
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

1	Introduction	1
2	The Standard Model and beyond	2
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
3	The CMS Experiment at the LHC	3
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
I	Search for diboson resonances with CMS	17
4	Diboson resonances as signature for new physics	18
5	Data sets and simulated samples	21
5.1	Simulation of proton-proton collisions	21
5.1.1	Monte Carlo event generators	21
5.1.2	CMS detector simulation	26
5.2	Simulated samples	26
5.2.1	Simulation of signal processes	26
5.2.2	Simulation of background processes	29
5.3	Data sets	30
6	Object and event reconstruction	33
6.1	Tracks and vertices	33
6.1.1	Track reconstruction	33
6.1.2	Vertex reconstruction	35
6.2	Electrons	36
6.2.1	Electron reconstruction	36
6.2.2	Electron trigger	38
6.2.3	Electron identification	41
6.3	Muons	44
6.3.1	Muon reconstruction	44

6.3.2	Muon trigger	48
6.3.3	Muon identification	49
6.4	Jets	52
6.4.1	Jet clustering algorithms	52
6.4.2	Jet reconstruction and calibration	54
6.4.3	Identification of b jets	60
6.5	Missing transverse energy	65
6.6	$W \rightarrow \ell\nu$ reconstruction	67
7	Identification of highly boosted $W/Z \rightarrow q\bar{q}^{(\prime)}$ and $H \rightarrow b\bar{b}$	70
7.1	Jet substructure observables	70
7.1.1	Pruned jet mass	70
7.1.2	N-subjettiness	71
7.2	The V tagging algorithm	73
7.3	The H tagging algorithm	75
8	Analysis strategy	79
8.1	Final event selection and categorization	79
8.2	Modelling of top quark production	83
8.3	$W+jets$ background estimate with α ratio method	83
8.3.1	Description	83
8.3.2	Extraction of the $W+jets$ normalization	85
8.3.3	Extraction of the $W+jets$ shape	86
8.3.4	Validation of the α method	90
8.4	Signal modeling	91
8.5	Systematic uncertainties	94
8.5.1	Systematic uncertainties in the background estimation	95
8.5.2	Systematic uncertainties in the signal prediction	97
8.6	Testing new resonance hypothesis	98
8.6.1	Limit setting procedure	99
8.6.2	The asymptotic approximation	102
8.6.3	Quantifying an excess of events	103
9	Results with 8 TeV data	105
9.1	Final m_{WH} distribution	105
9.2	Significance of the data	105
9.3	Cross section limits	107
10	Results with 13 TeV data	109
10.1	Final m_{WV} distribution	109
10.2	Cross section limits	110
11	Combination of searches for diboson resonances at $\sqrt{s} = 8$ and 13 TeV	112
11.1	Inputs to the combination	112
11.1.1	Reinterpretations	114
11.2	Combination procedure	114
11.3	Results	116
11.3.1	Limits on W' and Z' singlets	116
11.3.2	Limits on heavy vector triplet ($W' + Z'$)	117
11.3.3	Limits on bulk graviton	119

11.3.4 Significance at 2 TeV	119
12 Conclusions	121
II Calibration and Upgrade of the CMS Pixel Barrel Detector	122
13 introduction chapter: why pixels are so important for physics	123
14 The CMS Pixel Barrel Detector	124
14.1 Design of the CMS Pixel Barrel Detector	124
14.2 Detector modules	124
14.2.1 Sensor	124
14.2.2 Readout Chip	124
14.2.3 Token Bit Manager	124
14.3 Readout and control system	124
14.3.1 Analog readout chain	124
14.3.2 Front End Driver	124
14.3.3 Supply Tube	124
14.3.4 Communication and Control Unit	124
14.3.5 Front End Controller	124
14.4 Pixel Online Software	124
14.5 Performance at $\sqrt{s} = 8$ and 13 TeV	124
15 Optimization and commissioning for LHC Run II	125
15.1 Radiation damage after LHC Run I	125
15.2 Optimization for LHC Run II	125
15.2.1 Overview of pixel calibrations	125
15.2.2 Temperature dependence	125
15.3 Commissioning for LHC Run II	125
15.3.1 Installation into CMS	125
15.3.2 Check out of optical connections	125
15.3.3 Adjustment of readout chain settings	125
15.3.4 Optimisation of signal performance	125
16 Phase I Upgrade of the CMS Pixel Barrel Detector	126
16.1 Motivations	126
16.2 Summary of changes	126
16.3 The digital readout chain	126
16.4 The Phase I supply tubes	126
16.5 The test stand	126
16.6 Supply tubes assembly and commissioning	126
16.7 Detector commissioning	126
17 Conclusions	127
III Summary	128
A Studies on track reconstruction problems	129

Bibliography	133
---------------------	------------

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 μm 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 σ , 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, γ_r the relativistic gamma factor, while σ_x and σ_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 μm 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 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 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.

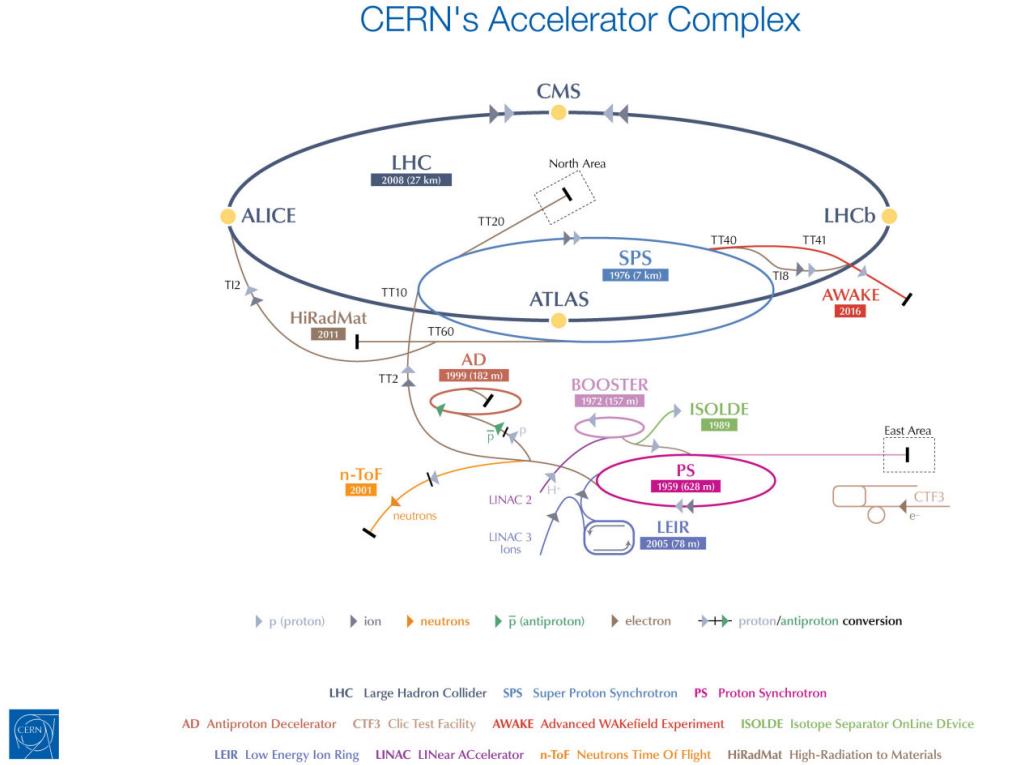


Figure 3.1: The LHC scheme together with its injection chain and the locations of the four main experiments ATLAS, CMS, LHCb and ALICE [6].

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 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 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 fb^{-1} . The data collected in 2016 are not considered in this work.

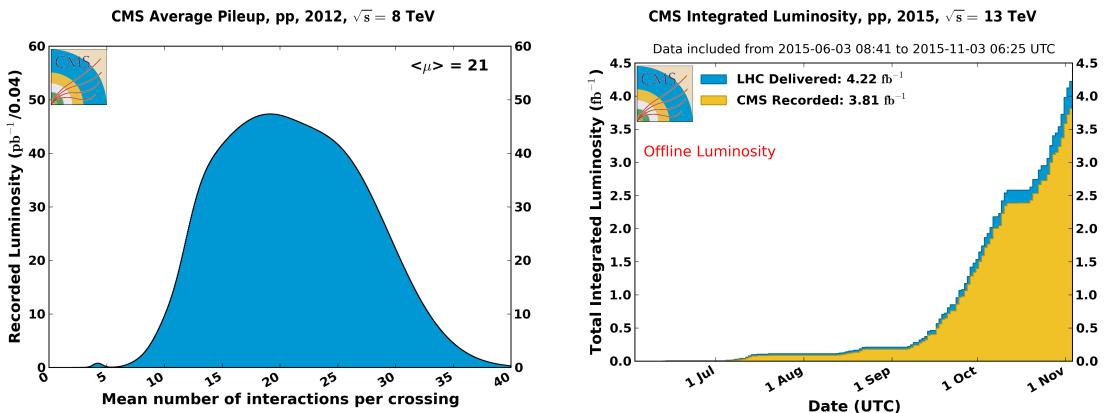


Figure 3.2: Mean number of interactions per bunch crossing in data collected in 2012 by the CMS experiment at LHC (left). Cumulative luminosity versus day (right) delivered by LHC (blue) in 2015; the off-line luminosity recorded by the CMS experiment is also reported (orange). [7]

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 τ -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 ϕ is measured from the x -axis in the x - y plane. The polar angle θ 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.

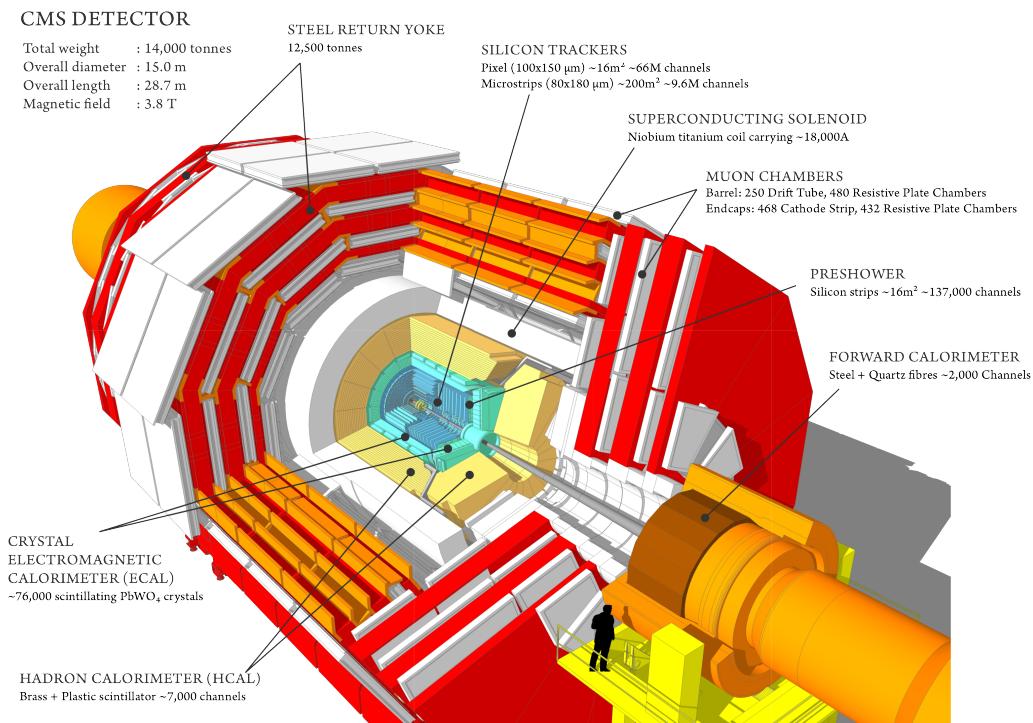


Figure 3.3: Layout of the CMS experiment and its sub-detectors.

