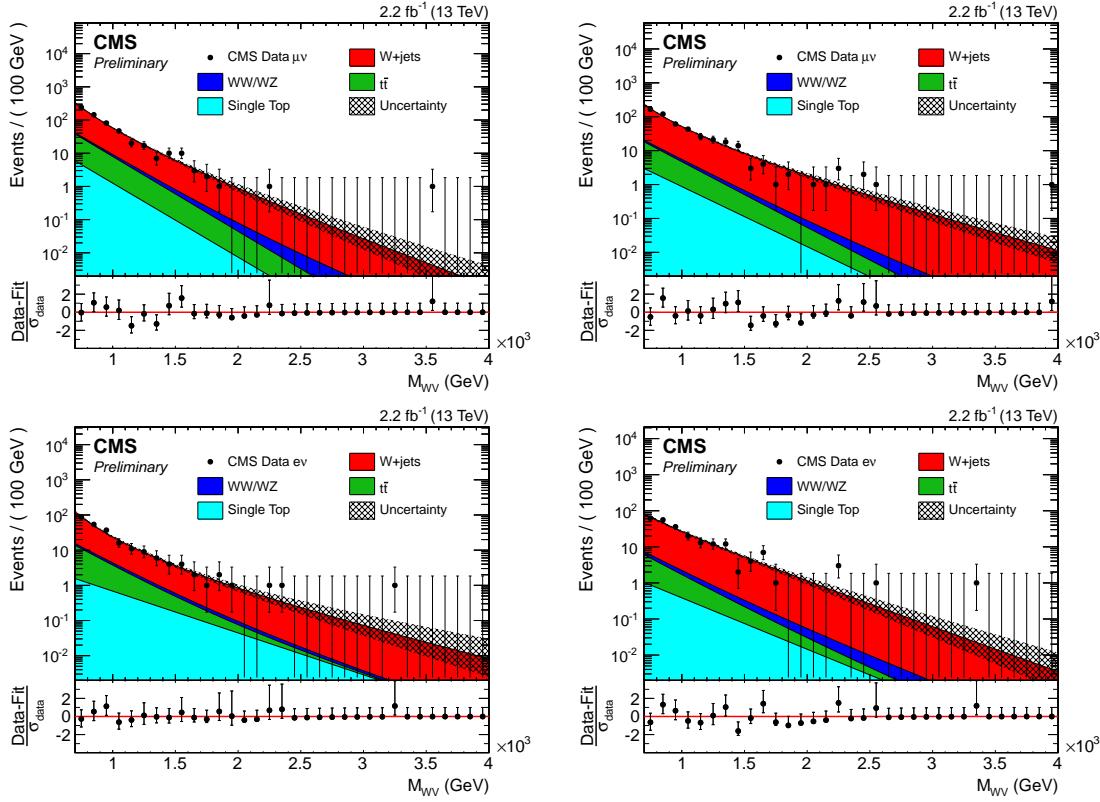
**Figure 9.7:** MC fits of dominant W+jets background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  sideband: high purity (left) and low purity (right) category for the muon channel.



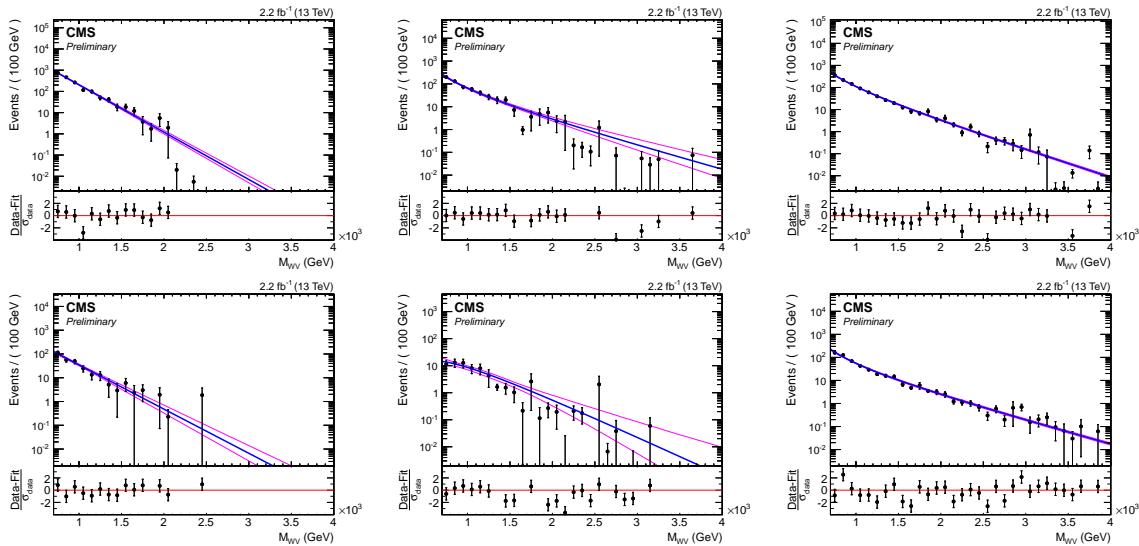
**Figure 9.8:** MC fits of non-dominant background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  sideband: on top (bottom) high purity (low purity) categories for the electron channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



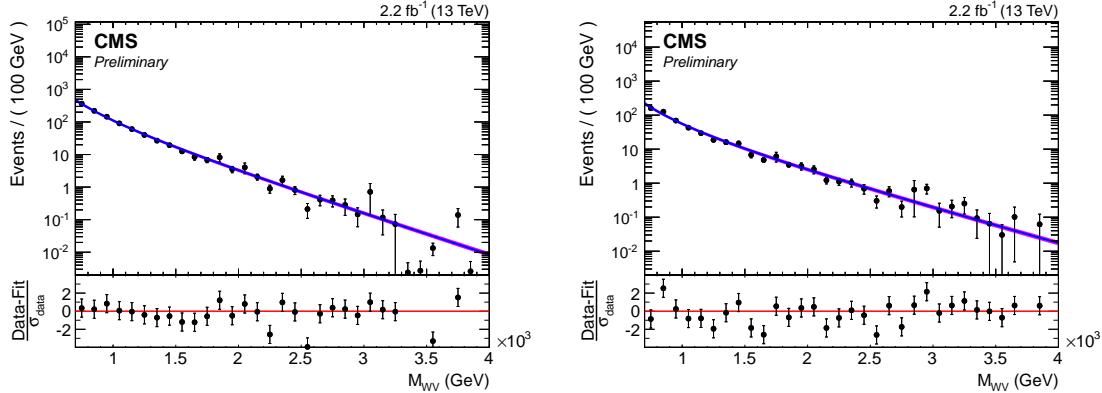
**Figure 9.9:** MC fits of dominant W+jets background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  sideband: high purity (left) and low purity (right) category for the electron channel.



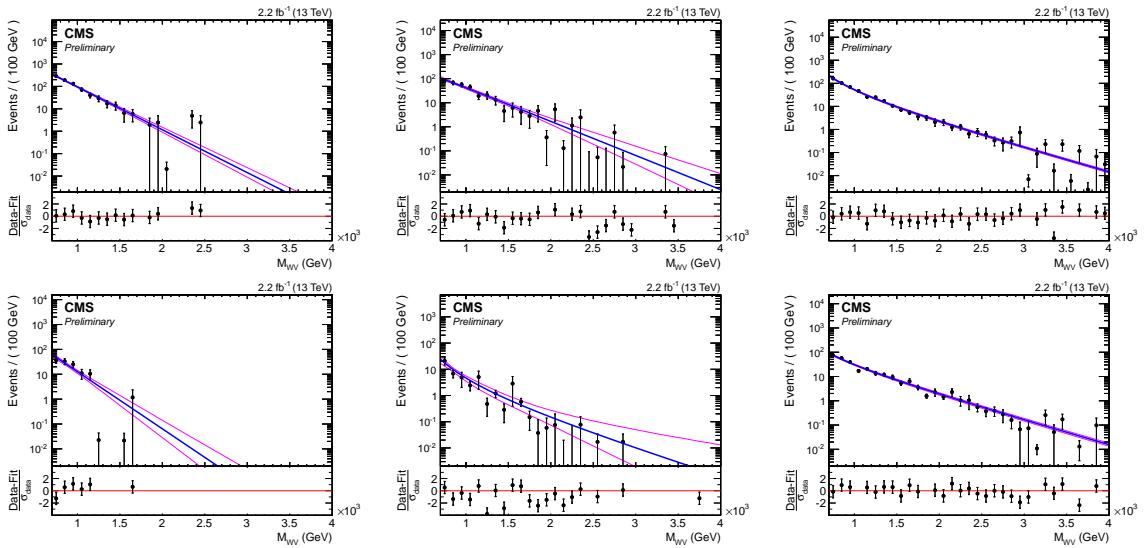
**Figure 9.10:** The fits for  $F_{data,LSB}(m_{\ell\nu j})$  for both electron (bottom) and muon (top) channels, high purity (left) and low purity (right) categories.



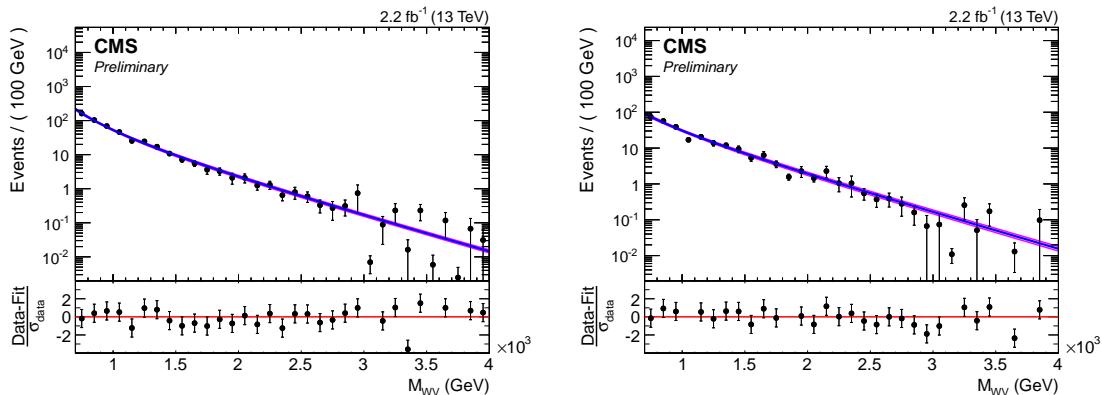
**Figure 9.11:** MC fits of non-dominant background  $m_{\ell\nu j}$  spectra in the  $m_{jet}$  signal region for events in the WW category: on top (bottom) high purity (low purity) categories for the muon channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



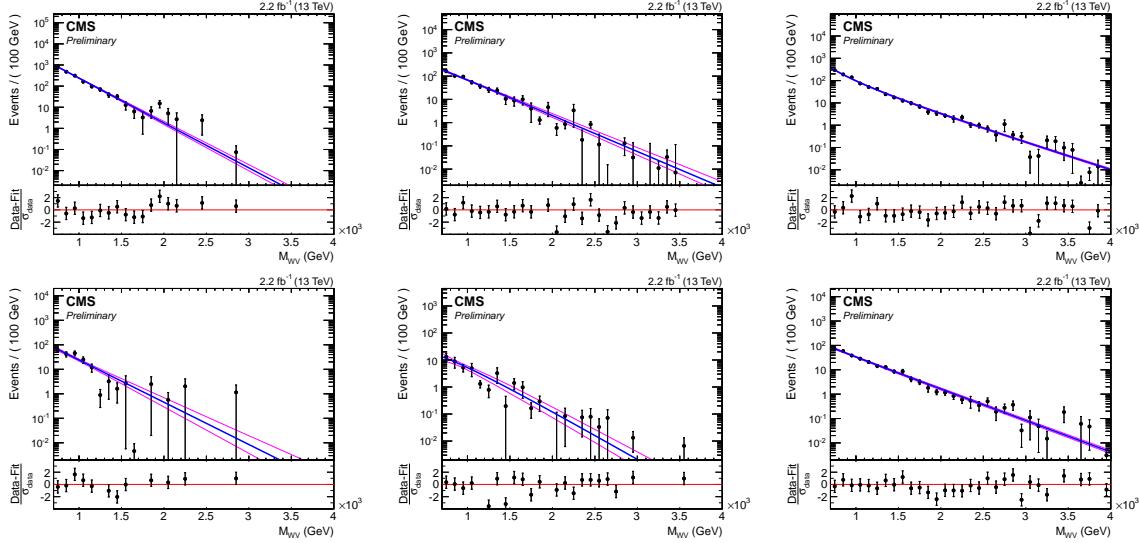
**Figure 9.12:** MC fits of dominant W+jets background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  signal region for events in the WW category: high purity (left) and low purity (right) category for the muon channel.