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 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 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 kHz/mm^2 at a radius of 22 cm and to 3 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 C_6F_{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°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 m^2 . The pixel detector is essential for the reconstruction of secondary vertices from bottom quarks and τ 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° 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

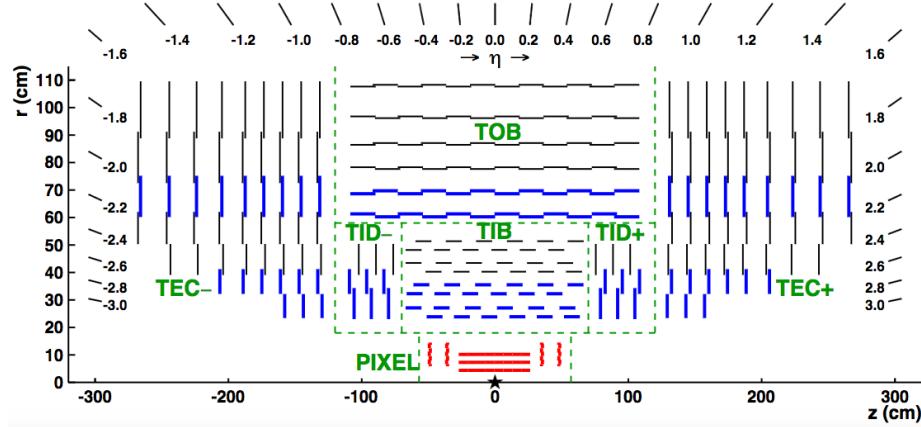


Figure 3.4: Longitudinal section of half of the CMS Tracker system; the different detector types are indicated.

the functioning of the CMS pixel barrel detector is given in Chapter 14.

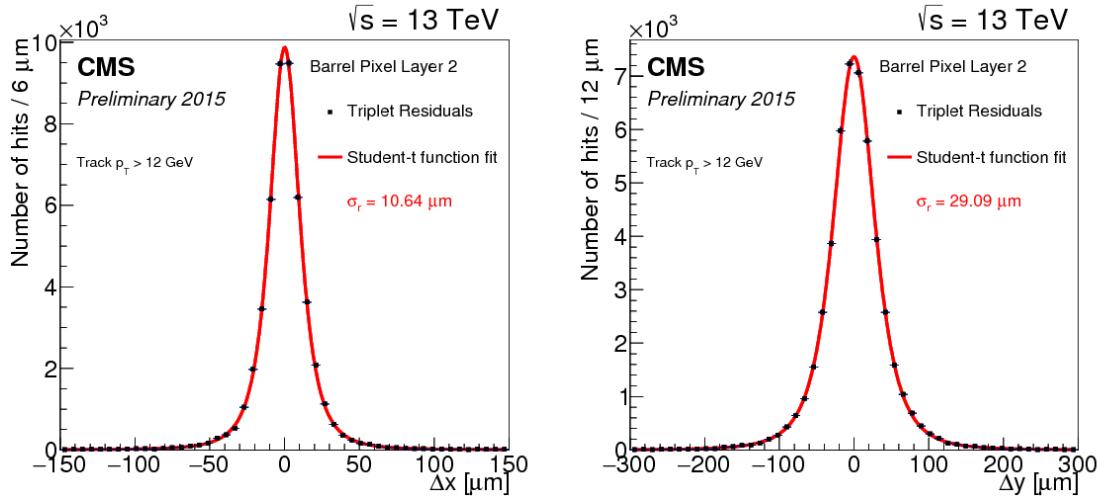


Figure 3.5: Distribution of hit residuals on pixel barrel layer 2 in the transverse (left) and longitudinal (right) direction with respect to the beam. The distributions are fitted with a Student's t-function for which sigma is shown on the plot [10].

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 m², 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 (PbWO₄). The choice of this material is motivated by its high density (8.28 g/cm³), 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 m³ and 2.9 m³, 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 mm² and a length of 2.3 cm (25.8 X_0). They are organised in 36 identical supermodules each covering 20° in ϕ . The crystals are contained in a thin-walled glass-fibre alveola structures (“submodules”) with 2(ϕ)×5(η) crystals per each resulting in a granularity 360-fold in ϕ and 2×85-fold in η . 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 mm² 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].

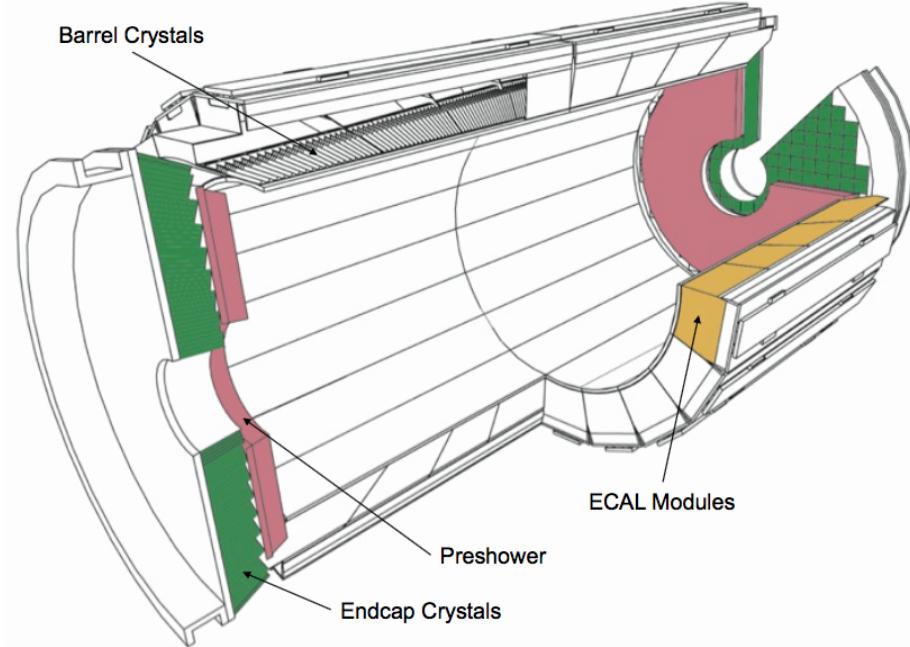


Figure 3.6: Schematic view of the CMS Electromagnetic Calorimeter [3].

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 (λ_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$)
$\Delta\phi \times \Delta\eta$	0.087×0.087	0.087×0.087	0.175×0.175	0.175×0.175	0.175×0.35

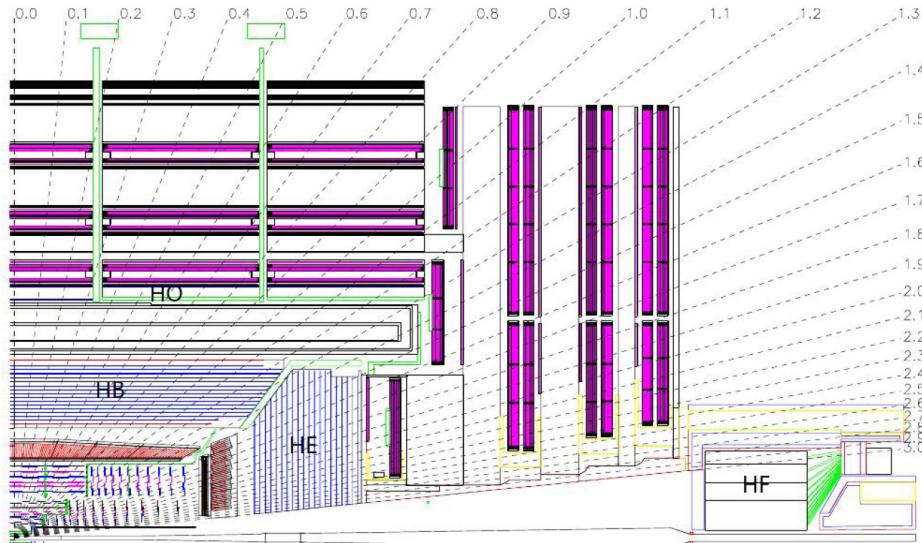


Figure 3.7: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [3].

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° 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 ϕ 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

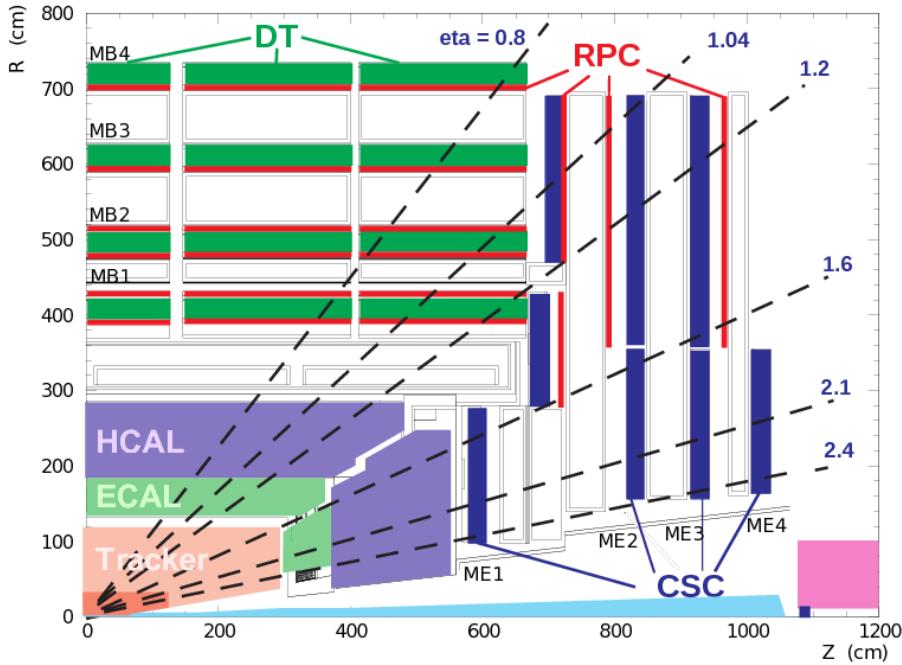


Figure 3.8: A longitudinal view of one quarter of the CMS experiment; the three muon detectors detector types are highlighted.

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 ~ 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 η , 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.

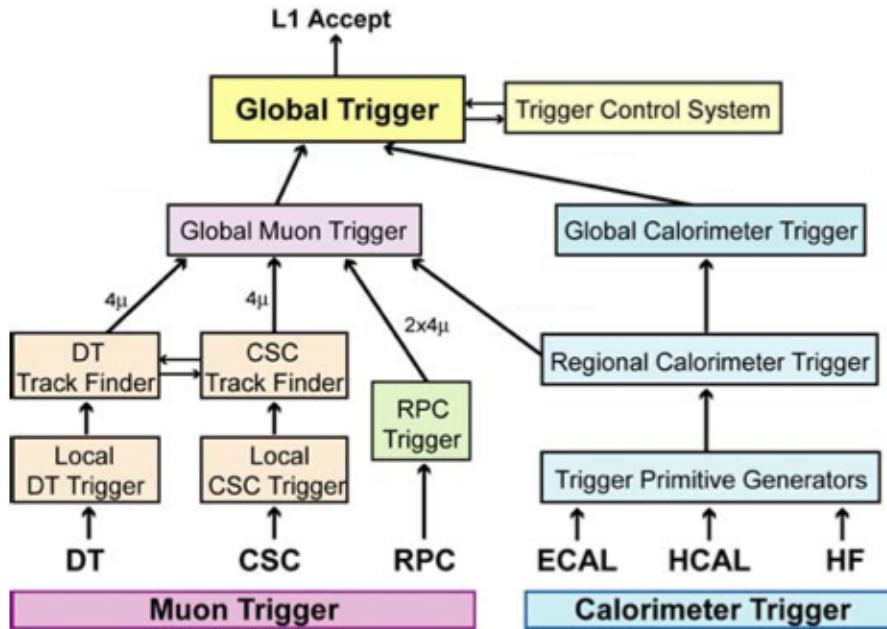


Figure 3.9: Architecture of the Level-1 Trigger [3].

Part I

Search for diboson resonances with CMS

Diboson resonances as signature for new physics

This part of the thesis is dedicated to the description and discussion of searches for new physics in proton-proton collision data collected with the CMS experiment at LHC. As pointed out in Chapter 2, the remarkable compatibility of the discovered scalar resonance by the ATLAS and CMS collaborations with the SM predictions for the Higgs boson, force physicists to deeply understand the role of naturalness in the dynamics of this particle. Several theoretical extensions to the SM have been proposed offering a concrete realization of naturalness, where new particles with masses in the TeV range generate loop corrections with the necessary cancellations to stabilize the Higgs boson mass. This means that the attention can be restricted to direct experimental manifestations of new physics which consist of the production of reasonably narrow new particles. More natural solutions can therefore be probed at the LHC through the direct discovery of these new particles in final states with SM objects with known properties. The research described in this work follows exactly this approach and it is focused on the direct search for new massive resonances ($M_X > 800 \text{ GeV}$) decaying to pairs of vector bosons (WW, WZ, or ZZ) or to a vector boson and a Higgs boson (WH or ZH). These decay modes are preferred for resonances predicted in several BSM models. Popular examples of such models include the bulk scenario of the Randall–Sundrum warped extra-dimensions described in Section 2.3.1, as well as the composite Higgs and Littlest Higgs models discussed in Section 2.3.2. Furthermore, the HVT model (Section 2.3.3) generalizes a large class of explicit models that predict new heavy spin-1 vector bosons, adopting a simplified model strategy. The properties of the described benchmark models studied in this thesis are summarized in Table 4.1.