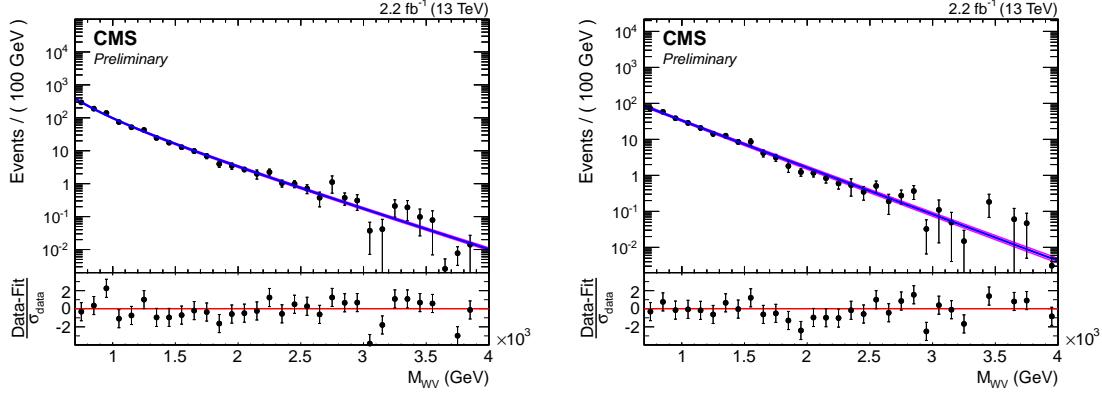
**Figure 9.13:** MC fits of non-dominant background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  signal region for events in the WW category: on top (bottom) high purity (low purity) categories for the electron channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



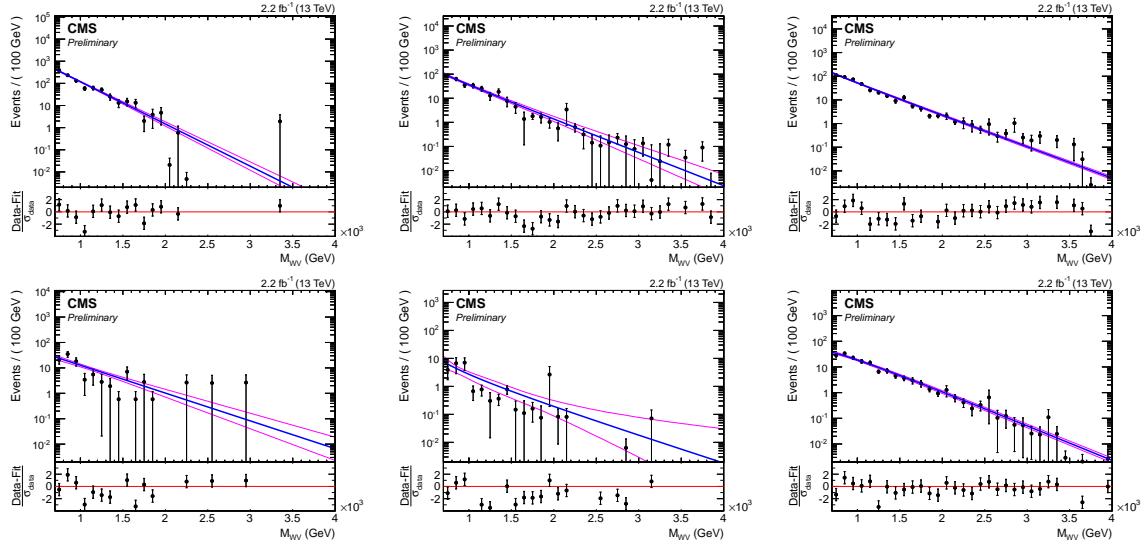
**Figure 9.14:** MC fits of dominant W+jets background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  signal region for events in the WW category: high purity (left) and low purity (right) category for the electron channel.



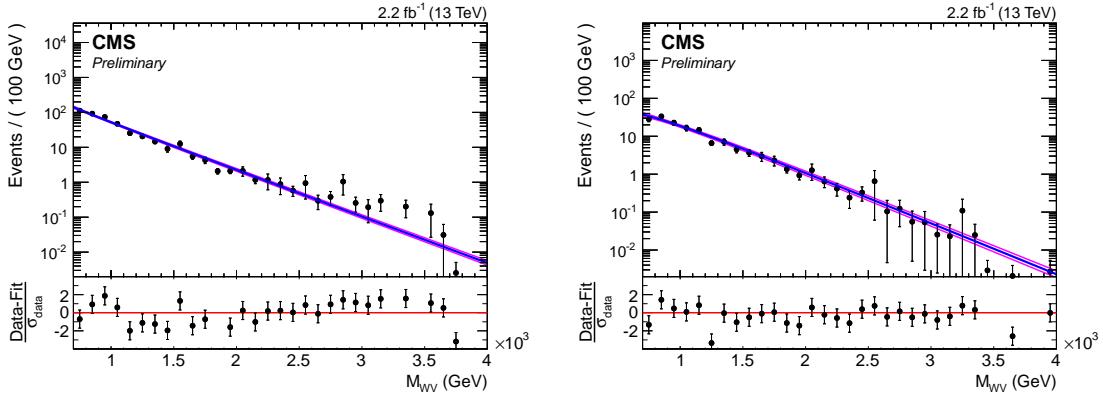
**Figure 9.15:** MC fits of non-dominant background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  signal region for events in the WZ category: on top (bottom) high purity (low purity) categories for the muon channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



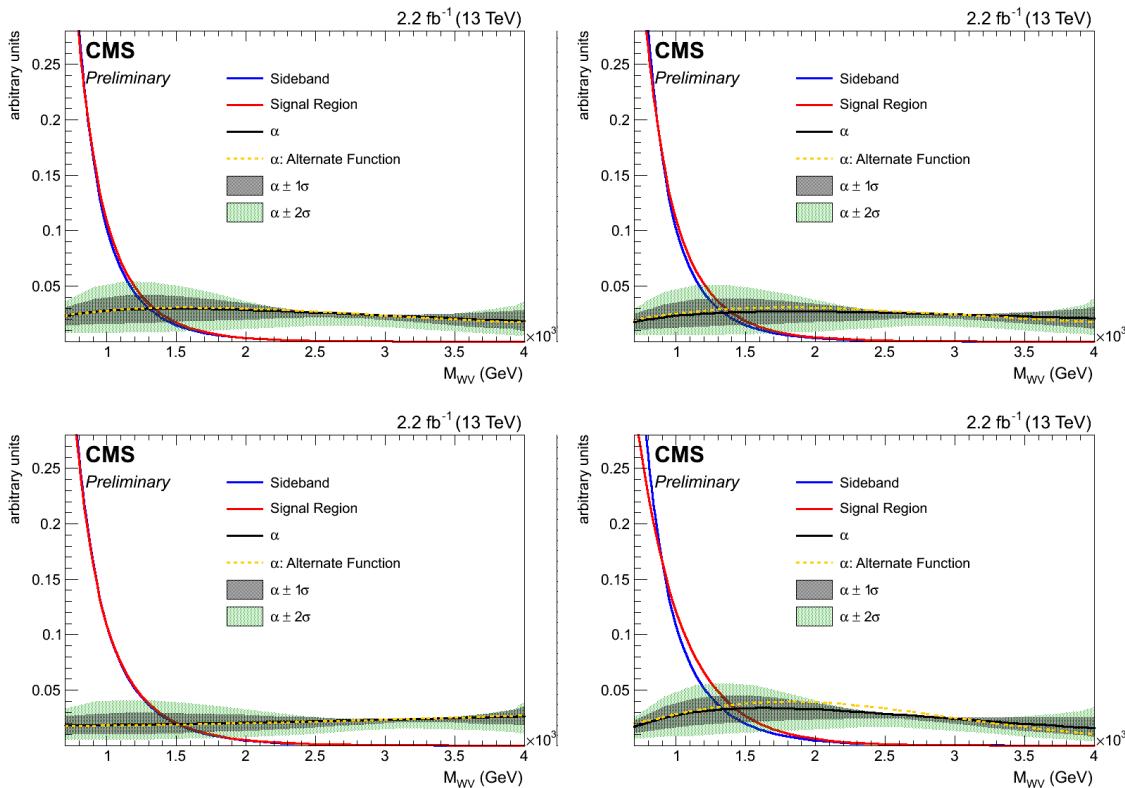
**Figure 9.16:** MC fits of dominant W+jets background  $m_{\ell\nu j}$  spectra in the  $m_{\text{jet}}$  signal region for events in the WZ category: high purity (left) and low purity (right) category for the muon channel.



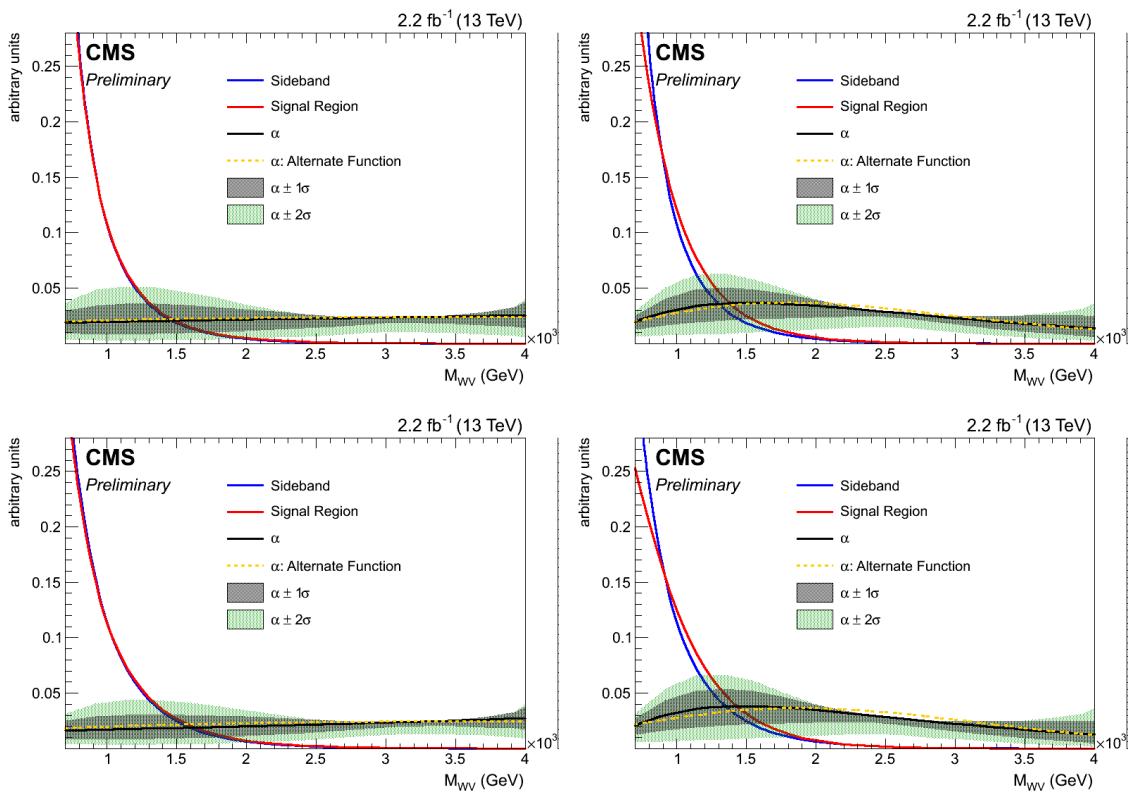
**Figure 9.17:** MC fits of non-dominant background  $m_{\ell\nu j}$  spectra in the  $m_{jet}$  signal region for events in the WZ category: on top (bottom) high purity (low purity) categories for the electron channel. Left to right are the  $t\bar{t}$ , diboson (WW/WZ/ZZ) and Single Top processes.



**Figure 9.18:** MC fits of dominant W+jets background  $m_{\ell\nu j}$  spectra in the  $m_{jet}$  signal region for events in the WZ category: high purity (left) and low purity (right) category for the electron channel.



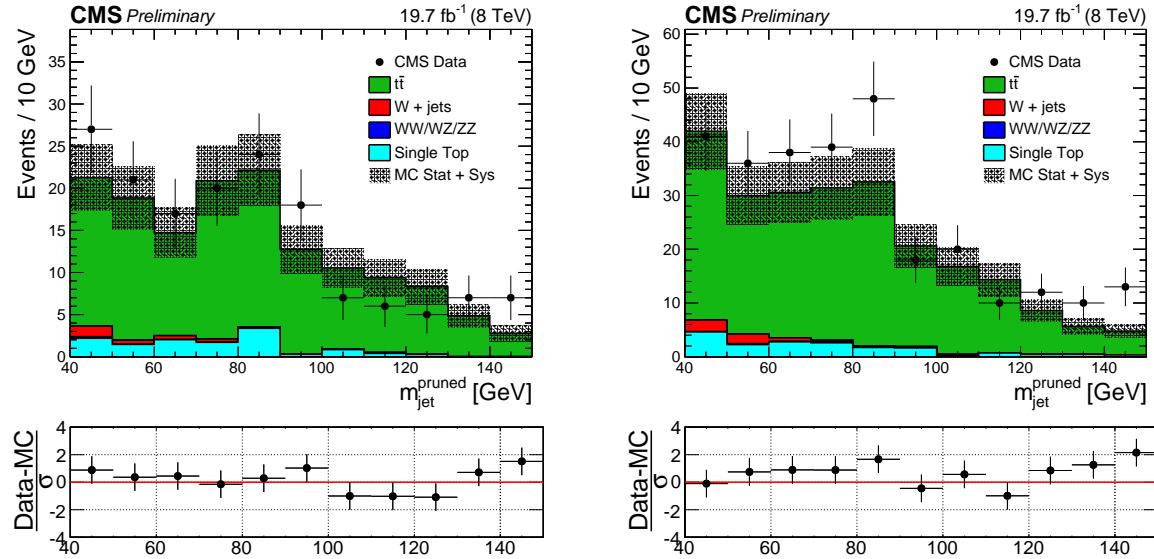
**Figure 9.19:** The functions  $\alpha_{MC}(m_{\ell\nu j})$  for the muon high purity category (top), low purity (bottom) categories used for extrapolate W+jets  $m_{\ell\nu j}$  shape in the WW (left) and WZ (right) signal regions.



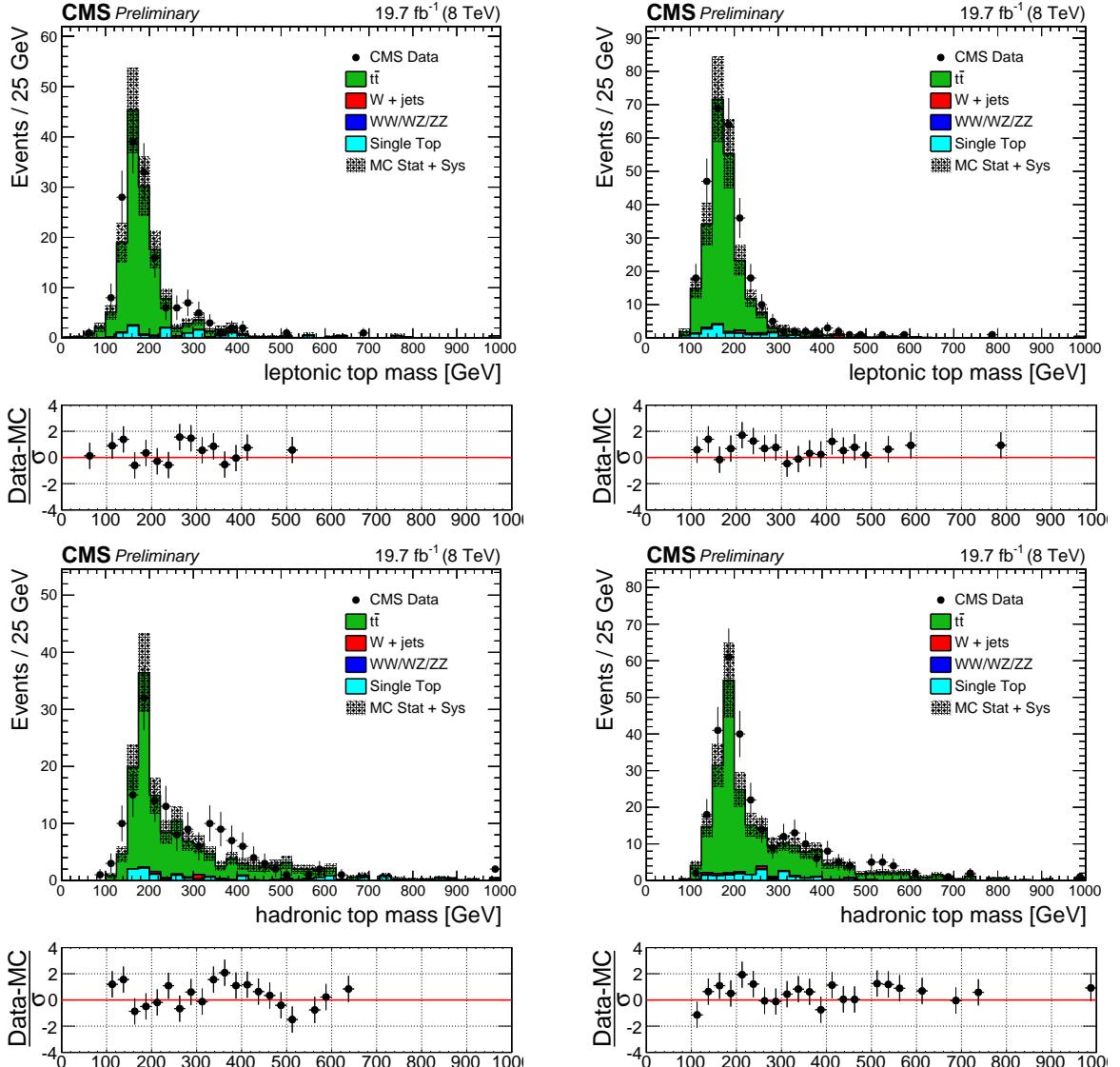
**Figure 9.20:** The functions  $\alpha_{MC}(m_{\ell\nu j})$  for the electron high purity category (top), low purity (bottom) categories used for extrapolate W+jets  $m_{\ell\nu j}$  shape in the WW (left) and WZ (right) signal regions.

## 9.2 Top quark production

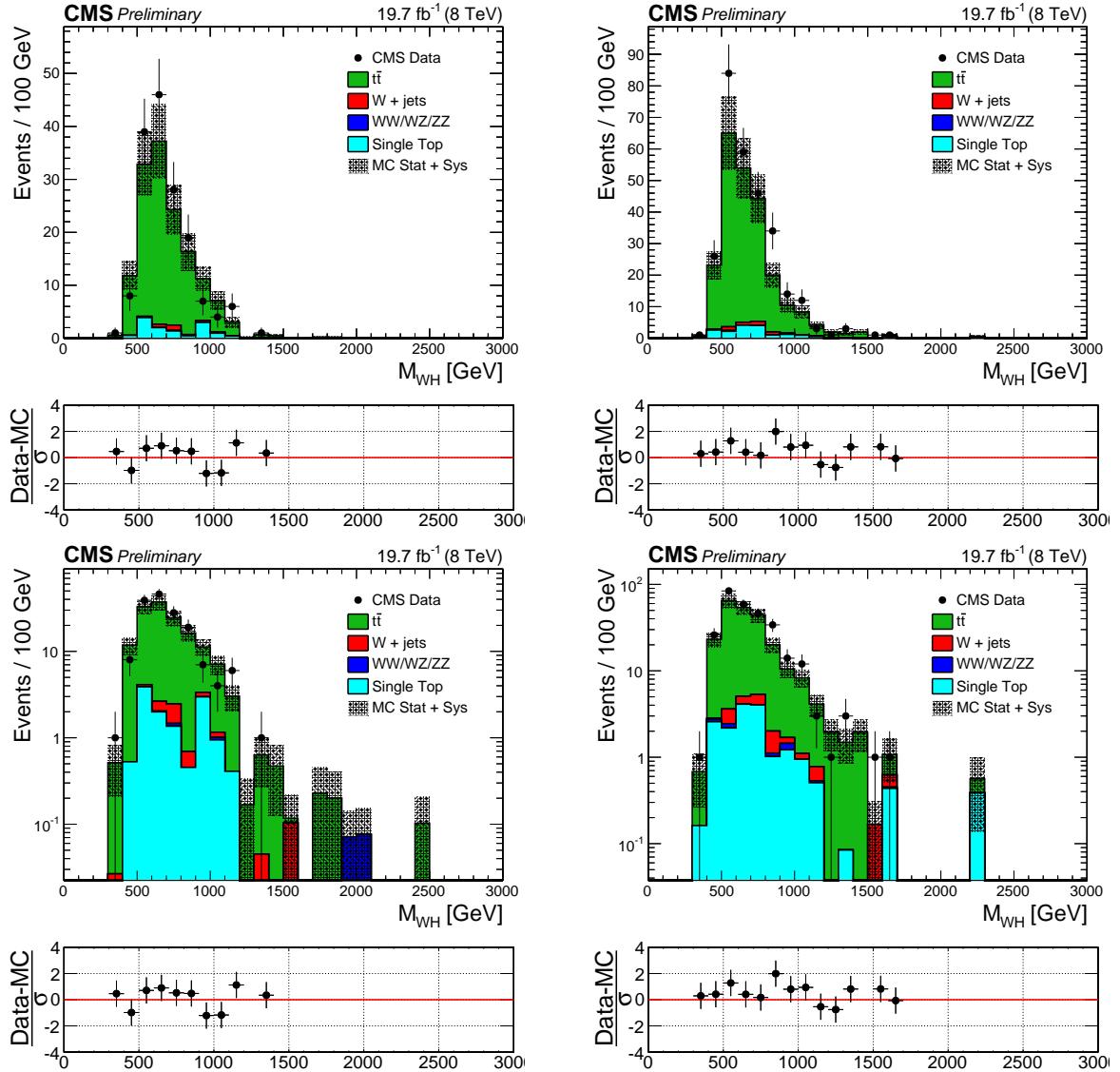
### 8 TeV analysis



**Figure 9.21:** Pruned jet mass for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 150$  GeV in the  $t\bar{t}$  control sample.

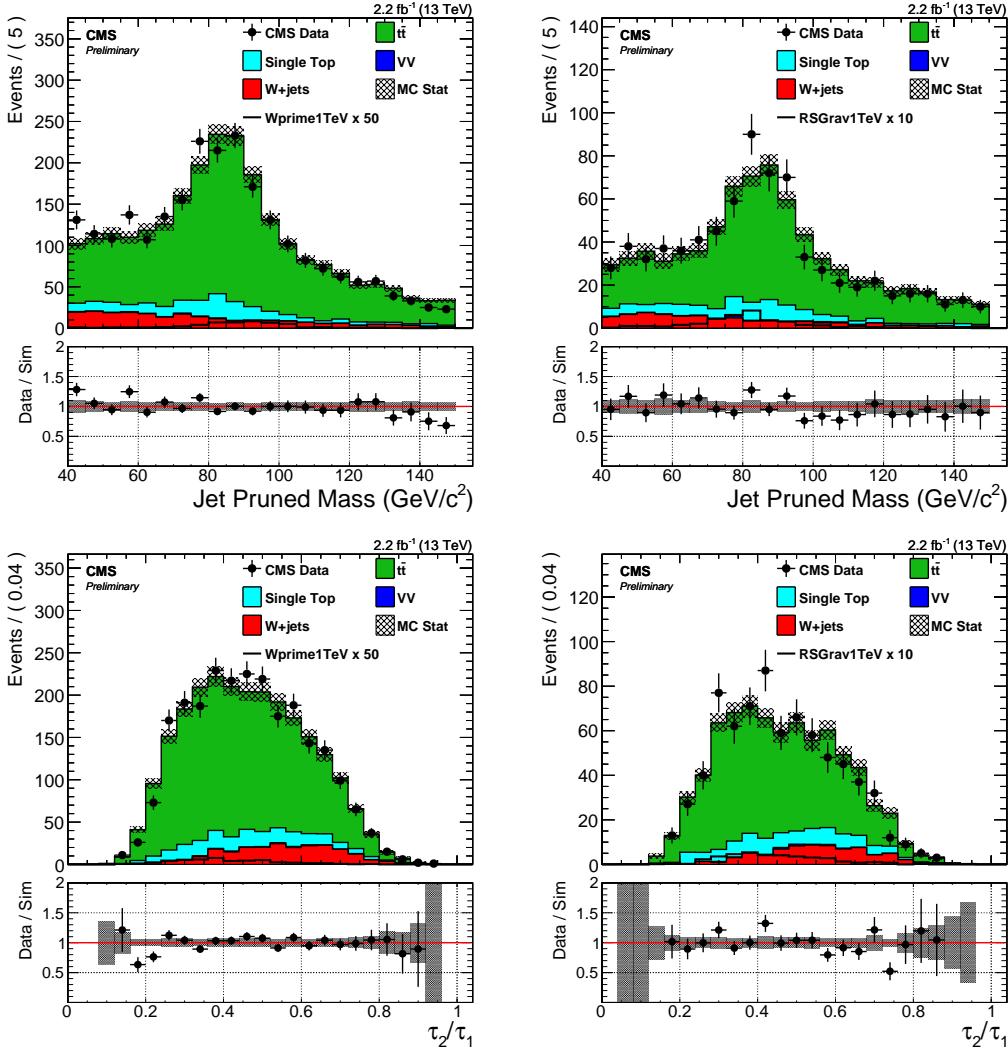


**Figure 9.22:**  $m_{top}^{leptonic}$  (top) and  $m_{top}^{hadronic}$  (bottom) for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 150$  GeV in the  $t\bar{t}$  control sample.

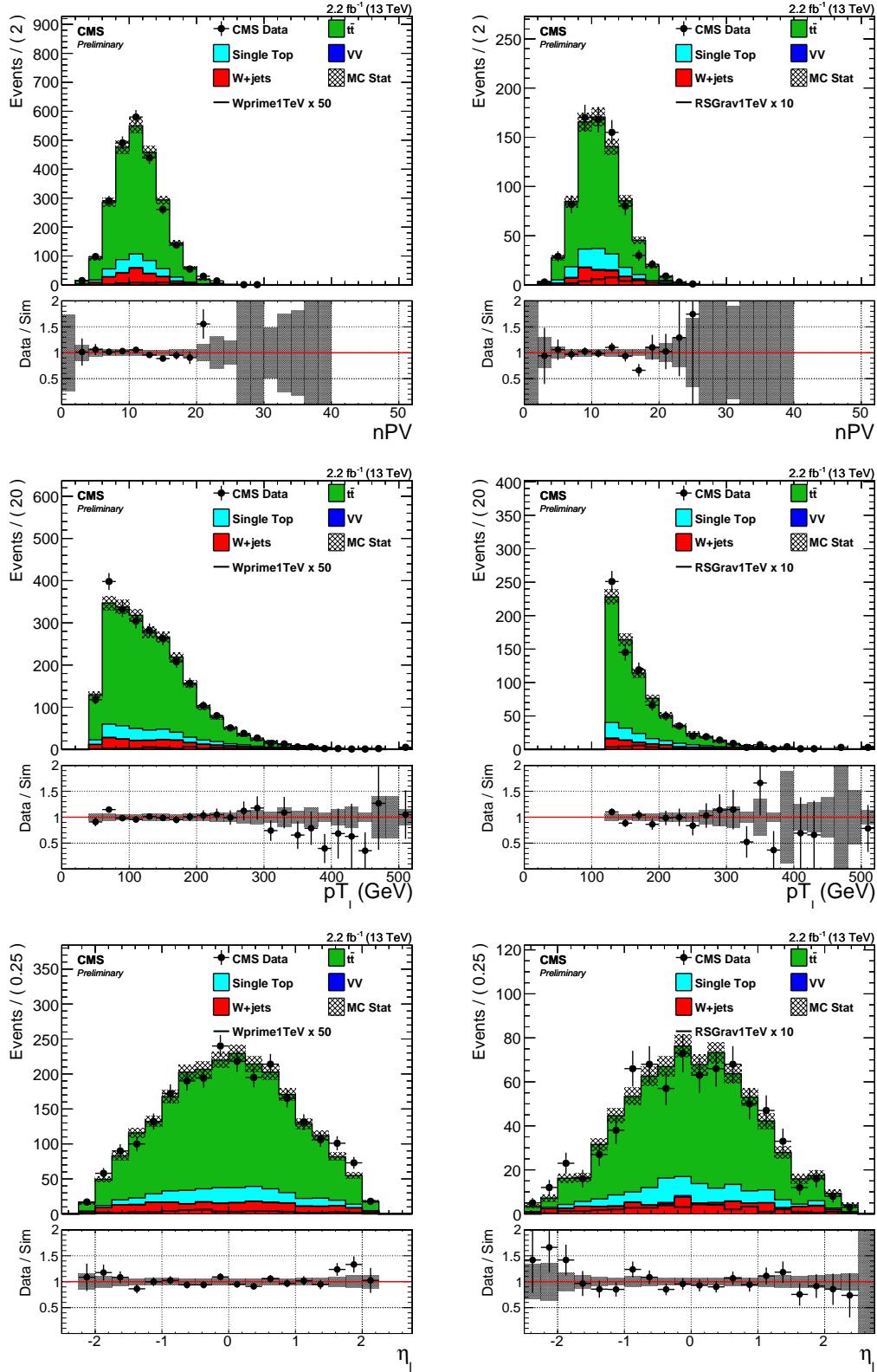


**Figure 9.23:**  $m_{WH}$  in linear (top) and log (bottom) scale for electron channel (left) and muon channel (right) for events with  $40 < m_{jet}^{pruned} < 150$  GeV in the  $t\bar{t}$  control sample.