Table 4.1: Summary of the properties of the heavy resonance models considered in this work. The polarization of the produced W/Z boson in all considered models is mostly longitudinal.

model	particles	spin	charge	main production	main decay
HVT model A, $g_V = 1$	W' singlet	1	± 1	$q\bar{q}^{(\prime)}$	$q\bar{q}^{(\prime)}$
HVT model A, $g_V = 1$	Z' singlet	1	0	$q\bar{q}$	$q\bar{q}$
HVT model A, $g_V = 1$	W' + Z' triplet	1	$0, \pm 1$	$q\bar{q}/q\bar{q}^{(\prime)}$	$q\bar{q}/q\bar{q}^{(\prime)}$
HVT model B, $g_V = 3$	W' singlet	1	± 1	$q\bar{q}^{(\prime)}$	WZ, WH
HVT model B, $g_V = 3$	Z' singlet	1	0	$q\bar{q}$	WW, ZH
HVT model B, $g_V = 3$	W' + Z' triplet	1	$0, \pm 1$	$q\bar{q}/q\bar{q}^{(\prime)}$	WW, WZ, WW, ZH
RS bulk scenario, $\tilde{k} = 0.5$	G_{bulk}	2	0	gg	WW, ZZ

The signal under investigation is a narrow resonance, referring to the assumption that the resonance's natural width is smaller than the experimental resolution, covering a large fraction of the parameter space of the reference models considered. **FIXME: why we do that?** The search considered here is conducted in semi-leptonic final states, where one of the two bosons is a W boson decaying into a charged lepton (ℓ) and a neutrino (ν). The lepton can be either a muon (μ) or an electron (e), however, the results include the $W \rightarrow \tau\nu$ contribution from the decay $\tau \rightarrow \nu\nu$. The gain in sensitivity from τ leptons is limited by the

small branching fractions involved. The second boson in the final state decays into hadrons, and can be either a vector boson $V = W$ or Z , or a Higgs boson. In the first case, the final state is labelled as $\ell\nu q\bar{q}$ including $W \rightarrow q\bar{q}^{(\prime)}$ and $Z \rightarrow q\bar{q}$ decays (Figures 4.1(b), 4.1(c) and 4.1(d)). For the Higgs boson, the final state is labeled as $\ell\nu b\bar{b}$ referring to the Higgs boson decay into a bottom quark-antiquark pair (Fig. 4.1(a)).

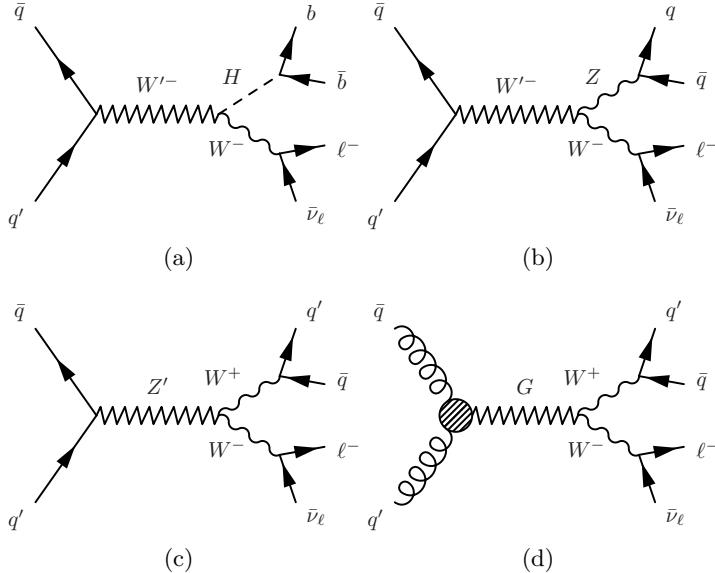


Figure 4.1: Feynman diagrams for the production of a spin-1 charged resonance W' decaying to the $\ell\nu b\bar{b}$ (a) and $\ell\nu q\bar{q}$ (b) final states. The latter final state is also given by the Feynman diagrams for the production of a spin-1 Z' and spin-2 G neutral resonances.

The search in the $\ell\nu b\bar{b}$ final state is based pp collision data at $\sqrt{s} = 8$ TeV collected in 2012 and corresponding to an integrated luminosity of 19.7 fb^{-1} . The second analysis described in this thesis and focused on the $\ell\nu q\bar{q}$ final state is instead based on the pp collision data at $\sqrt{s} = 13$ TeV collected in 2015 and corresponding to an integrated luminosity of 2.3 fb^{-1} . The analysis strategy is similar in the two searches and specific differences will be emphasized throughout the text.

The key challenge of these analyses is the reconstruction of the highly energetic decay products. Since the resonances under study have masses of \approx TeV, their decay products, i.e. the bosons, have on average transverse momenta of several hundred GeV or more. As a consequence, the particles emerging from the boson decays are very collimated. In particular, the jet-decay products of the bosons cannot be resolved using the standard algorithms, but are instead reconstructed as a single jet object. Dedicated techniques, so-called jet “V tagging” and “H tagging” techniques, are applied to exploit the substructure of such jet objects, and can help resolve jet decays of massive bosons. These techniques also help to suppress SM background, which mainly originates from the production of W bosons in association with jets ($W+\text{jets}$).

The aim is to reconstruct the full event to be able to search for a localized enhancement in the invariant mass of the WV or WH system on the top of a smoothly falling SM background distribution. The background mainly comprises $W+\text{jets}$ production, while another significant contribution is represented by events involving pair produced top quarks ($t\bar{t}$). Other minor backgrounds are represented by single top quark and SM diboson (WW , WZ or ZZ) production processes.

The invariant mass of the WV and WH system is determined by estimating the neutrino

transverse momentum with the measured missing transverse energy (E_T^{miss}) in the event, while an estimate of the neutrino longitudinal momentum is derived by imposing the constraint of the W mass on the invariant mass of the $\ell\nu$ system. In the following the diboson invariant mass will be labelled either $m_{\ell\nu j}$ system, or m_{WV} and m_{WH} for the $\ell\nu q\bar{q}$ and $\ell\nu b\bar{b}$ final states, respectively. The mass spectrum for the dominant W+jets background is determined from observed events with a reconstructed jet mass not compatible with the V or H hypothesis. This strategy partially relies on the simulation of the background processes. Furthermore, simulated events are used for the optimization of the analysis selection aimed at maximizing the discrimination of the signal against the background and hence the analysis sensitivity.

This part of the thesis is organized as follows. Chapter 5 gives an overview of the methods used to simulate the physics processes happening in pp collisions at the LHC together with a description of the specific simulated background and signal events used in these analyses, as well as a discussion about the data sets analyzed. Chapter 6 provides a detailed description of the algorithms used in CMS for the reconstruction of the event and of the physics objects expected in the semi-leptonic final states under investigation. Particular attention is given to the V and H tagging algorithms representing the key feature of these analyses and separately discussed in Chapter 7. The analysis strategy, already outlined here, is explained in details in Chapter 8. This includes the final event selection and categorization optimized to enhance the analysis sensitivity, as well as the strategy for the estimation of the expected background, the modelling of the signal and the related systematic uncertainties which will be used as input to the statistical analysis of the diboson invariant mass distribution observed in data. The final results are discussed in Chapters 9 and 10 for the 8 and 13 TeV data analysis, respectively.

Eventually, the results of these analyses are combined with limits derived in companion CMS searches for resonances decaying to a pair of bosons in several different final states during both LHC Run 1 and Run 1. These analyses use the same V and H tagging techniques as presented here to separate the signal from the large multijet or V+jets background. The statistical combination represent the last piece of this work and it is presented in Chapter 11.

Data sets and simulated samples

The simulation of proton-proton collisions is usually performed by means of Monte Carlo (MC) event generators, providing an accurate modelling of the event kinematics and topology at parton and hadron level. The hard inelastic scattering has to be fully calculated, namely, from the hard interaction between the partons inside the protons, where perturbative QCD calculations (**FIXME: ref**) can be used, to the formation of particle jets from the outgoing partons. Furthermore, it is fundamental to understand the exact response of the detector to the outgoing particles produced in proton-proton collisions. Consequently, the stable outgoing particles are fed to a full detector simulation that models the interaction of those particles with the detector material and the corresponding detector response. The raw detector data are then subject to the same reconstruction algorithms that are also used for real data. In this chapter, MC event generators are described in detail, followed by a brief description of the CMS detector simulation.

5.1 Simulation of proton-proton collisions

5.1.1 Monte Carlo event generators

The generation of hard inelastic proton-proton collisions is factorized into different steps ordered by the timescale on which they happen, as illustrated in Fig. 5.1, and described in the following.

The basis of theoretical event generation at the LHC is a parametrisation of the incoming partons (quarks, anti-quarks and gluons) stemming from the proton, which is given by the parton density functions (PDF). They describe the probability to find a quark or gluon with a given proton momentum fraction x from the pp collision taking place at the LHC. In QCD the PDFs depend on a factorization scale μ_F^2 at which the proton is probed. All interactions between quarks and gluons happening at scales below the scale μ_2^F are absorbed into the PDFs. Therefore at small μ_2^F the proton is observed basically as a combination of its three valence quarks uud . At higher scales, however, it is dominated by sea quarks and gluons.

A collision between two partons, one from each side, gives the hard process of interest, which can be due to physics within or beyond the standard model. Using the incoming partons as input, the simulation of the hard process is performed by the event generator. It produces hypothetical events with the distributions and rates predicted by theory based on the cross section formulae of the physics process of interest. Using the cross section formula the phase space is sampled and candidate events are defined by choosing values for the degrees of freedom from a uniformly distributed random number generator.

The cross section can be calculated by means of the so called *factorization theorem* [15]. According to the theorem, the hadron itself is described by the whole particle composition interacting on a soft binding energy scale, whereas the collisions occur between the partons on a hard energy scale with large transverse momenta. The cross section for the process is then given by the convolution of the PDF, $f_i(x, Q^2)$ integrated over the proton momentum

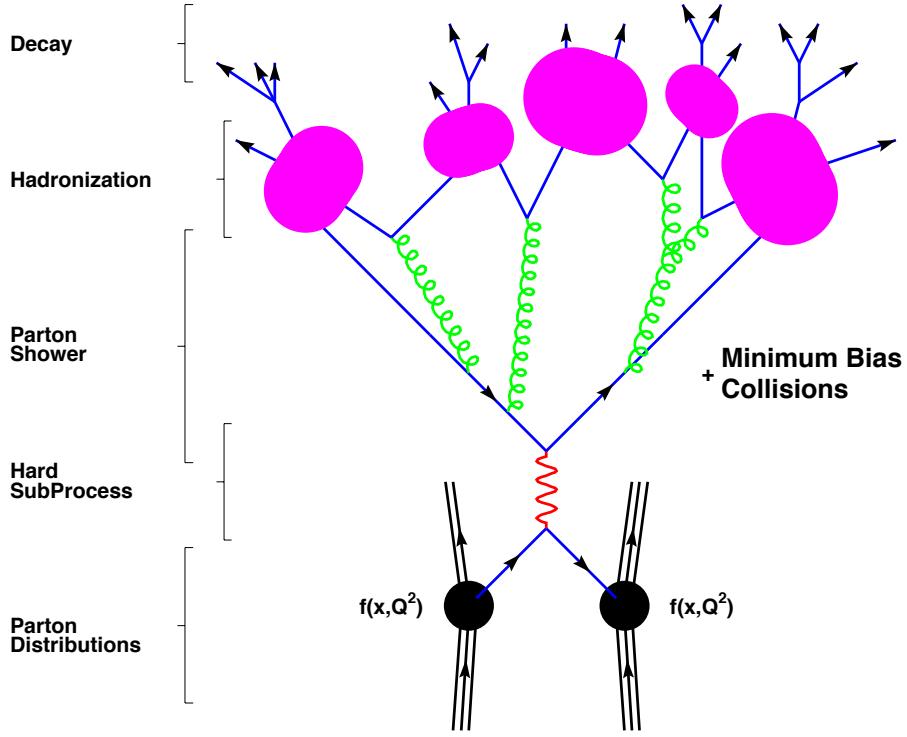


Figure 5.1: Steps of Monte Carlo event generation as described in the text evolving in time from bottom to top [14].

fraction x , for the colliding protons (A, B) at an energy scale Q^2 and the hard parton-parton cross sections $\hat{\sigma}_{ij} \rightarrow X$ of all combinations of two partons i and j:

$$\sigma(AB \rightarrow X) = \sum_{q,g=0}^n \alpha_S^n(\mu_R^2) \sum_{ij} \int dx_i dx_j f_{i,A}(x_i, \mu_F^2) f_{j,B}(x_j, \mu_F^2) \cdot \hat{\sigma}_{ij \rightarrow X}^{(n)}(s; x_i, x_j, \mu_R^2, \mu_F^2) \quad (5.1)$$

where the index n runs over the perturbative order and s is the squared center of mass energy of the collision. The tree-level process, where no emission of gluons or quarks holds, is called “Leading Order” (LO) and takes place when $n = 0$. Further orders are called “Next-to-Leading Order” (NLO, $n = 1$), “Next-to-Next-to-Leading Order” (NNLO, $n = 2$) and so on.