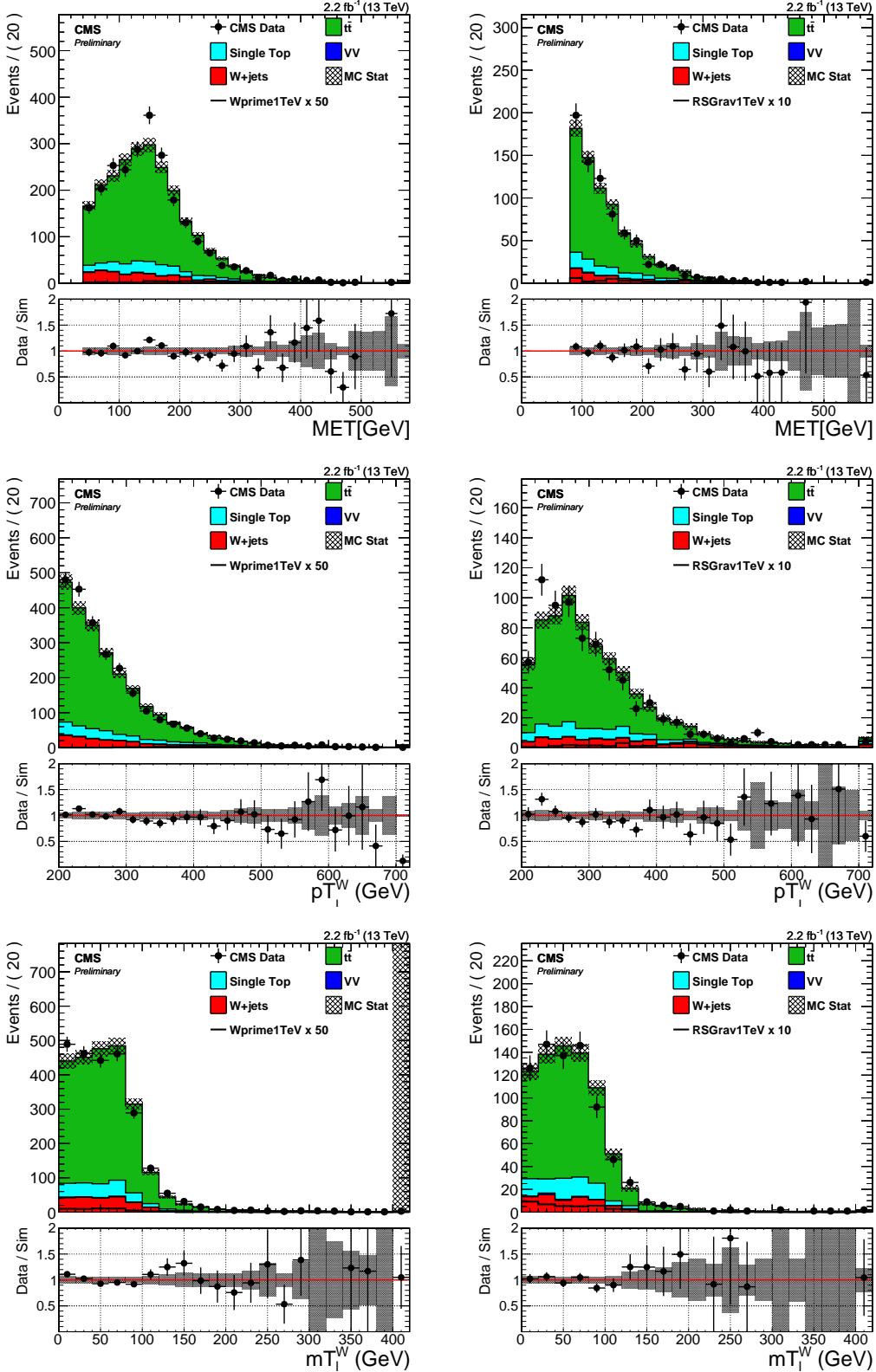
## 13 TeV analysis



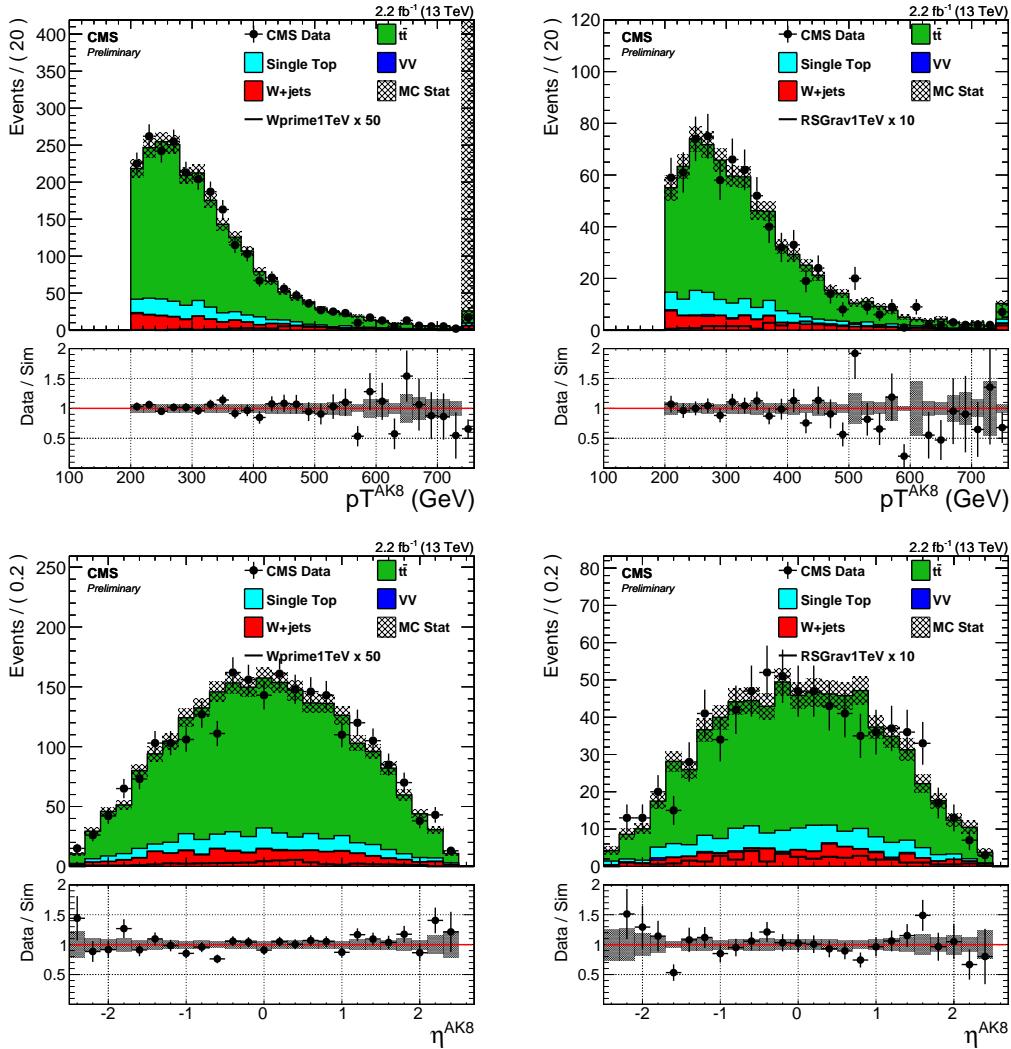
**Figure 9.24:** Comparison plots between data and MC for different observables, in the  $t\bar{t}$  control region. Top: pruned jet mass. Bottom: N-subjettiness. Left: muon channel, right: electron channel.



**Figure 9.25:** Comparison plots between data and MC for different observables, in the  $t\bar{t}$  control region. From top to bottom: number of primary vertices, lepton  $p_T$ , lepton  $\eta$ . Left: muon channel, right: electron channel.



**Figure 9.26:** Comparison plots between data and MC for different observables, in the  $t\bar{t}$  control region. From top to bottom:  $E_T^{\text{miss}}$ , leptonic W  $p_T$ , transverse mass of the leptonic W. Left: muon channel, right: electron channel.



**Figure 9.27:** Comparison plots between data and MC for different observables, in the  $t\bar{t}$  control region. Top:  $p_T$  of the leading AK8 jet. Bottom:  $\eta$  of the leading AK8 jet. Left: muon channel, right: electron channel.

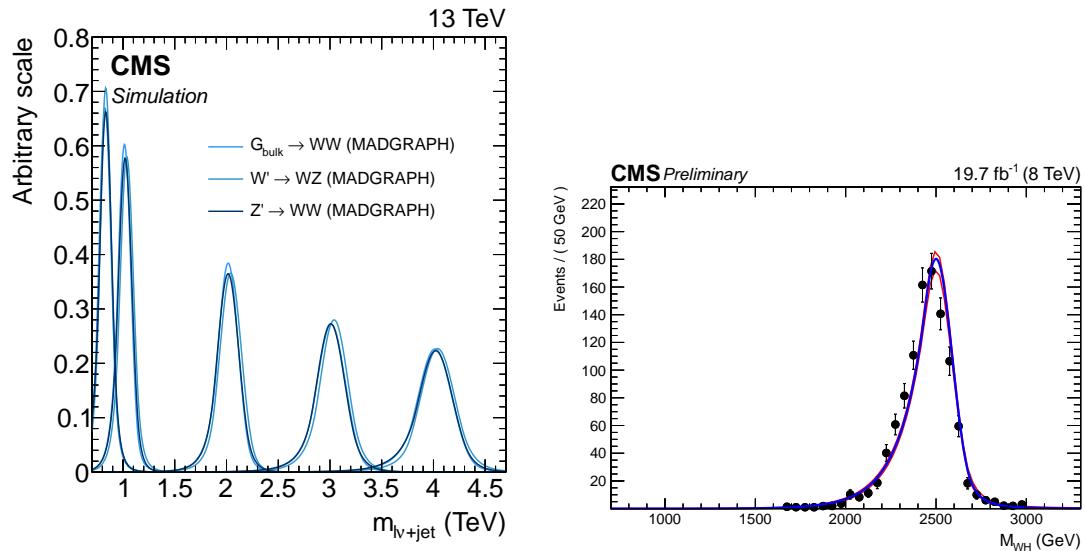
### 9.3 Systematic uncertainties in the background estimation

# Signal modeling and statistical treatment

---

## 10.1 Signal modeling

### 10.1.1 Parametrization of the resonance mass



**Figure 10.1:**  $ml^?+jet$  (right) distribution for different signal mass hypotheses used to extract the signal shape.

### 10.1.2 Signal efficiency

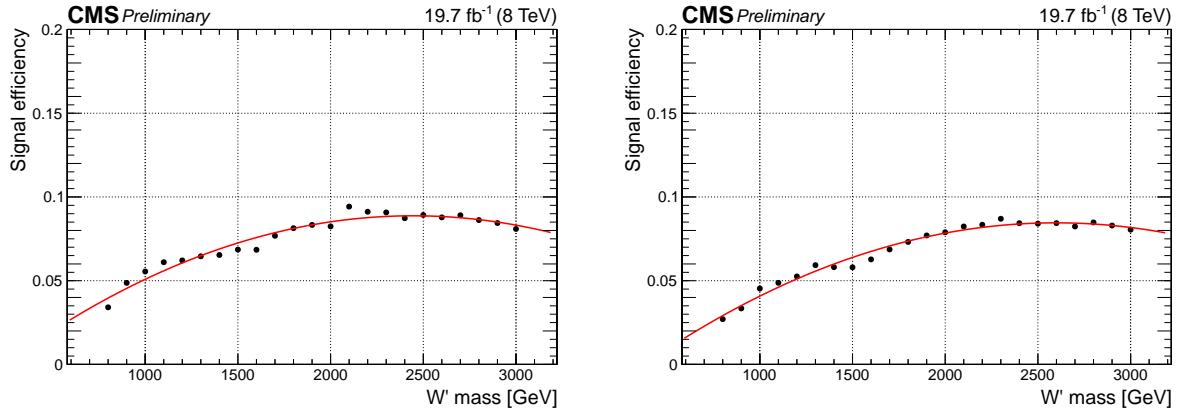
## 10.2 Systematic uncertainties in the signal prediction