As it can be seen from the formula, the PDFs play a fundamental role in the description of the hard process, and it is very important to have several experimental tests accessible for their values. Perturbative QCD cannot predict PDFs, since they contain also the low energy (non-perturbative) information about the scattering. As a consequence, PDFs distributions are extracted from data, in deep-inelastic scattering experiments. Most of the parametrizations of proton PDFs now used for the LHC have been extracted from the ZEUS [16] and H1 [17] experiments in electron-proton collisions at the HERA collider and fixed target experiments. The more recent parametrizations also take into account vector boson production and single-inclusive jet production from the Tevatron experiments, as well as LHC data. Once measured for a certain momentum fraction x_i at an energy scale Q^2 , they can be extrapolated to another scale using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) evolution equation [18]. The PDF sets used for signal simulation samples in these analyses

are provided by the CTEQ/CT group [19, 20]. This set especially incorporates the effects of Tevatron Run I jet production data on the gluon distribution and is therefore expected to describe the mainly gluon based LHC processes realistically. The CT sets additionally include combined data set of HERA-1 data and new data on the asymmetry in the rapidity distribution of the charged lepton from W boson decay from CDF and rapidity distributions of Z bosons from both CDF and D0. The NNPDF sets [21] are calculated with an approach based on neural network and the newest versions include LHC data as well. An example of the most important parton distributions inside the proton is shown in Fig. 5.2.

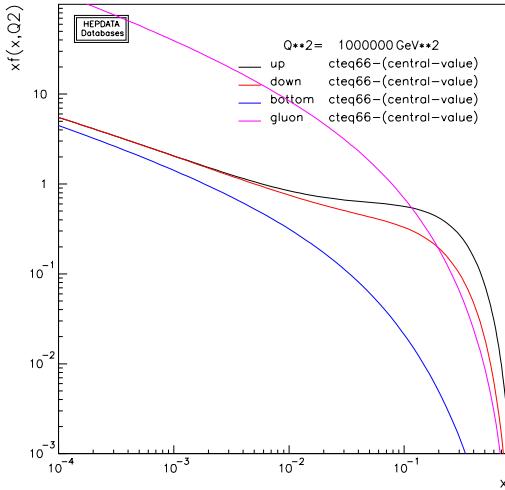


Figure 5.2: CTEQ6.6 central value parton distribution functions at the typical mass scale of a new diboson resonance ($Q^2 = (1000 \text{ GeV})^2$) for up, down and bottom quarks and gluons in the proton in double-logarithmic scale.

An accurate description of the process must take into account radiative corrections to the tree-level or LO description of the process of interest. In particular, one has to include the effects of real and virtual higher-order corrections in perturbation theory. This is achieved by computing the matrix element between the initial and final states as the sum of contributions with increasing powers of $\alpha_S^n(\mu_R^2)$. For instance, the LO contribution to the W boson production process can be calculated from the diagram in Fig. 5.3. The diagrams contributing at NLO to this process and corresponding to the real and virtual radiative corrections at the first order are shown in Fig. 5.4.

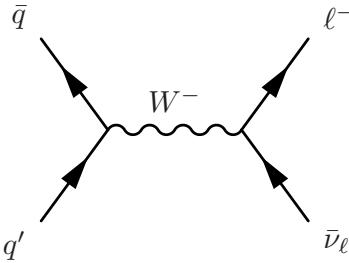


Figure 5.3: (top) Feynman diagram contributing to the W boson production at leading order.

Perturbative calculations in QCD are limited to processes in which the coupling constant $\alpha_S^n(\mu_R^2)$ is small, and by the complexity of higher order calculations preventing their evaluation. Consequently, the current generators are only able to treat a limited number of partons in

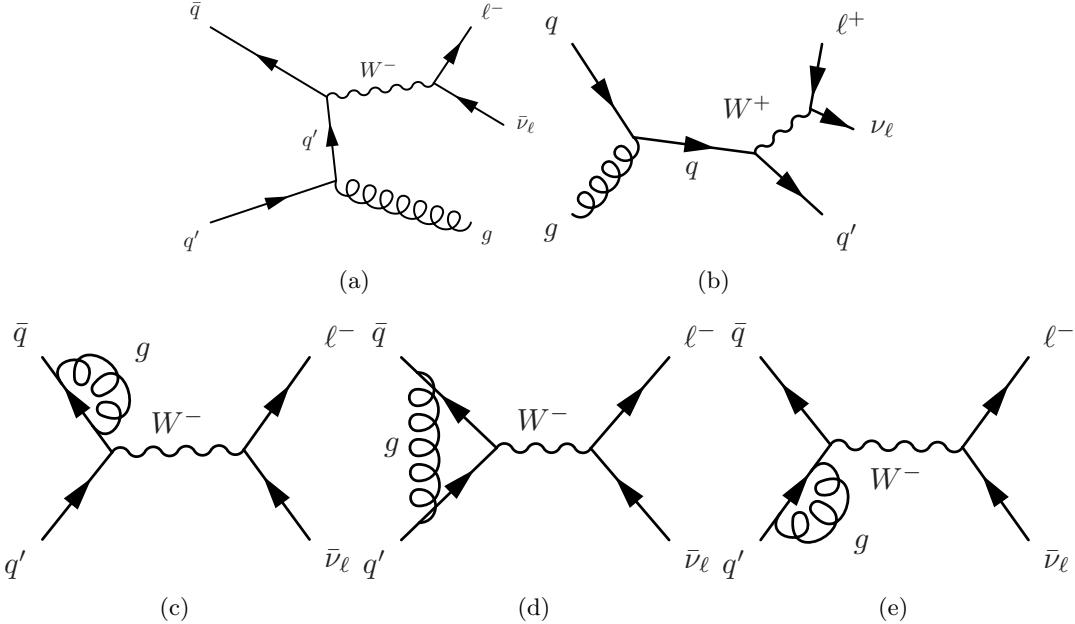


Figure 5.4: Feynman diagrams contributing at next-to-leading order to the W boson production and corresponding to the first order real (top) and virtual (boson) radiative corrections. The charge conjugate production modes are implied. Only the leptonic decay of the W boson is considered.

the final state. Parton showering algorithms extend the fixed order calculations beyond these limiting factors by calculating emissions of additional partons from the in- and outgoing partons of the main interaction. This approach in principle takes into account emissions of an unlimited number of partons, but, as opposed to full higher order calculations, does not take into account loop diagrams.

Parton showering algorithms start from the hard process allowing the partons to split (or branch) into pairs of other partons. These again may also branch and so on, so that an event then consists of a large number of elementary particles, including quarks and gluons. The cascade of splittings is stopped once the energy scale reaches values where the coupling constant $\alpha_S^n(\mu_R^2)$ becomes large.

At this stage, quarks and gluons, which carry colour, cannot be considered as free anymore and recombine to form neutral hadrons, through the so called hadronization process. The formation of color-neutral hadrons from the colored partons is treated in phenomenological non-perturbative models. Eventually, many short-lived resonances will be present after hadronization which are then decayed.

The showering and hadronization programs often bring along the possibility to add underlying events. The underlying event arises from the coloured remains of the protons that did not take part in the hard collisions, the so-called beam remnants. They are usually included in the hadronization process, because they might be colour-connected to the hard subprocess. The produced hadrons will however carry a very small transverse momentum and will be very forward. The probability for colour reconnection to take place between two partons can also be adjusted based on experimental data. It is also possible that more than one parton interact with the other proton. This phenomenon, called multiple parton interaction, and it is usually added to the description of the process.

As last step the pileup due to additional pp collisions in the same bunch crossing is also accounted for. Additional simulated minimum-bias interactions are added to the generated

events to match the additional particle production due to pileup. The exact number of average collisions per bunch crossing in the data is estimated by multiplying the instantaneous luminosity, continuously monitored, by the total inelastic cross section. One can then reconstruct the distribution of the number of pileup interactions in the data for the complete data taking. The corresponding distributions for the 2012 and 2015 data are shown in Figures 5.5(a) and 5.5(c), respectively, together with the corresponding simulated pileup scenarios. Simulated events are then reweighted such that they match the data distribution. The description of the pileup by the simulation can be verified by counting the number of reconstructed vertices in the event as illustrated in Figures 5.5(b) and 5.5(d).

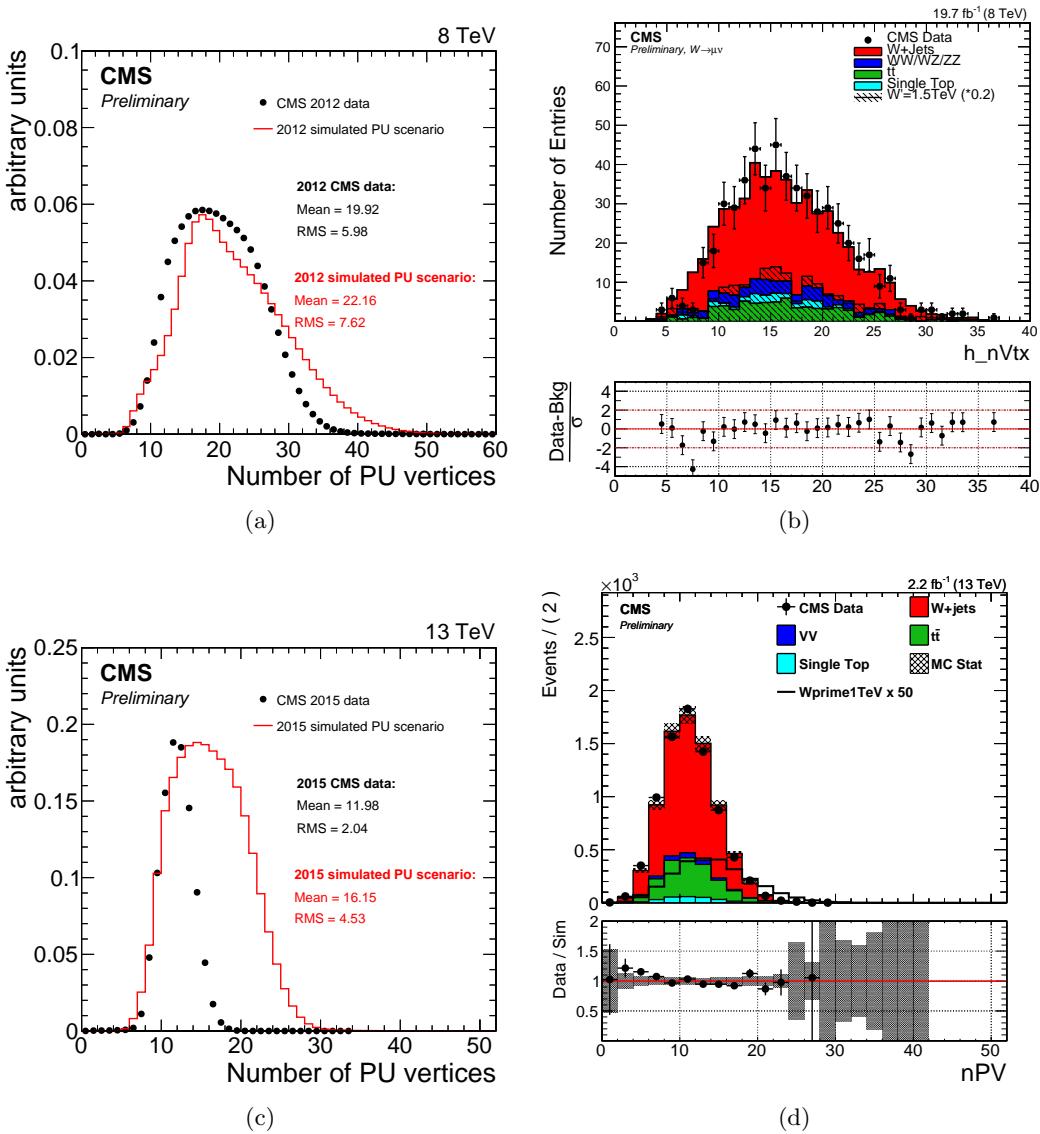


Figure 5.5: blabla.

Currently, NLO available calculations included in MC event generators cover a wide range of physics processes, starting with two particles annihilation to a maximum of five final state objects. A popular generator is PYTHIA [22, 23], a general purpose program which, in addition to the hard process, also takes care of the parton showering, the hadronization,

and the description of the underlying event. For the matrix element calculation, PYTHIA only considers the leading order hard subprocess (diagram in Fig. 5.3 for the W production case), and higher order effects are added by “evolving” the event using the parton shower. A more accurate approach is followed by MADGRAPH [24] where the hard real radiative corrections are included in the matrix element (Fig. 5.4). This generator is well suited to study processes such as W or Z produced in association with hard jets. Since it does not completely simulate the events, it needs an additional program, typically PYTHIA, to perform the parton shower after the computing of the matrix element. It has to be noted that matrix element generators as well as shower and hadronization generators are usually treated independently: the matrix element generators compute the hard process at fixed-order and the parton shower processes the soft and collinear emissions. However, this fails to correctly represent higher order processes in which an additional parton is emitted at the hard scale because parts of this process overlap with the soft one. Combining an NLO matrix element program with a parton shower program therefore leads to double-counting of events. The NLO matrix element generators, such as POWHEG [25] and MC@NLO [26], take special care of the merging of soft and collinear emissions and hard ones. They allow to calculate any amplitudes at NLO accuracy, and include the matching to the parton shower to produce events ready for the hadronization.

5.1.2 CMS detector simulation

For a detailed understanding on how interactions in pp collisions at the LHC are observed by the CMS detector, a dedicated simulation of the whole detector is needed. Both the propagation of particles through the detector material as well as the response of the active detector components and their digital output need to be simulated. The input to the detector simulation are collections of particles produced by MC event generators. The output is the digital signal from all detector components in the same format that is used for real data.

The CMS simulation is based on the GEANT4 [27] toolkit. The program calculates the trajectory of the various particles generated during the collision, simulates their electromagnetic and hadronic interaction with the crossed material and the signal they will produce in the various subdetectors. The detector geometry is given as an input to the program and to obtain a description as close as possible to the reality, any available information such as the existence of insensitive materials or dead channels and their position, is included. The electronic readout of the hits produced by particles is simulated, taking into account resolution and detector response effects. The same code is then used to reconstruct the various physical objects described in Chapter 6.

5.2 Simulated samples

5.2.1 Simulation of signal processes

For the 8 TeV data analysis, the signal hypothesis has been simulated at LO accuracy with a W' boson produced via quark-antiquark annihilation and decaying into W and Higgs bosons in the $\ell\nu q\bar{q}$ final state with $q = b, c$ or g and $\ell = e, \mu$ or τ . Resonance masses in the range 0.8–2.5 TeV are considered in this analysis. The events are generated at parton level using a model of a generic narrow spin-1 W' resonance implemented with MADGRAPH. Showering and hadronization are performed using PYTHIA6 using the Z2* tune [28, 29]. The full simulation of the detector has been done privately following the standard CMS procedure described in Section 5.1.2. This emulation of the official procedure has been validated comparing the private production with the official MC campaign usually carried out in the

collaboration. Furthermore, it has been verified that the kinematic distributions obtained with the implementation of the generic model agree with those predicted by implementations of the LH, composite Higgs and HVT models in MADGRAPH. The resonance width differs in the three models, but in each case it is found to be negligible with respect to the experimental resolution.

The following parameters are used to compute the cross sections: $g_V = 3$, $c_H \simeq -1$, and $c_F \simeq 1$ in the HVT model B and $\cot 2\theta = 2.3$, $\cot \theta = -0.20799$ in the LH model, where θ is a mixing angle parameter that determines W' couplings ([FIXME: ref](#)) and that $\cot 2\theta$ and $\cot \theta$ can be directly related to c_H and c_F .

The intrinsic width and cross section for both models are listed in Table 5.1 for the resonance masses considered. The widths for the HVT model B are computed by means of Equation (2.31) in Ref. [30][FIXME: ref to the theory chapter](#), while the cross sections were obtained using the online tools provided by the authors of Ref. [30].

Table 5.1: Intrinsic total widths (Γ) and cross sections (σ) for the LH model and HVT model B for different resonance masses. The $WH \rightarrow \ell\nu b\bar{b}$ branching fraction is not included in the calculation.

Resonance mass [TeV]	LH model		HVT model B	
	Γ [GeV]	σ [pb]	Γ [GeV]	σ [pb]
0.8	7.22	5.09×10^{-1}	24.1	3.37×10^{-1}
0.9	8.12	3.03×10^{-1}	27.1	2.48×10^{-1}
1.0	9.02	1.87×10^{-1}	30.1	1.71×10^{-1}
1.1	9.92	1.18×10^{-1}	33.1	1.16×10^{-1}
1.2	10.8	7.65×10^{-2}	36.1	8.05×10^{-2}
1.3	11.7	5.06×10^{-2}	39.1	5.59×10^{-2}
1.4	12.6	3.39×10^{-2}	42.2	3.88×10^{-2}
1.5	13.5	2.29×10^{-2}	45.2	2.51×10^{-2}
1.6	14.4	1.56×10^{-2}	48.2	1.87×10^{-2}
1.7	15.3	1.08×10^{-2}	51.2	1.30×10^{-2}
1.8	16.2	7.43×10^{-3}	54.2	9.03×10^{-3}
1.9	17.1	5.17×10^{-3}	57.2	6.27×10^{-3}
2.0	18.0	3.61×10^{-3}	60.2	4.25×10^{-3}
2.1	19.0	2.53×10^{-3}	63.2	3.02×10^{-3}
2.2	19.8	1.76×10^{-3}	66.2	2.10×10^{-3}
2.3	20.8	1.24×10^{-3}	69.2	1.46×10^{-3}
2.4	21.6	8.67×10^{-4}	72.2	1.01×10^{-3}
2.5	22.6	6.07×10^{-4}	75.3	7.31×10^{-4}

Figure 5.6 shows the ratio between the resonance natural width and mass for a W' in the Little Higgs and the HVT model B. The width is less than 5% for the following parameter values: $0.95 < g_V < 3.76$, $c_H = -1$, and $c_F = 1$; $g_V < 3.9$, $c_H = -1$, and $c_F = 0$; or $g_V < 7.8$, $c_H = 0.5$, and $c_F = 0$. The widths for the LH model have been computed by means of Eq. (15) in Ref. [31], and they are less than 5% for values of $0.084 < |\cot \theta| < 1.21$. Hence, in both models the width can be considered to be negligible compared to the experimental resolution.

For the 13 TeV data analysis, the bulk graviton model and HVT models are used as benchmark signal processes. In these models, a resonance is simulated which decays only to pairs of vector gauge bosons in the $\ell\nu q\bar{q}^{(\prime)}$ final state, with $\ell = e, \mu$, and τ . The vector gauge bosons are produced with a longitudinal polarization in more than 99% of the cases. For each resonance hypothesis, masses are considered in the range 0.6 to 4.0 TeV. Simulated signal events are generated at LO accuracy with MADGRAPH5_AMC@NLO with a relative resonance

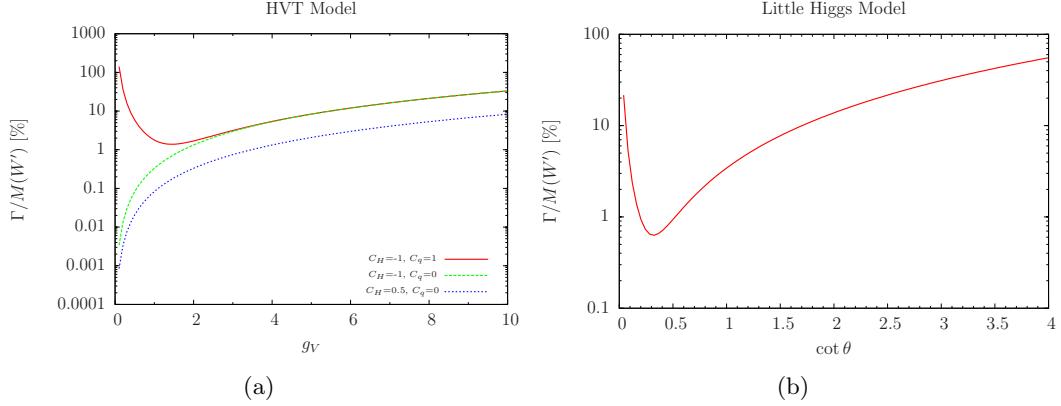


Figure 5.6: Ratio between the resonance natural width Γ and mass for a W' in the Little Higgs and the HVT model B.

width of 0.1%.

The natural width of for a bulk graviton as a function of the curvature parameter \tilde{k} and for different mass hypotheses is shown in Fig. 5.7. For cases in which $\tilde{k} \leq 0.5$ the relative width of the graviton resonance (Γ_G/M_G) is predicted to be below 1%. Hence, it can be neglected when compared to the detector resolution over the whole explored mass range.

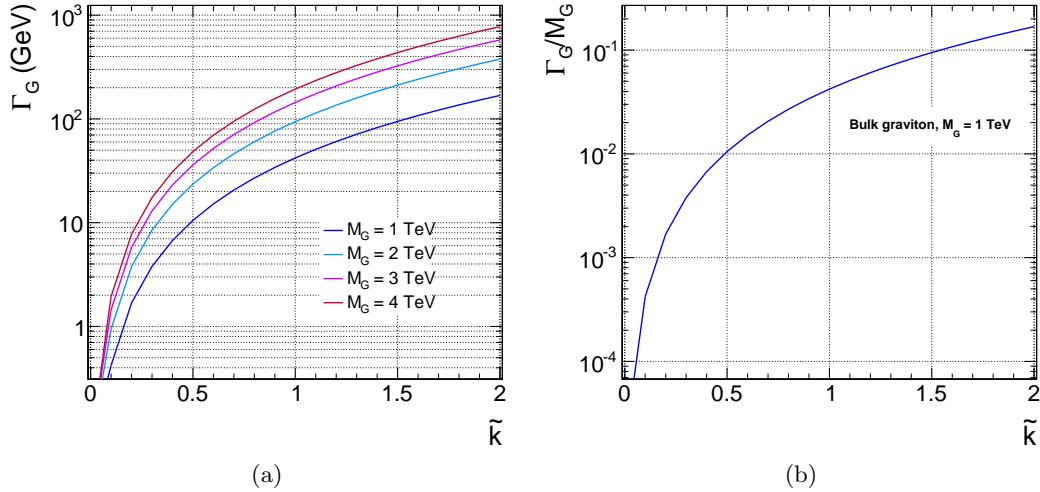


Figure 5.7: (a) Natural width of a bulk graviton as a function of the coupling constant \tilde{k} and for various mass hypotheses. (b) The same dependence is expressed as relative fraction of the signal width with respect to a reference graviton mass of 1 TeV.

Figure 5.8 compare the production cross sections $\sigma(pp) \rightarrow X$ of the resonance for $\sqrt{s} = 8$ and 13 TeV, for the bulk graviton with $\tilde{k} = 0.5$, and W' and Z' in the HVT model B, as a function of the resonance mass. Cross sections for the bulk graviton model are computed with MADGRAPH with the model used for the even generation, while values for the HVT model B are obtained using the online tools provided by the authors of Ref. [30] using the same parameters as for the 8 TeV data analysis.

For a resonance mass of 2 TeV, the production rates at for $\sqrt{s} = 13$ TeV are expected to

increase of a factor ≈ 17 for a resonance produced via gluon-gluon fusion such as the graviton; a smaller factor of ≈ 7 is expected instead for resonances produced via quark-antiquark annihilation such as W' and Z' .