## 10.3 Testing new resonance hypothesis

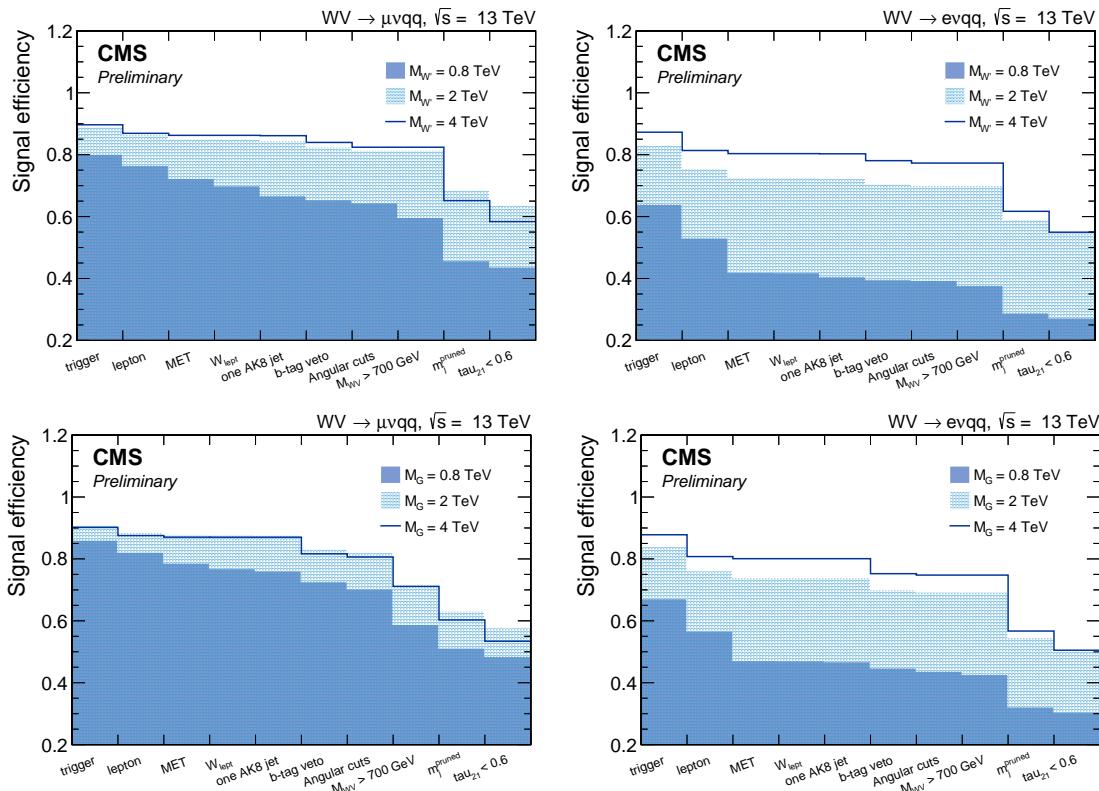
### 10.3.1 Profile likelihood procedure

### 10.3.2 The $CL_s$ method

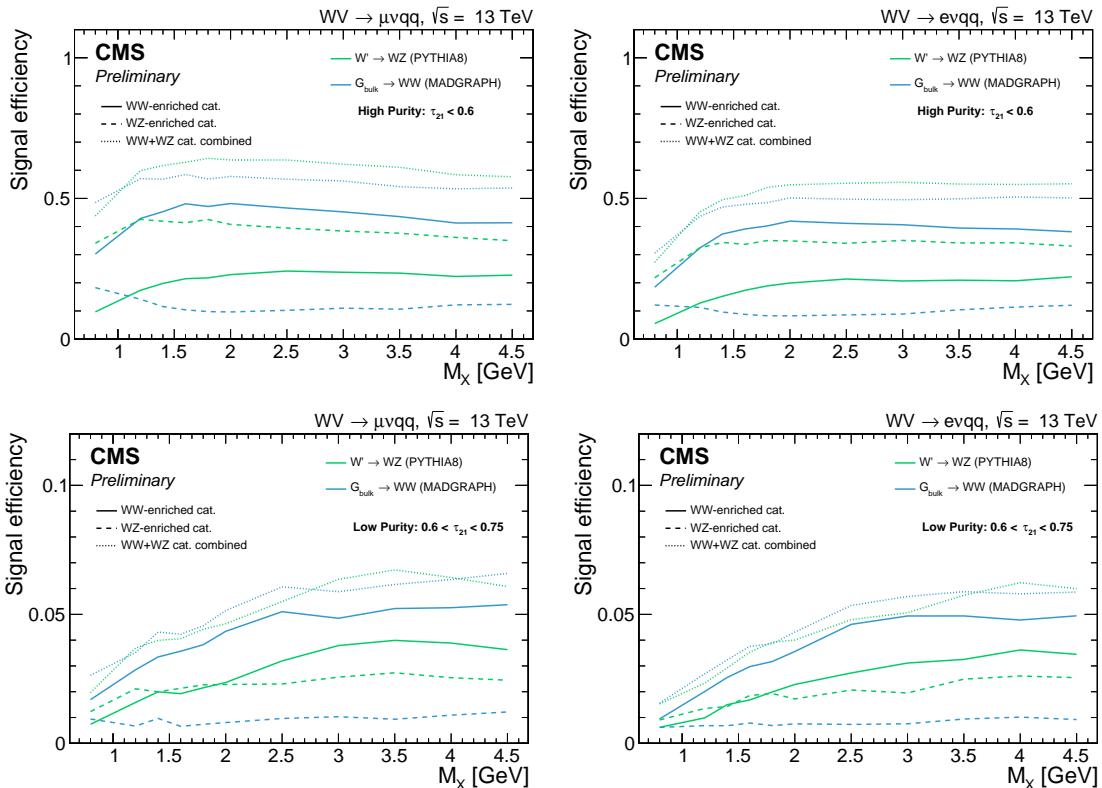
### 10.3.3 Treatment of uncertainties



**Figure 10.2:** Signal efficiency for the final selection criteria as a function of the  $W'$  mass hypothesis in the muon (left) and electron (right) channel.



**Figure 10.3:** Expected signal efficiency of each selection in the muon (left) and electron (right) channels for 3 mass points, for  $WV'$  (top) and BulkG (bottom) signal models.

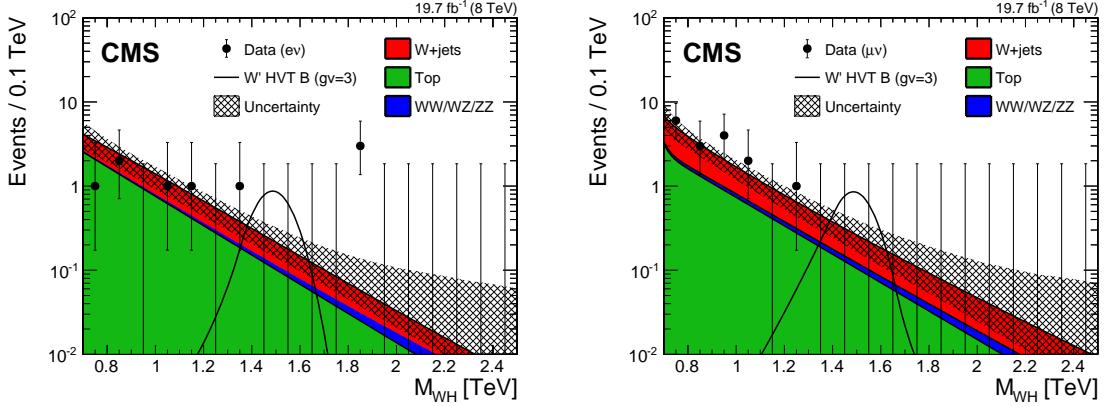


**Figure 10.4:** Expected signal efficiency for high-purity (top) and low-purity (bottom) categories, in electron (left) and muon (right) channel separately, for several resonance mass hypothesis. The efficiencies for the W-mass cut, the Z-mass cut and for the mass categories combined are also shown.

# Results with 8 TeV data

---

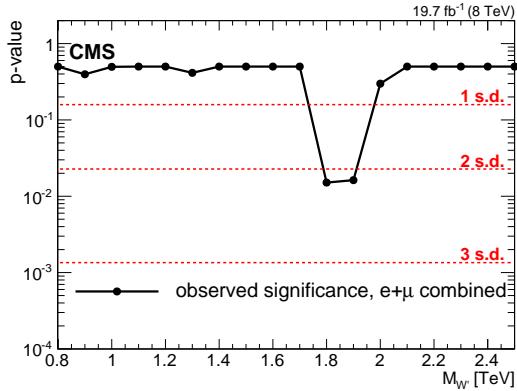
## 11.1 Final $m_{\text{WH}}$ distribution



**Figure 11.1:** Final distributions in  $m_{\text{WH}}$  for data and expected backgrounds for electron (left) and muon (right) categories. The 68% error bars for Poisson event counts are obtained from the Neyman construction [70]. The hatched region indicates the statistical uncertainty of the fit combined with the systematical uncertainty in the shape. This figure also shows a hypothetical  $\text{W}'$  signal with mass of 1.5 TeV, normalized to the cross section predicted by the HVT model B with parameter  $g_V = 3$  as described in Section.

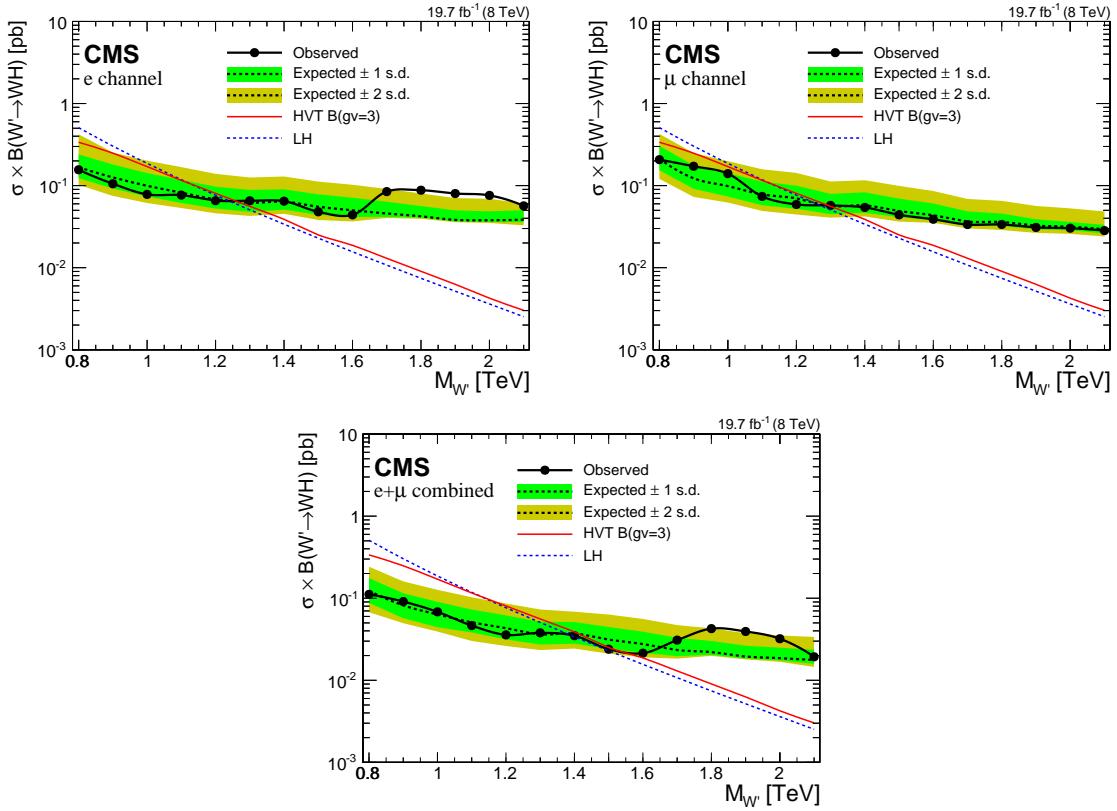
## 11.2 Studies on the excess

### 11.3 Significance of the data



**Figure 11.2:** Local p-value of the combined electron and muon data as a function of the  $\text{W}'$  boson mass, probing a narrow WH resonance.

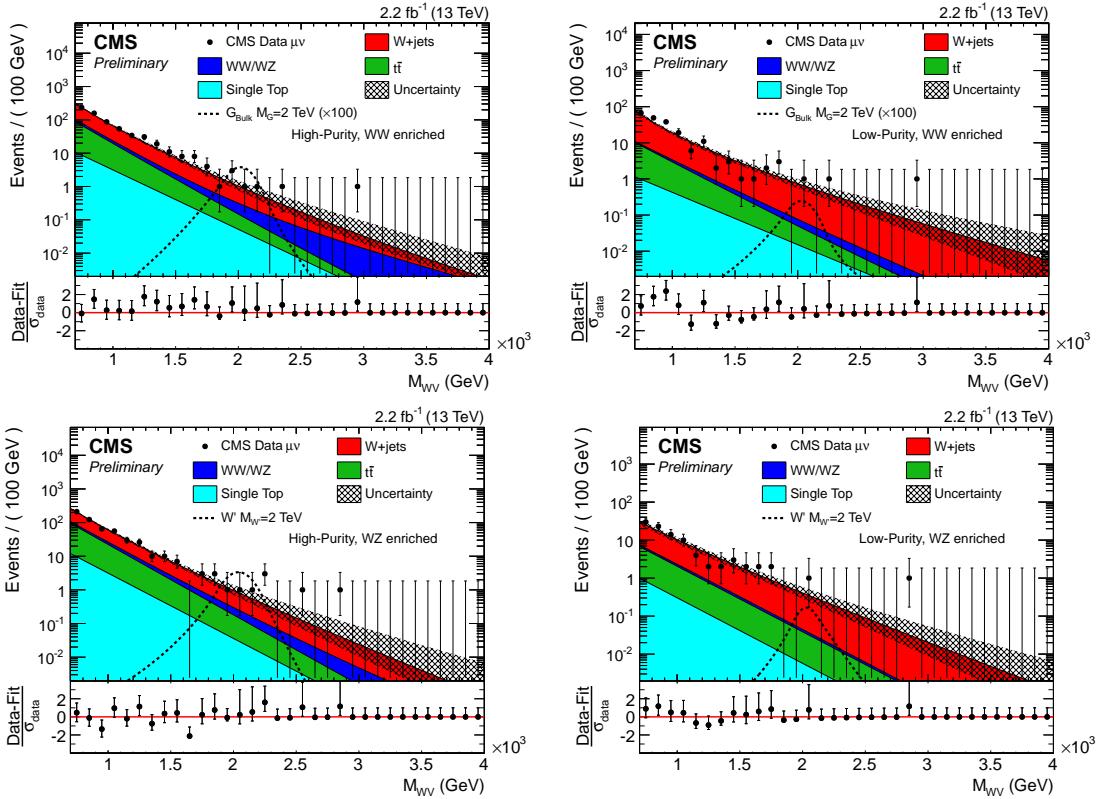
## 11.4 Cross section limits



**Figure 11.3:** Observed (solid) and expected (dashed) upper limits at 95% CL on the product of the  $W'$  production cross section and the branching fraction of  $W' \rightarrow WH$  for electron (left) and muon (right) channels, and the combination of the two channels (lower plot). The products of cross sections and branching fractions for  $W'$  production in the LH and HVT models are overlaid.