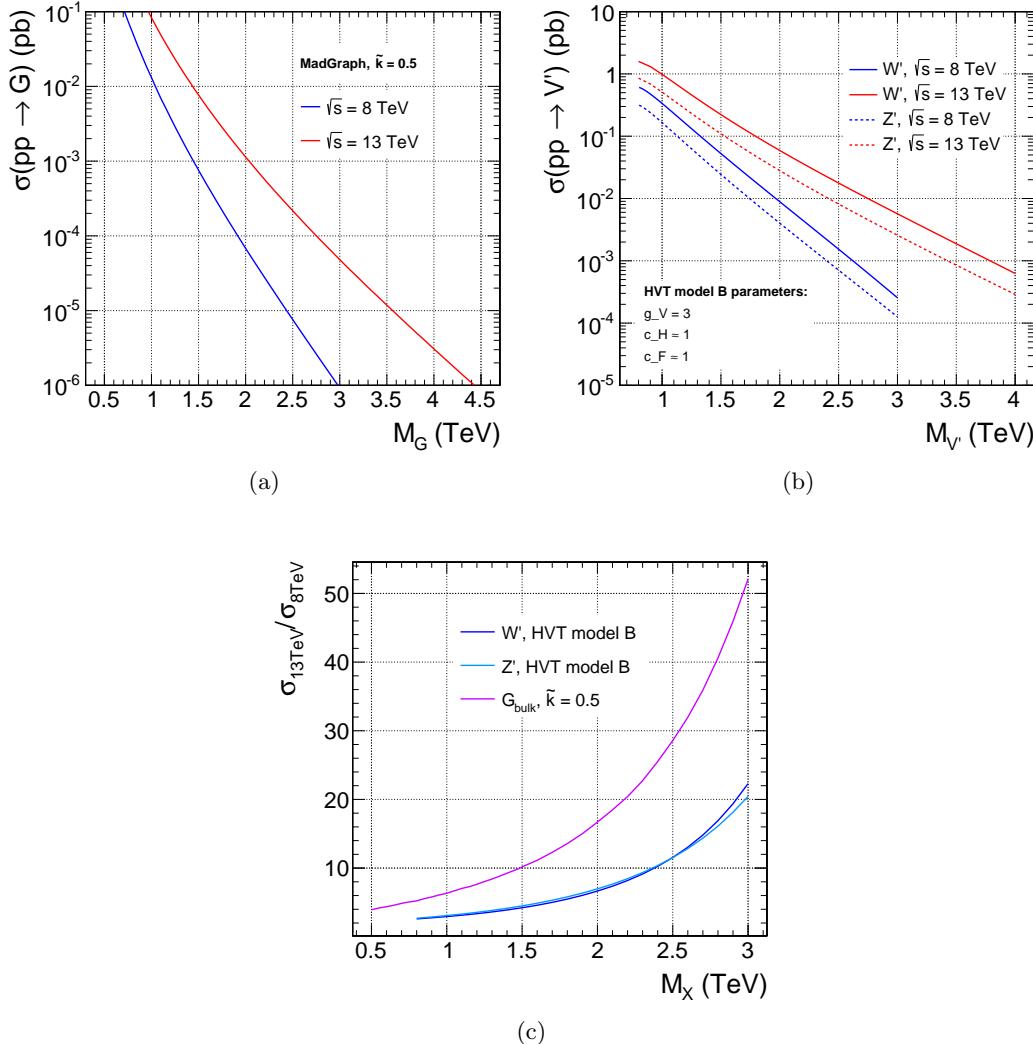


Figure 5.8: Comparison of the production cross sections of the resonance for $\sqrt{s} = 8$ and 13 TeV for the bulk graviton (a), and W' and Z' (b) in the HVT model B, as a function of the resonance mass. (c) Ratio of the production cross sections for $\sqrt{s} = 8$ and 13 TeV for all models.

5.2.2 Simulation of background processes

For the 8 TeV data analysis, the background is modelled using the MADGRAPH5 v1.3.30 event generator to simulate the production of W boson in association with jets at LO, the POWHEG 1.0 r1380 package to generate $t\bar{t}$ and single top quark events at NLO accuracy, and PYTHIA6 v424 for SM diboson (WW , WZ , and ZZ) production at LO. All simulated event samples are generated using the CTEQ6L1 PDF set with α_S also at leading order, except for the POWHEG $t\bar{t}$ sample, for which the CT10 NNLO PDF set is used. All the samples are then processed further by PYTHIA6, using the Z2* tune for simulation of parton showering and subsequent hadronization, and for simulation of the underlying event. All simulated

background samples are normalized to the integrated luminosity of the recorded data, using inclusive cross sections determined at next-to-leading order, or next-to-next-to-leading order when available, calculated with the cross section integrators MCFM [32–35] and FEWZ [36], except for the $t\bar{t}$ sample, for which TOP++ [37] is used. The NNLO cross section for the W+jets process is obtained by rescaling the LO value given by the generator to the NNLO cross section derived from the inclusive production by means of a flat k -factor = NNLO/LO = 1.3. The simulated samples used in the 8 TeV data analysis described in this work are listed in Table 5.2 together with the corresponding cross sections.

Table 5.2: Summary of the MC generated samples for background processes used for the 8 TeV data analysis. The cross sections used to normalize the samples are also indicated.

Process	Cross section (pb)	Generator	PDF set
W+jets, $W \rightarrow \ell\nu$, $p_T^W > 180$ GeV	29.0 (NNLO)	MADGRAPH	CTEQ6L1
$t\bar{t}$ (inclusive)	252.9 (NNLO+NNLL)	POWHEG	CT10
single t quark (t-channel, inclusive)	54.9 (NNLO)	POWHEG	CTEQ6L1
single \bar{t} quark (t-channel, inclusive)	29.7 (NNLO)	POWHEG	CTEQ6L1
single t quark (tW-channel, inclusive)	11.2 (NNLO)	POWHEG	CTEQ6L1
single \bar{t} quark (tW-channel, inclusive)	11.2 (NNLO)	POWHEG	CTEQ6L1
single t quark (s-channel, inclusive)	3.8 (NNLO)	POWHEG	CTEQ6L1
single \bar{t} quark (s-channel, inclusive)	1.8 (NNLO)	POWHEG	CTEQ6L1
WW (inclusive)	54.8 (NLO)	PYTHIA6	CTEQ6L1
WZ (inclusive)	33.2 (NLO)	PYTHIA6	CTEQ6L1
ZZ (inclusive)	8.1 (NLO)	PYTHIA6	CTEQ6L1

For the 13 TeV analysis, the W+jets SM process is simulated with MADGRAPH5_AMC@NLO at LO accuracy, The $t\bar{t}$, single top quark and diboson events are generated with both POWHEG and MADGRAPH5_AMC@NLO at NLO accuracy. Parton showering and hadronization are implemented through PYTHIA8 using the CUETP8M1 tune [38,39]. The NNPDF 3.0 PDFs with α_S at NLO, are used for all simulated samples. The simulated background is normalized using inclusive cross sections calculated at NLO, or NNLO order in QCD where available, using MCFM and FEWZ, except for the $t\bar{t}$ sample, for which TOP++ [37] is used. The NNLO cross section for the W+jets process is obtained by rescaling the LO value given by the generator to the NNLO cross section derived from the inclusive production by means of a flat k -factor = NNLO/LO = 1.21.

The simulated samples used in the 13 TeV data analysis described in this work are listed in Table 5.3 together with the corresponding cross sections.

5.3 Data sets

Two independent dataset are analyzed in this work to search for diboson resonances decaying to two different final states.

The analysis focused on the $\ell\nu b\bar{b}$ final state is performed with the complete dataset recorded in 2012 by the CMS detector and corresponding to an integrated luminosity of 19.7 fb^{-1} of proton-proton collisions at $\sqrt{s} = 8$ TeV. The recorded events are divided into 4 run periods (runs A, B, C, D).

Table 5.3: Summary of the MC generated samples for background processes used for the 13 TeV data analysis. The cross sections used to normalize the simulated events are also indicated. The NNPDF 3.0 PDFs are used for all simulated samples

Process	Cross section (pb)	Generator
W+jets, $W \rightarrow \ell\nu$, $100 < H_T < 200$ GeV	1627.5 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$, $200 < H_T < 400$ GeV	435.2 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$, $400 < H_T < 600$ GeV	59.2 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$, $600 < H_T < 800$ GeV	14.6 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$, $800 < H_T < 1200$ GeV	6.7 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$, $1200 < H_T < 2500$ GeV	1.6 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$, $H_T > 2500$ GeV	0.04 (NNLO)	MADGRAPH5_AMC@NLO
$t\bar{t}$ (inclusive)	831.8 (NNLO+NNLL)	POWHEG
single t quark (t-channel), $W \rightarrow \ell\nu$	44.5 (NNLO)	POWHEG
single \bar{t} quark (t-channel), $W \rightarrow \ell\nu$	26.5 (NNLO)	POWHEG
single t quark (tW-channel, inclusive)	35.9 (NNLO)	POWHEG
single \bar{t} quark (tW-channel, inclusive)	35.9 (NNLO)	POWHEG
single $t+\bar{t}$ quark (s-channel), $W \rightarrow \ell\nu$	3.7 (NNLO)	MADGRAPH5_AMC@NLO
$WW \rightarrow \ell\nu q\bar{q}'$	50.0 (NNLO)	POWHEG
$WZ \rightarrow \ell\nu q\bar{q}$	10.7 (NLO)	MADGRAPH5_AMC@NLO
$ZZ \rightarrow \ell\ell q\bar{q}$	3.22 (NLO)	MADGRAPH5_AMC@NLO

The second analysis described in this work is focused on the $\ell\nu q\bar{q}'$ final state and it is performed with only a part of the full dataset recorded in 2015 by the CMS detector corresponding to an integrated luminosity of 2.3 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$. During 2015, there have been three running periods labeled from B to D. In fact, after a short period of 50 ns operation (period B), the machine collected data with a bunch spacing of 25 ns (period C and D). However, since the first two periods only add a tiny contribution to the total integrated luminosity of 2015 collisions, the decision was made to base the analysis on period D only, corresponding to the largest data set.

All events that are accepted by a specific set of high level triggers enter one specific dataset, so that the choice of a trigger for the analysis defines which dataset has to be used. As discussed in the next Chapter (Sections 6.3.2 and 6.2.2), events are collected with a trigger requiring either one muon or one electron passing given p_T and η selections. Hence, the datasets used in these analyses are the so called “SingleMuon” and ”SingleElectron” primary data sets listed in Table 5.4.

Even though stable run periods are chosen for the analyses, not all runs can be used. These analyses require the whole detector to be functional since the objects employed are reconstructed from all parts of the detector as discussed in the next Chapter. Therefore, only data-taking runs and luminosity blocks during which the detector was in a state sufficiently good for further analysis are used.

Table 5.4: Data sets used in these analyses.

\sqrt{s}	Year	Data set	Run period	Run range	$\mathcal{L} [\text{pb}]^{-1}$
8 TeV	2012	SingleMuon	A	190456–193621	889.362
			B	193833–196531	4424
			C	198022–203742	7144
			D	203777–208686	7307
		Total		190456–208686	19764
	2012	SingleElectron	A	190456–193621	889.362
			B	193833–196531	4422
			C	198022–203742	7080
			D	203777–208686	7314
		Total		190456–208686	19705
13 TeV	2015	SingleMuon	D	256630–260627	2320
		SingleElectron	D	256630–260627	2320

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 τ -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 [40] 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 ϕ and polar angle θ , 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) [41], 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 [42, 43] 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 χ^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 χ^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 $1 < p_T < 100 \text{ GeV}$, the track reconstruction efficiency is $> 99\%$ over the full η -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

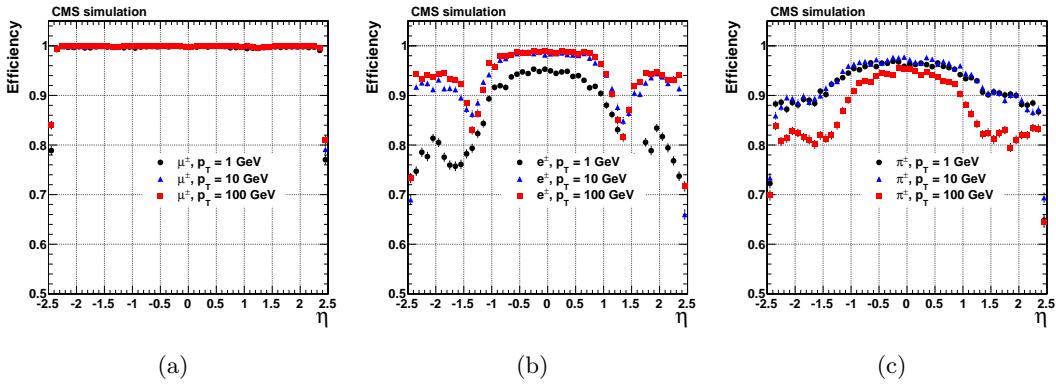


Figure 6.1: Track reconstruction efficiency for simulated muons (a), electrons (b), and pions (c) passing the high-purity quality requirements as a function of η and for $p_T = 1, 10$, and 100 GeV [40].