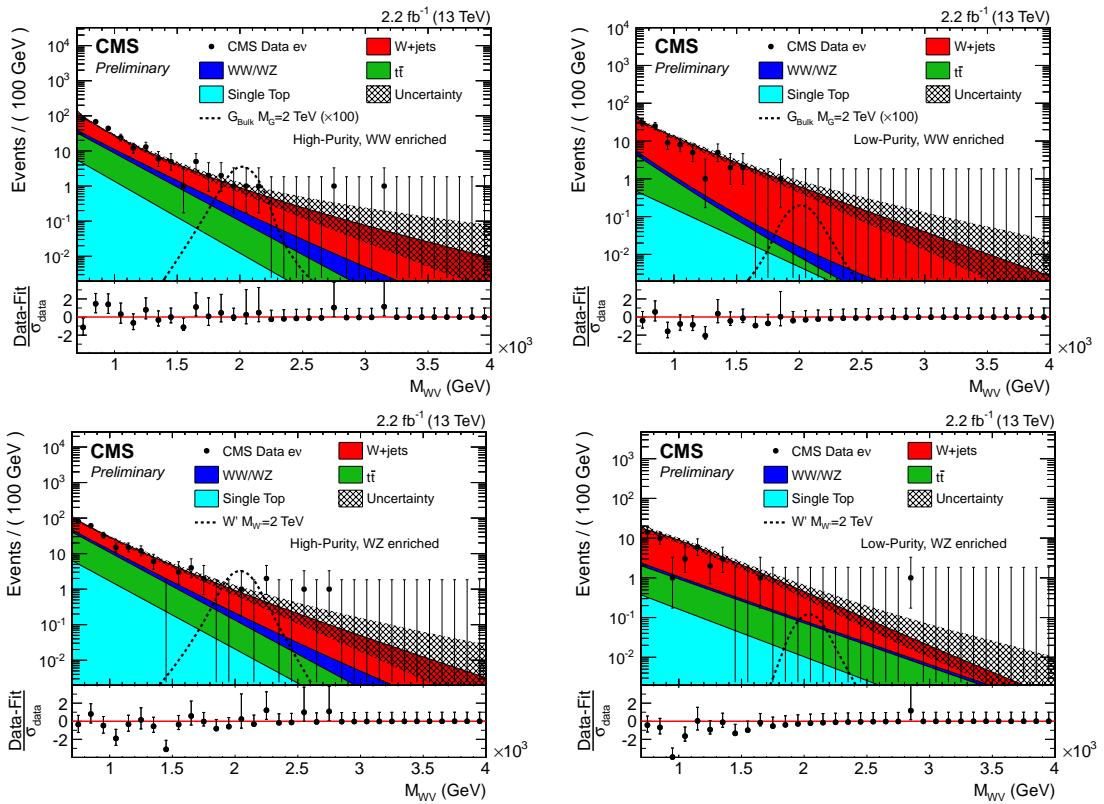
# Results with 13 TeV data

## 12.1 Final $m_{WV}$ distribution

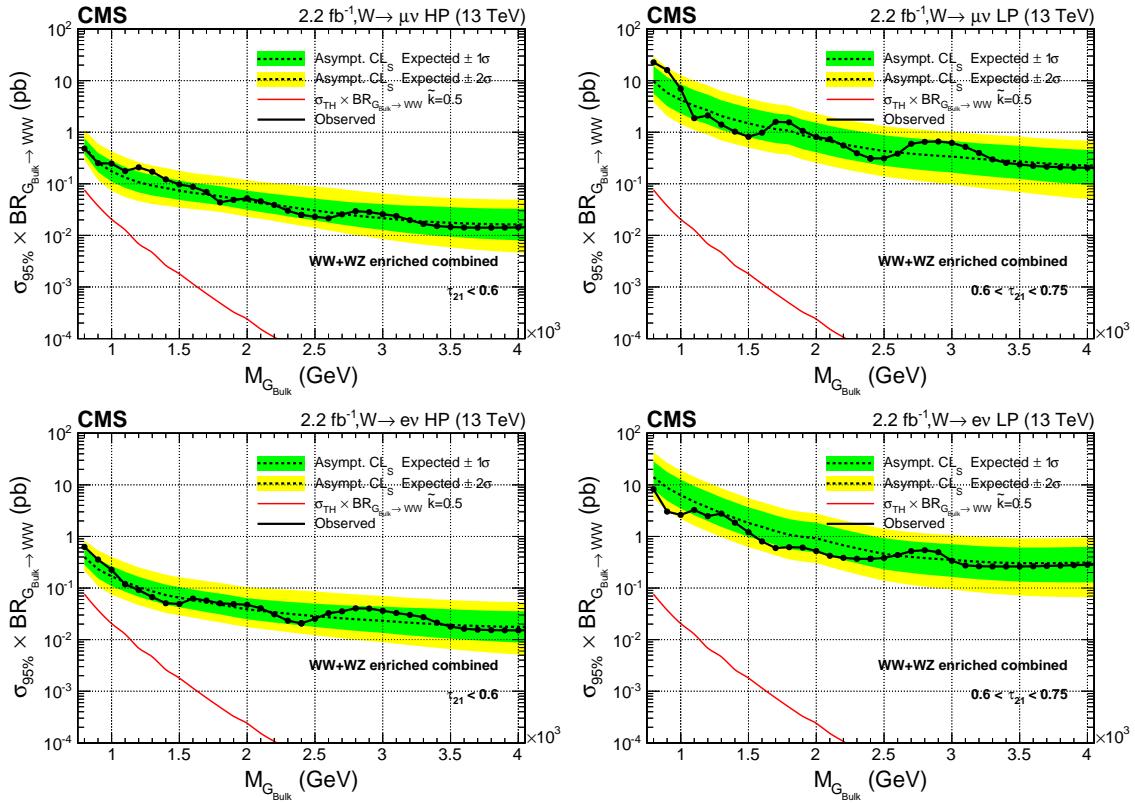


**Figure 12.1:** Examples of  $m_J$  extrapolation of the  $m_{\ell\nu j}$  shape into the Signal Region for the muon channel for the HP (left) and LP (right) categories in the WW (top) and WZ (bottom) signal regions. The expected shape for a Bulk Graviton and for a  $W'$  with a mass of 2 TeV is also shown in the WW-enriched and WZ-enriched category, respectively.

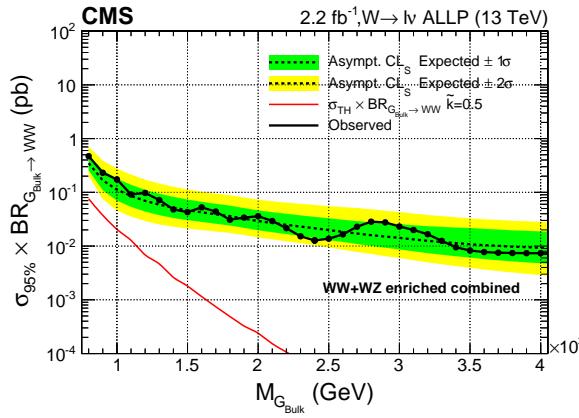
## 12.2 Cross section limits



**Figure 12.2:** Examples of  $m_J$  extrapolation of the  $m_{\ell\nu j}$  shape into the Signal Region for the electron channel for the HP (left) and LP (right) categories in the WW (top) and WZ (bottom) signal regions. The expected shape for a Bulk Graviton and for a  $W'$  with a mass of 2 TeV is also shown in the WW-enriched and WZ-enriched category, respectively.



**Figure 12.3:** Expected 95% CL upper limit on graviton production cross section times the branching fraction of  $G_{\text{bulk}} \rightarrow WW$  assuming  $2.1 \text{ fb}^{-1}$  of data. The limit is obtained with the Asymptotic CLs technique. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The expected product between the Bulk graviton production cross section and the branching fraction is shown as a red solid curve for  $\tilde{k} = 0.5$ . Top panel: results for muon channel, HP category on the left, LP on the right. Bottom panel: results for electron channel, HP category on the left, LP on the right.



**Figure 12.4:** Expected 95% CL upper limit on the graviton production cross section times the branching fraction of  $G_{\text{bulk}} \rightarrow WW$  assuming  $2.1 \text{ fb}^{-1}$  of data. The limit is obtained with the Asymptotic CLs technique. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The expected product between the Bulk graviton production cross section and the branching fraction is shown as a red solid curve for  $\tilde{k} = 0.1$ .

# Combination of searches for diboson resonances at $\sqrt{s} = 8$ and 13 TeV

---

## 13.1 Inputs to the combination

### 13.1.1 8 TeV VV searches

### 13.1.2 13 TeV VV searches

### 13.1.3 8 TeV VH searches

### 13.1.4 13 TeV VH searches

## 13.2 Combination procedure

## 13.3 Results

### 13.3.1 Limits on $W'$

### 13.3.2 Limits on $Z'$

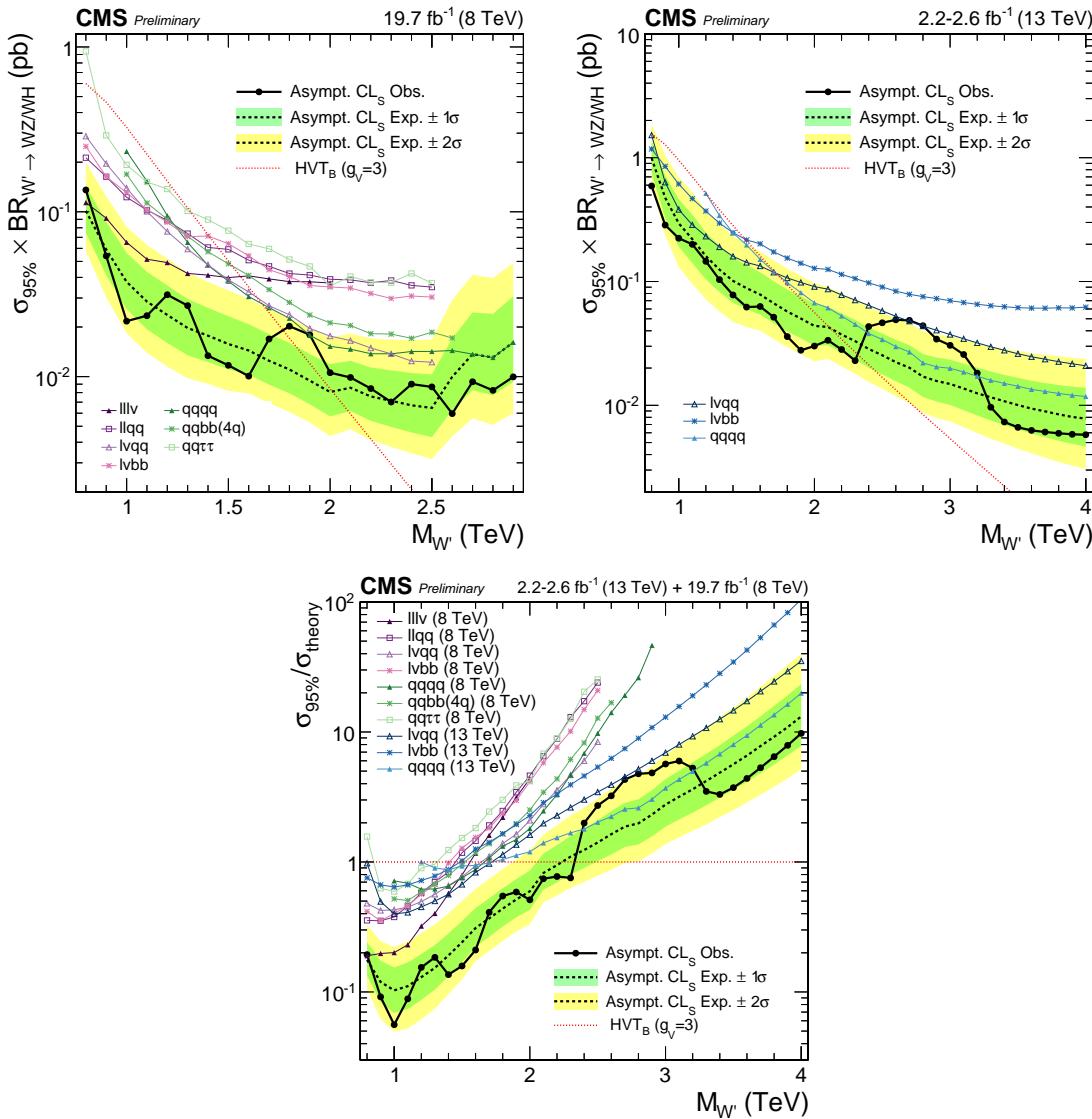
### 13.3.3 Limits on heavy vector triplet ( $W' + Z'$ )

### 13.3.4 Limits on Bulk Graviton

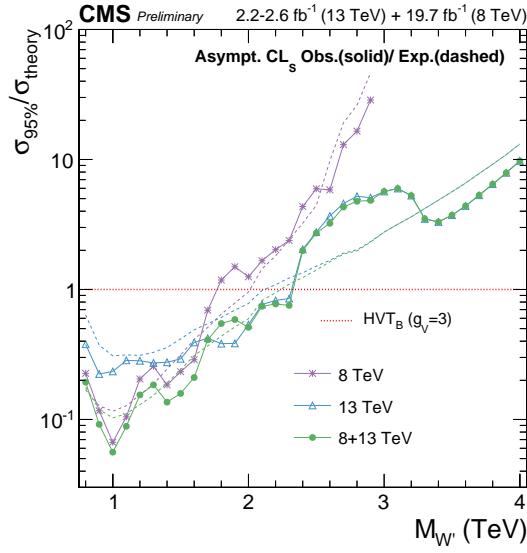
### 13.3.5 Significance at 2 TeV

**Table 13.1:** Statistical significance of excesses observed at 1.8 TeV in the various searches, expressed in standard deviations.