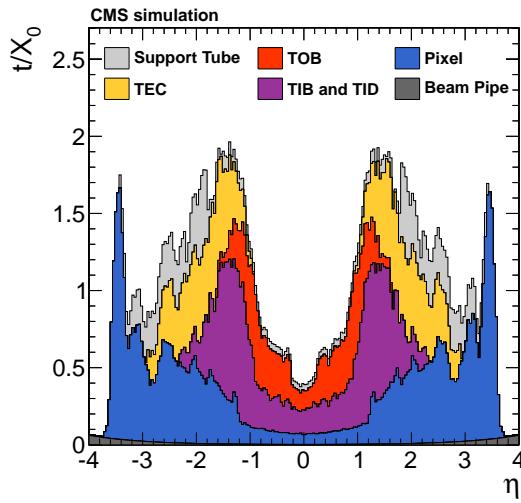


Figure 6.2: Material budget of the CMS tracker in units of radiation length X_0 as a function of pseudorapidity divided into the contributions of the different subdetectors [40].

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 [44], 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 χ^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 [45]. This clustering allows

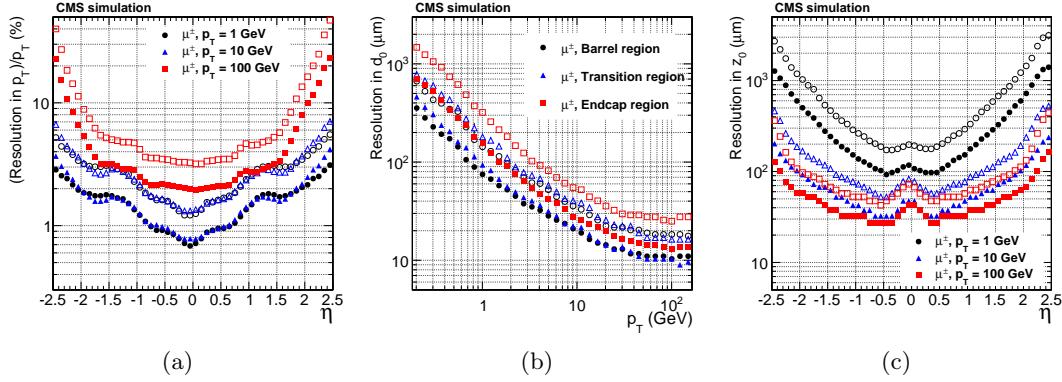


Figure 6.3: Resolution of track transverse momentum (a), transverse (b) and longitudinal (c) impact parameter for simulated muons passing the high-purity quality requirements as a function of η and for $p_T = 1, 10$, and 100 GeV [40].

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* [46] 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.

In the analyses described in this work, all events are required to have at least one primary vertex reconstructed within a 24 cm window along the beam axis, with a transverse distance from the nominal pp interaction region of less than 2 cm.

6.2 Electrons

6.2.1 Electron reconstruction

The electron reconstruction in CMS [48] 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

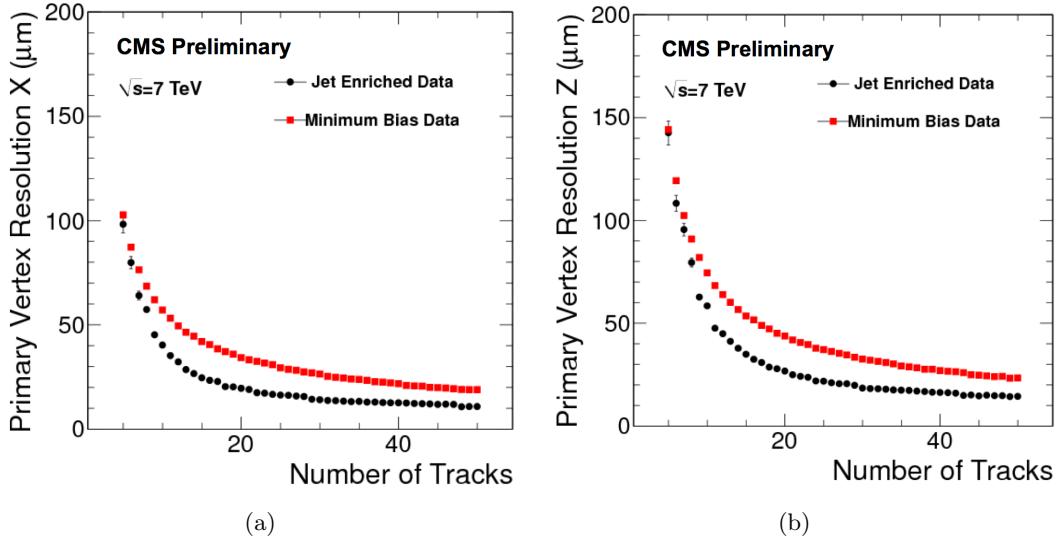


Figure 6.4: Primary-vertex resolution in x (a) and z (b) as a function of the number of tracks at the fitted vertex, for two kinds of events with different average track p_T values; the results in y is almost identical to the one in x [47].

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 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 ϕ coordinate. To recover this radiated energy, ECAL superclusters are formed, by merging clusters of similar η over some range of ϕ . 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 χ^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 χ^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) [49]. 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 η and ϕ 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 θ 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 [50]. 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 5×5 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

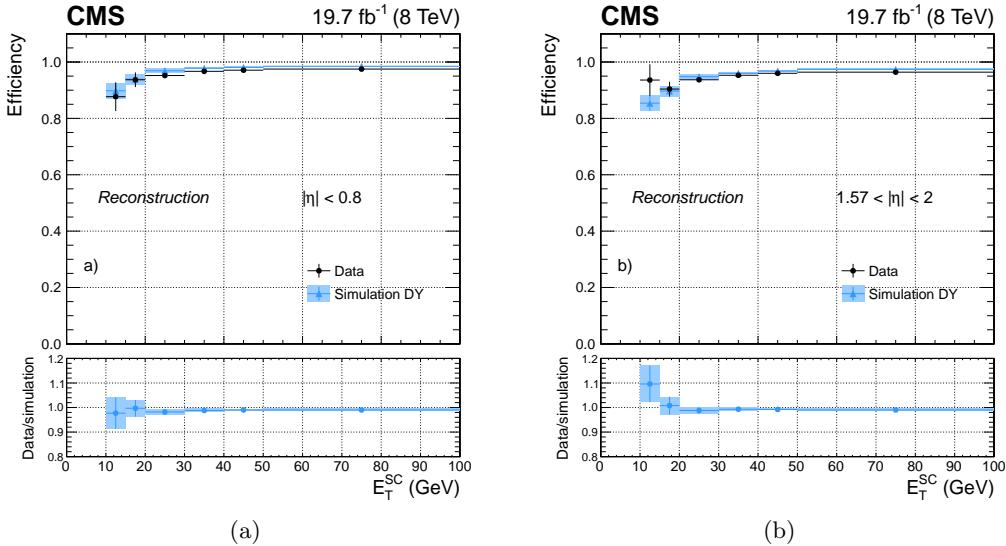


Figure 6.5: Electron reconstruction efficiency measured in dielectron events in data (dots) and Drell-Yan simulation (triangles), as a function of the E_T of the electron supercluster for electrons reconstructed in the ECAL barrel (a) and endcaps (b). The bottom panels show the corresponding data-to-simulation scale factors [51].

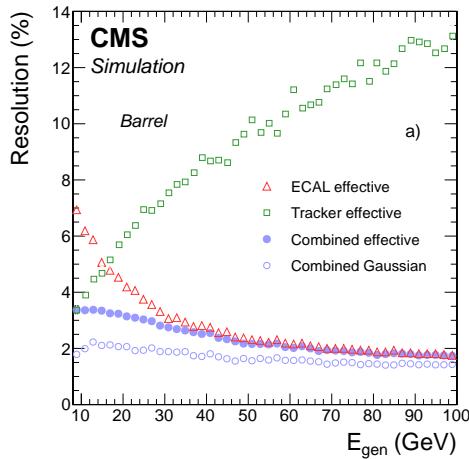


Figure 6.6: Expected resolution in p_T for isolated electrons in the ECAL barrel as a function of the electron generated energy, obtained from the ECAL, the tracker and the combined estimates [51].

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 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 γ) 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/ γ 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) $H/E < 0.05$ $\sigma_{in\eta} < 0.011$ (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/ γ 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) $H/E < 0.05$ $\sigma_{in\eta} < 0.011$ (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

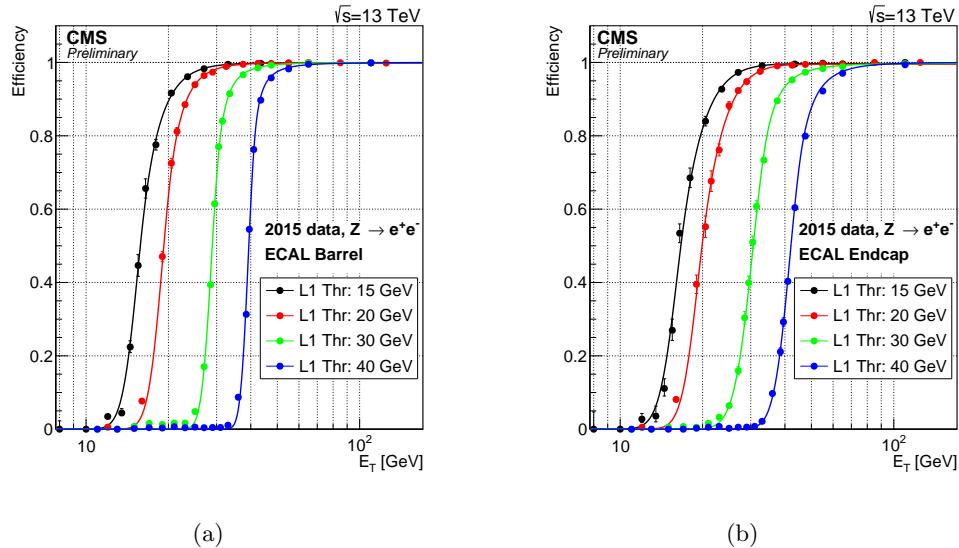


Figure 6.7: L1 electron triggering efficiency in ECAL barrel (a) and endcaps (b) as a function of the offline reconstructed electron E_T . The efficiency is shown for the 15, 20, 30, 40 GeV EG trigger thresholds [52].

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

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 weak 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 [53]. 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 η 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 η_{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	≤ 1	≤ 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	≤ 1	≤ 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 η , $\Delta\eta_{in}$, and in ϕ , $\Delta\phi_{in}$, 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 ϕ 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 η of the seed cluster of the supercluster which is found to provide a more accurate indication of the η 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 η -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, ρ , 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 1×5 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 5×5 matrix centered on the seed crystal;
- $E_{2\times 5}/E_{5\times 5}$: ratio of the energy contained in the most energetic 2×5 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 5×5 matrix centered on the seed crystal;
- $\sigma_{i\eta i\eta}$: measure of the spread in η in units of crystals of the electrons energy in the 5×5 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 $\sigma_{i\eta i\eta}$.

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 electrons, so a specific muon detection system (see Sec. 3.2.3) is needed in order to identify

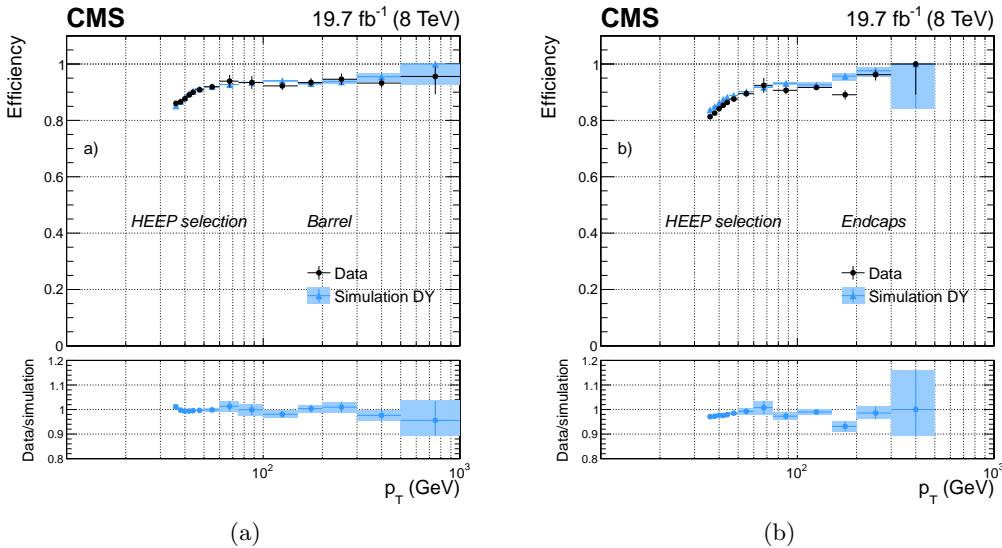


Figure 6.8: Efficiency of the high- E_T electron selection as a function of electron p_T for dielectron events in 8 TeV data (dots) and DY simulation (triangles) in the ECAL barrel (a), and endcaps (b) [51].