Combination	$W'$	$Z'$	HVT ( $W' + Z'$ )	$G_{\text{bulk}}$
VV 13 TeV	0.00	0.10	0.00	0.00
VV+VH 13 TeV	0.00	0.00	0.00	-
VV 8 TeV	1.22	0.56	1.03	1.61
VV 8+13 TeV	0.20	0.46	0.33	0.35
VH 8 TeV	2.05	0.56	1.79	-
VV+VH 8 TeV	2.22	0.77	1.95	-
VV+VH 8+13 TeV	0.86	0.00	0.83	-



**Figure 13.1:** (top left) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow W' \rightarrow WZ/WH)$  as a function of the resonance mass obtained by combining the 8 TeV diboson searches. The curve corresponding to the cross sections predicted by the HVT model B is overlaid. (top right) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow W' \rightarrow WZ/WH)$  as a function of the resonance mass obtained by combining the 13 TeV diboson searches. The curve corresponding to the cross sections predicted by the HVT model B is overlaid. (bottom) Exclusion limits at 95% CL on the signal strength as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In each of the three plots the different colored lines correspond to the searches entering the combination.



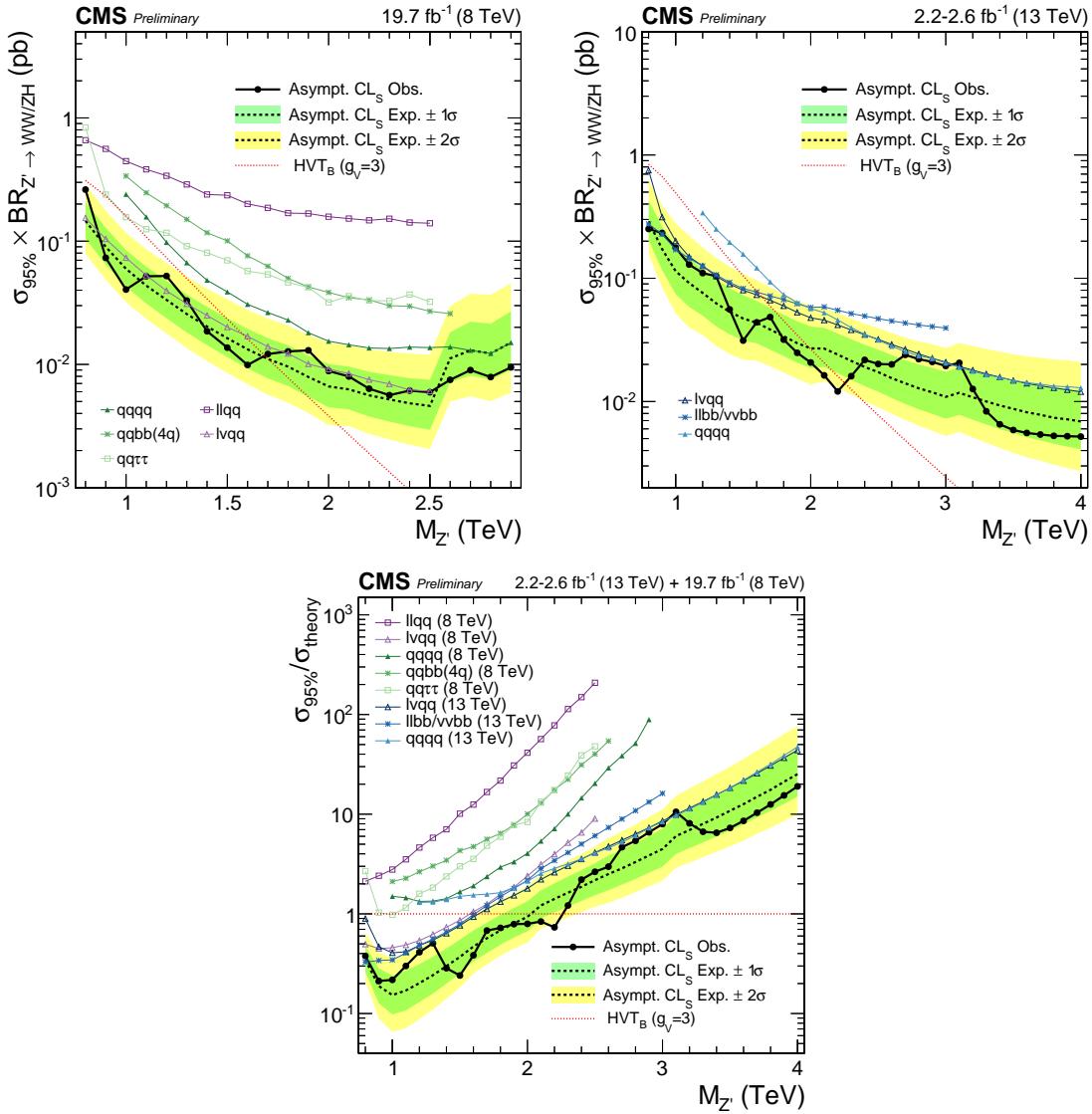
**Figure 13.2:** Comparison of the observed (solid) and expected (dashed) exclusion limits at 95% CL obtained by combining only 8 TeV or only 13 TeV searches to the results from the combination of all the 8 and 13 TeV results.

**Table 13.2:** Statistical significance of excesses observed at 1.9 TeV in the various searches, expressed in standard deviations.

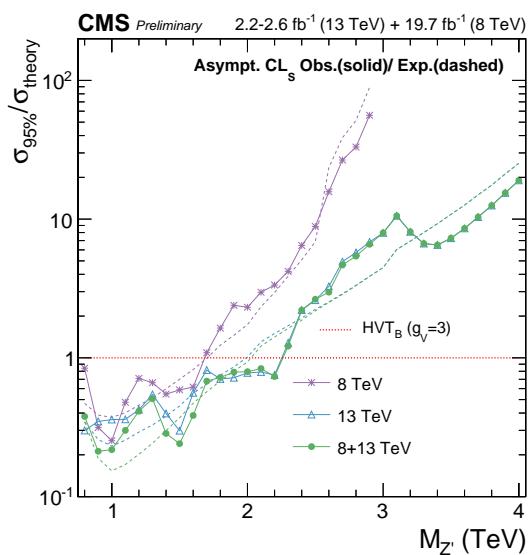
Combination	W'	Z'	HVT (W' +Z')	G <sub>bulk</sub>
VV 13 TeV	0.00	0.05	0.00	0.00
VV+VH 13 TeV	0.00	0.00	0.00	-
VV 8 TeV	1.20	0.46	0.91	1.05
VV 8+13 TeV	0.00	0.30	0.00	0.00
VH 8 TeV	2.17	1.41	1.78	-
VV+VH 8 TeV	2.32	1.02	1.89	-
VV+VH 8+13 TeV	0.33	0.00	0.20	-

**Table 13.3:** Statistical significance of excesses observed at 2 TeV in the various searches, expressed in standard deviations.

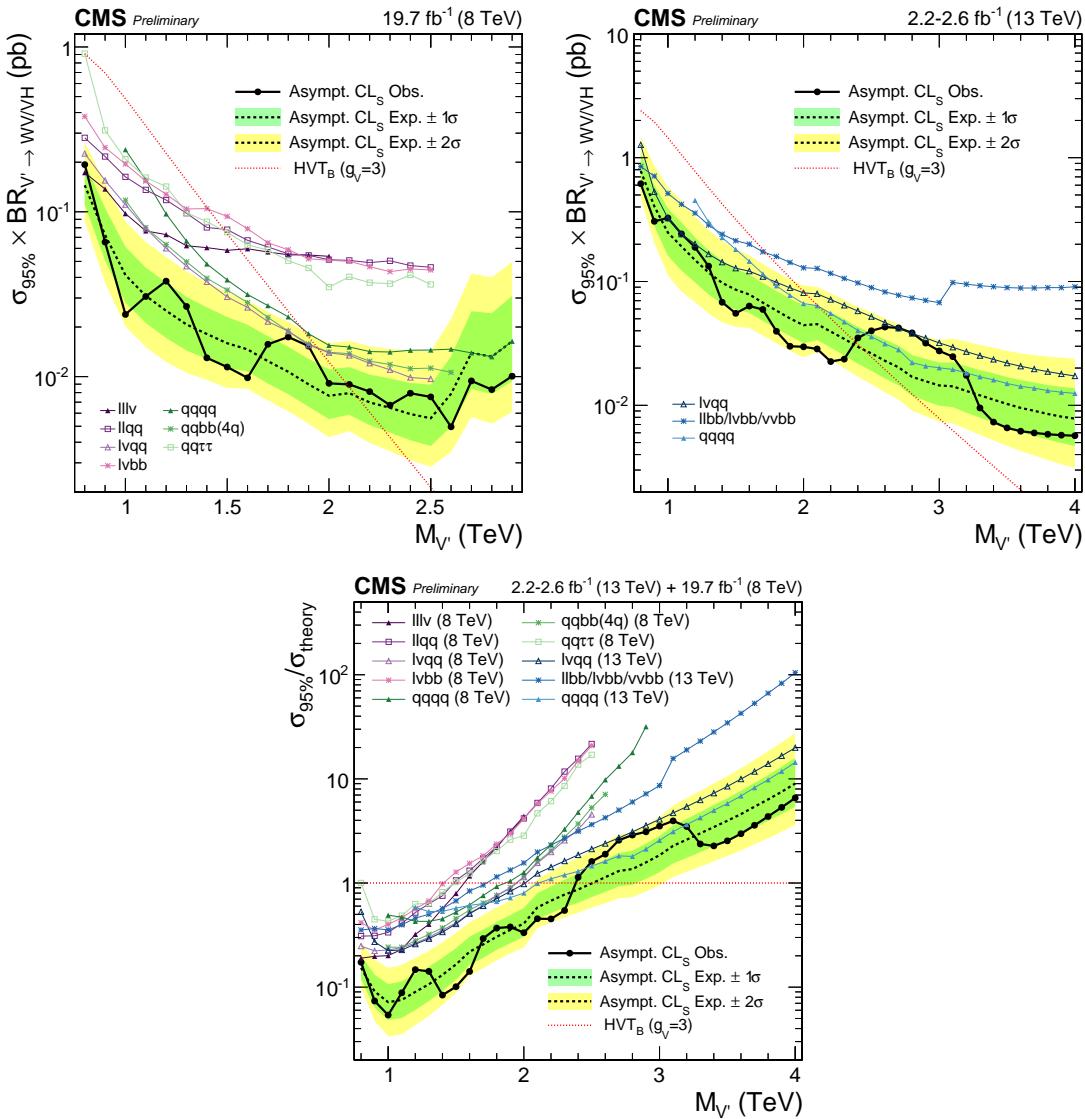
Combination	W'	Z'	HVT (W' +Z')	G <sub>bulk</sub>
VV 13 TeV	0.00	0.07	0.00	0.00
VV+VH 13 TeV	0.00	0.00	0.00	-
VV 8 TeV	0.77	0.75	0.76	0.44
VV 8+13 TeV	0.23	0.45	0.29	0.06
VH 8 TeV	0.00	0.00	0.00	-
VV+VH 8 TeV	0.58	0.60	0.48	-
VV+VH 8+13 TeV	0.00	0.00	0.00	-



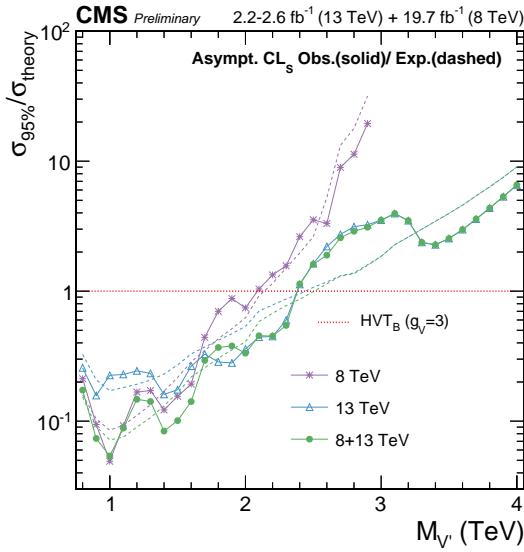
**Figure 13.3:** (top left) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow Z' \rightarrow WW/ZH)$  as a function of the resonance mass obtained by combining the 8 TeV diboson searches. The curve corresponding to the cross sections predicted by the HVT model B is overlaid. (top right) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow Z' \rightarrow WW/ZH)$  as a function of the resonance mass obtained by combining the 13 TeV diboson searches. The curve corresponding to the cross sections predicted by the HVT model B is overlaid. (bottom) Exclusion limits at 95% CL on the signal strength as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In each of the three plots the different colored lines correspond to the searches entering the combination.



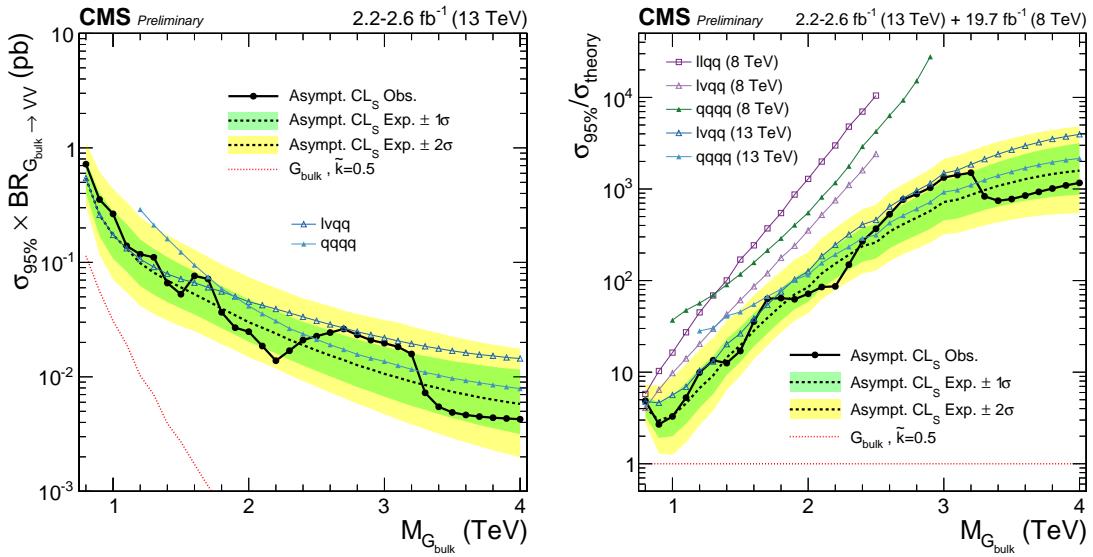
**Figure 13.4:** Comparison of the observed (solid) and expected (dashed) exclusion limits at 95% CL obtained by combining only 8 TeV or only 13 TeV searches to the results from the combination of all the 8 and 13 TeV results.



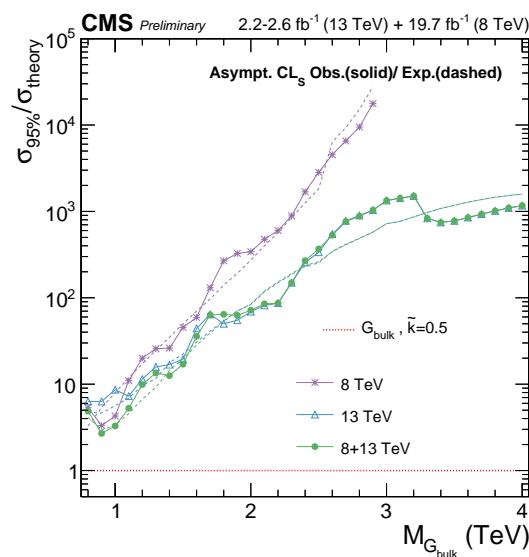
**Figure 13.5:** (top left) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow V' \rightarrow WV/VH)$  ( $V' = W', Z'$  and  $V=W,Z$ ) as a function of the resonance mass obtained by combining the 8 TeV diboson searches. The curve corresponding to the cross sections predicted by the HVT model B is overlaid. (top right) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow V' \rightarrow WV/VH)$  as a function of the resonance mass obtained by combining the 13 TeV diboson searches. The curve corresponding to the cross sections predicted by the HVT model B is overlaid. (bottom) Exclusion limits at 95% CL on the signal strength as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In all the three plots the different colored lines correspond to the searches entering the combination.



**Figure 13.6:** Comparison of the observed (solid) and expected (dashed) exclusion limits at 95% CL obtained by combining only 8 TeV or only 13 TeV searches to the results from the combination of all the 8 and 13 TeV results.



**Figure 13.7:** (left) Observed (black solid) and expected (black dashed) exclusion limits at 95% CL on  $\sigma(pp \rightarrow G_{\text{bulk}} \rightarrow VV)$  ( $V=W,Z$ ) as a function of the resonance mass obtained by combining the 13 TeV diboson searches. The curve corresponding to the cross sections predicted by bulk graviton model is overlaid. (right) Exclusion limits at 95% CL on the signal strength as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In all the three plots the different colored lines correspond to the searches entering the combination.



**Figure 13.8:** Comparison of the observed (solid) and expected (dashed) exclusion limits at 95% CL obtained by combining only 8 TeV or only 13 TeV searches to the results from the combination of all the 8 and 13 TeV results.

CHAPTER 14

# Conclusions

---

## **Part II**

# **Calibration and Upgrade of the CMS Pixel Barrel Detector**

# **introduction chapter: why pixels are so important for physics**

---

# The CMS Pixel Barrel Detector

---

## 16.1 Design of the CMS Pixel Barrel Detector

### 16.2 Detector modules

#### 16.2.1 Sensor

#### 16.2.2 Readout Chip

#### 16.2.3 Token Bit Manager

### 16.3 Readout and control system

#### 16.3.1 Analog readout chain

#### 16.3.2 Front End Driver

#### 16.3.3 Supply Tube

#### 16.3.4 Communication and Control Unit

#### 16.3.5 Front End Controller

### 16.4 Pixel Online Software

### 16.5 Performance at $\sqrt{s} = 8$ and 13 TeV

# Optimization and commissioning for LHC Run II

---

## 17.1 Radiation damage after LHC Run I

## 17.2 Optimization for LHC Run II

### 17.2.1 Overview of pixel calibrations

### 17.2.2 Temperature dependence

## 17.3 Commissioning for LHC Run II

### 17.3.1 Installation into CMS

### 17.3.2 Check out of optical connections

### 17.3.3 Adjustment of readout chain settings

### 17.3.4 Optimisation of signal performance

# Phase I Upgrade of the CMS Pixel Barrel Detector

---

**18.1 Motivations**

**18.2 Summary of changes**

**18.3 The digital readout chain**

**18.4 The Phase I supply tubes**

**18.5 The test stand**

**18.6 Supply tubes assembly and commissioning**

**18.7 Detector commissioning**

CHAPTER 19

# Conclusions

---

# **Part III**

# **Summary**



# Bibliography

- [1] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001.
- [2] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [3] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [4] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.
- [5] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.
- [6] F. Marcastel, “CERN’s Accelerator Complex. La chane des accrateurs du CERN”., General Photo.
- [7] CMS Collaboration, “CMS Public Luminosity Results”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [8] N. Cartiglia, “Measurement of the proton-proton total, elastic, inelastic and diffractive cross sections at 2, 7, 8 and 57 TeV”, arXiv:1305.6131.
- [9] CMS Collaboration, “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, 1997.
- [10] CMS Collaboration, “CMS Pixel Detector 2015 Offline Performance Results”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflineLHCC2015>.
- [11] CMS Collaboration Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [12] CMS Collaboration Collaboration, “The CMS hadron calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [13] P. Adzic, “Energy resolution of the barrel of the CMS Electromagnetic Calorimeter”, *Journal of Instrumentation* **2** (2007), no. 04, P04004.
- [14] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [15] W. Adam, B. Mangano, T. Speer, and T. Todorov, “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Dec, 2006.
- [16] P. Billoir, “Progressive track recognition with a Kalman like fitting procedure”, *Comput. Phys. Commun.* **57** (1989) 390–394, doi:10.1016/0010-4655(89)90249-X.
- [17] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.

- [18] T. Speer et al., “Vertex Fitting in the CMS Tracker”, Technical Report CMS-NOTE-2006-032, CERN, Feb, 2006.
- [19] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239, doi:10.1109/5.726788.
- [20] W. Waltenberger, R. Frhwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 12, N343.
- [21] CMS Collaboration, “CMS Tracking Performance Results”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTRK>.
- [22] S. Baffioni et al., “Electron reconstruction in CMS”, *Eur. Phys. J.* **C49** (2007) 1099–1116, doi:10.1140/epjc/s10052-006-0175-5.
- [23] W. Adam, R. Frhwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.
- [24] CMS Collaboration, “Measurements of Inclusive  $W$  and  $Z$  Cross Sections in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV”, *JHEP* **01** (2011) 080, doi:10.1007/JHEP01(2011)080, arXiv:1012.2466.
- [25] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV”, *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [26] CMS Collaboration, “CMS ECAL Performance Results”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResults>.
- [27] CMS Collaboration, “Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *JHEP* **04** (2015) 025, doi:10.1007/JHEP04(2015)025, arXiv:1412.6302.
- [28] CMS Collaboration, “Performance of CMS muon reconstruction in  $pp$  collision events at  $\sqrt{s} = 7$  TeV”, *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.
- [29] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, 2006.
- [30] CMS Collaboration, “Additional Performance Results for EPS 2015”,.
- [31] R. Radogna, A. Colaleo, and S. Nuzzo, “Search for high-mass resonances decaying into muon pairs with the CMS experiment at LHC”. PhD thesis, Bari U., 2016. Presented 14 Mar 2016.
- [32] CMS Collaboration, “Performance of the CMS Level-1 Trigger”, Technical Report CMS-CR-2012-322, CERN, 2012. Comments: Presented at 36th International Conference on High Energy Physics, July 4-11, 2012, Melbourne, Australia.
- [33] G. P. Salam and G. Soyez, “A Practical Seedless Infrared-Safe Cone jet algorithm”, *JHEP* **05** (2007) 086, doi:10.1088/1126-6708/2007/05/086, arXiv:0704.0292.

- [34] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, “Longitudinally invariant  $k_t$  clustering algorithms for hadron hadron collisions”, *Nucl. Phys. B* **406** (1993) 187, doi:10.1016/0550-3213(93)90166-M.
- [35] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev. D* **48** (1993) 3160, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [36] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms”, *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.
- [37] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*. 1998. arXiv:hep-ph/9907280.
- [38] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [39] CMS Collaboration, “Identification techniques for highly boosted W bosons that decay into hadrons”, *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.
- [40] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.
- [41] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- [42] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *Submitted to: JINST* (2016) arXiv:1607.03663.
- [43] T. C. collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *Journal of Instrumentation* **6** (2011), no. 11, P11002.
- [44] CMS Collaboration, “Pileup Jet Identification”, Technical Report CMS-PAS-JME-13-005, CERN, 2013.
- [45] CMS Collaboration, “Pileup Removal Algorithms”, Technical Report CMS-PAS-JME-14-001, CERN, 2014.
- [46] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, 2014.
- [47] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [48] M. Cacciari and G. P. Salam, “Dispelling the  $N^3$  myth for the  $k_t$  jet-finder”, *Phys. Lett. B* **641** (2006) 57–61, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- [49] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

- [50] “Jet energy scale and resolution performances with 13TeV data”,.
- [51] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.
- [52] W. Waltenberger, “Adaptive Vertex Reconstruction”, Technical Report CMS-NOTE-2008-033, CERN, Geneva, Jul, 2008.
- [53] CMS Collaboration, “Measurement of  $B\bar{B}$  Angular Correlations based on Secondary Vertex Reconstruction at  $\sqrt{s} = 7$  TeV”, *JHEP* **03** (2011) 136, doi:10.1007/JHEP03(2011)136, arXiv:1102.3194.
- [54] CMS Collaboration Collaboration, “Performance of b tagging at  $\text{sqrt}(s)=8$  TeV in multijet,  $t\bar{t}$  and boosted topology events”, Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, 2013.
- [55] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, Technical Report CMS-PAS-BTV-15-001, CERN, 2016.
- [56] T. C. collaboration, “Missing transverse energy performance of the CMS detector”, *Journal of Instrumentation* **6** (2011), no. 09, P09001.
- [57] CMS Collaboration, “Performance of the CMS missing transverse momentum reconstruction in pp data at  $\sqrt{s} = 8$  TeV”, *JINST* **10** (2015), no. 02, P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.
- [58] CMS Collaboration, “Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector”, Technical Report CMS-PAS-JME-16-004, CERN, 2016.
- [59] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.
- [60] J. Bauer, “Prospects for the Observation of Electroweak Top-Quark Production with the CMS Experiment”. PhD thesis, Karlsruher Institut für Technologie (KIT), 2010.
- [61] CMS Collaboration, “Studies of jet mass in dijet and W/Z + jet events”, *JHEP* **05** (2013) 090, doi:10.1007/JHEP05(2013)090, arXiv:1303.4811.
- [62] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Techniques for improved heavy particle searches with jet substructure”, *Phys. Rev. D* **80** (2009) 051501, doi:10.1103/PhysRevD.80.051501, arXiv:0903.5081.
- [63] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches”, *Phys. Rev. D* **81** (2010) 094023, doi:10.1103/PhysRevD.81.094023, arXiv:0912.0033.
- [64] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- [65] CMS Collaboration, “Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at  $\sqrt{s} = 8$  TeV”, *JHEP* **08** (2014) 174, doi:10.1007/JHEP08(2014)174, arXiv:1405.3447.

- [66] CMS Collaboration, “Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at  $\sqrt{s} = 8$  TeV”, *JHEP* **08** (2014) 173, doi:10.1007/JHEP08(2014)173, arXiv:1405.1994.
- [67] CMS Collaboration, “Search for massive resonances decaying into pairs of boosted W and Z bosons at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-EXO-15-002, CERN, 2015.
- [68] CMS Collaboration, “V Tagging Observables and Correlations”, Technical Report CMS-PAS-JME-14-002, CERN, 2014.
- [69] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [70] F. Garwood, “Fiducial Limits for the Poisson Distribution”, *Biometrika* **28** (1936) 437, doi:10.1093/biomet/28.3-4.437.