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 ± 0.004 (stat.)	0.984 ± 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 ± 0.001 (stat.)	0.975 ± 0.004 (stat.)

and correctly measure its momentum. In the standard CMS reconstruction [54], 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 χ^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 losses 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 $p_T > 0.5 \text{ GeV}$ 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.

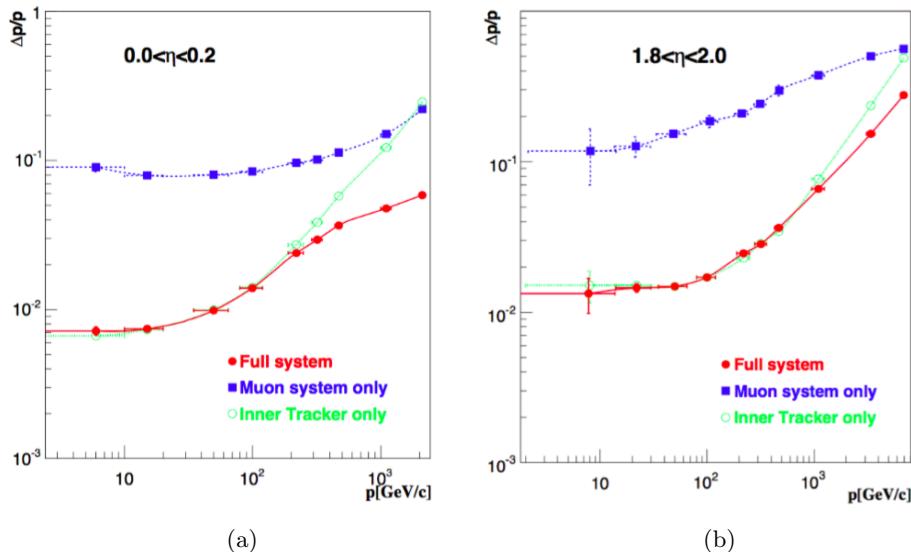


Figure 6.9: Relative resolution of the muon momentum measurement for the reconstruction with the inner tracker only, the muon system only and for the combination of the inner tracker and the muon system, for simulated muons emitted in the central (a) and forward (b) regions [55].

Figure 6.10 shows the muon tracking efficiency as a function of the η 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.

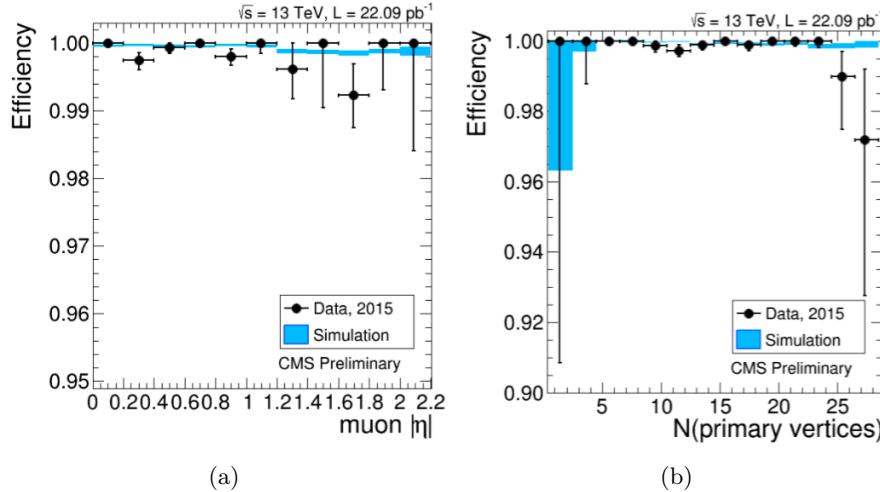


Figure 6.10: Tracking efficiency measured with a T&P technique, for muons from Z decays, as a function of the muon η (a) and the number of primary vertices (b), for 2015 data (black dots) and simulation (blue bands) [56].

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 χ^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 [54, 57].

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 η 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 η . In 2012 the efficiency of the L1 single-muon trigger was greater than 90%.

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 6.7: 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$

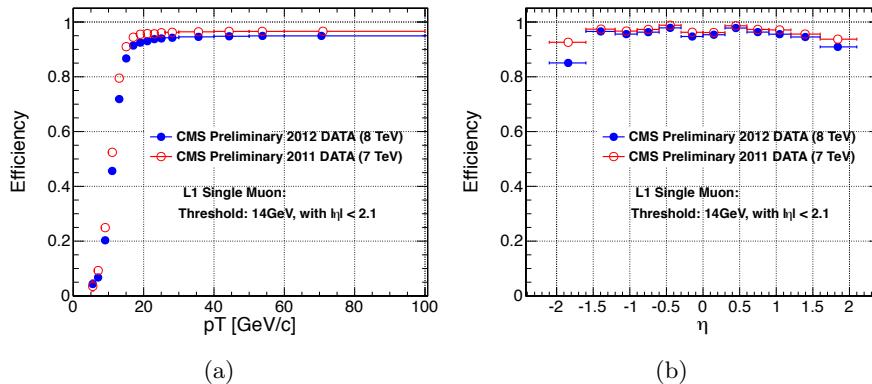


Figure 6.11: Efficiency of the L1 single-muon trigger with a 14 GeV threshold as a function of the muon p_T (a) and η (b) [58].

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 ± 0.0004	0.956 ± 0.001	0.984 ± 0.001

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 [54]. In general, particles detected as muons are produced in pp collision from different sources which lead to different

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 ± 0.002	0.961 ± 0.006	0.968 ± 0.003

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 χ^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 ≥ 1 . To further suppress muons from decays in flight.
- Number of tracker layers involved in the measurements ≥ 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 ≥ 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 ± 0.0003	0.992 ± 0.001	0.9949 ± 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 ± 0.0001	0.9994 ± 0.0001	0.9997 ± 0.0001

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

Muon $ \eta $	$0 < \eta < 1.2$	$1.2 < \eta < 2.1$
High- p_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 ± 0.005	1.002 ± 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 ± 0.001	1.001 ± 0.001

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 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 [59] and sequential clustering [60–64] algorithms. The main algorithms used by LHC experiments are the anti- k_t algorithm [64] (AK) and the Cambridge–Aachen (CA) [60, 62] algorithms, which are found to fulfil 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:

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

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_T values, the integer n depends on the specific jet algorithm, $\Delta R^2 ij$ is the distance between i and j in η and ϕ , 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 [64]. 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.

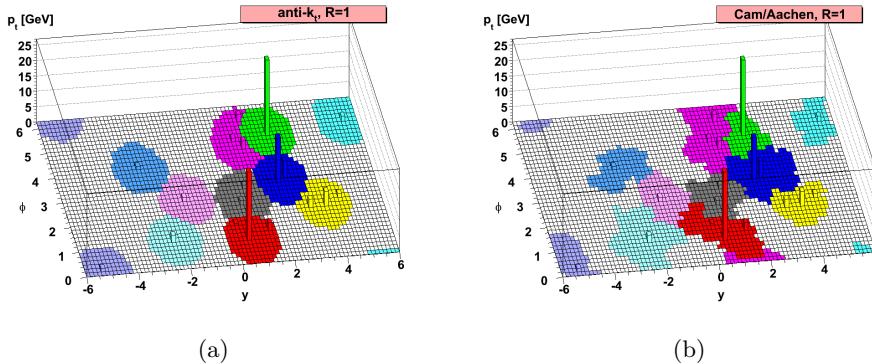


Figure 6.12: An example of jet clustering with the AK (a) and CA (b) algorithms. The reconstructed jets are shown as colored regions [64].

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 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 Section 8.1), along with the b tagging algorithm described in Section 6.4.3.

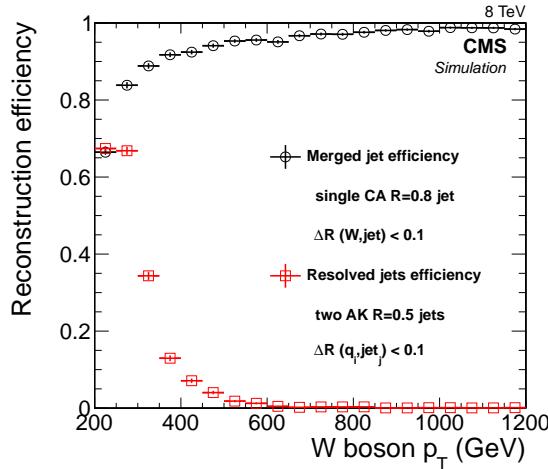


Figure 6.13: Efficiency to reconstruct a CA8 jet within $\Delta R < 0.1$ of a generated W boson, and the efficiency to reconstruct two AK5 jets within $\Delta R < 0.1$ of the generated quarks W bosons, as a function of the p_T of the W boson [65].

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 [66, 67] 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.

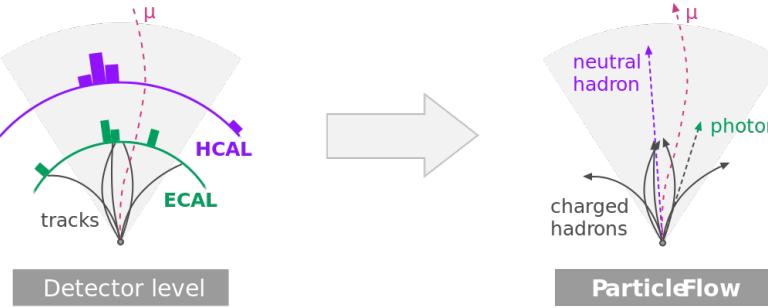


Figure 6.14: Sketch of the CMS particle-flow algorithm.

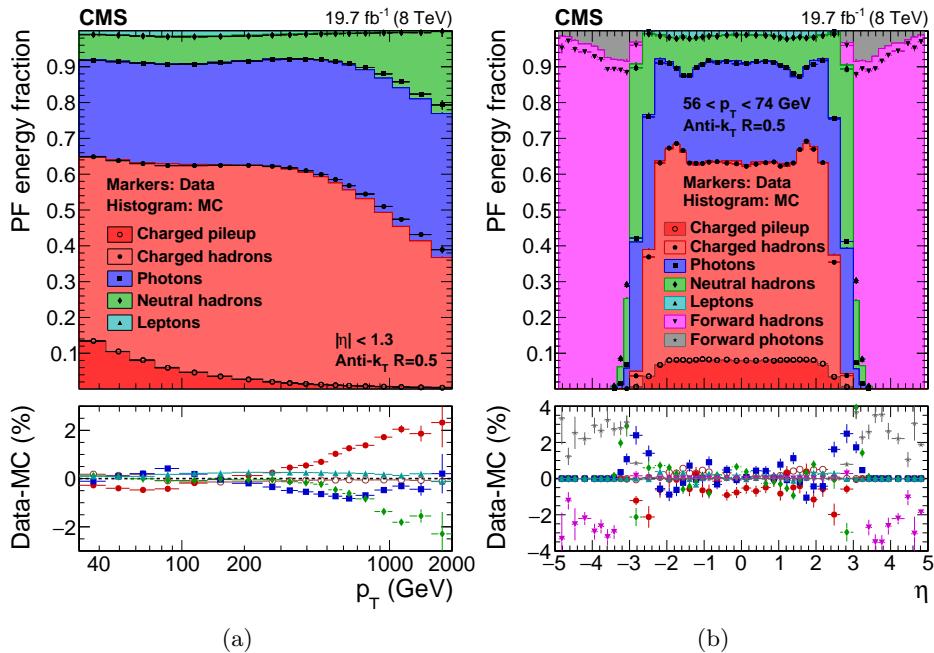


Figure 6.15: PF jet composition in data and simulation versus p_T for jets with $|\eta| < 1.3$ (a), and versus η at $56 < p_T < 74 \text{ GeV}$ (b) [68].

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 [69] 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^{miss} (see Section 6.5). This method is referred to as *charged-hadron subtraction* (CHS), and represents the reference benchmark 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 [70–72] 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 (ρ) [73, 74], the jet area (A) [75], jet p_T and η . 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 η 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.

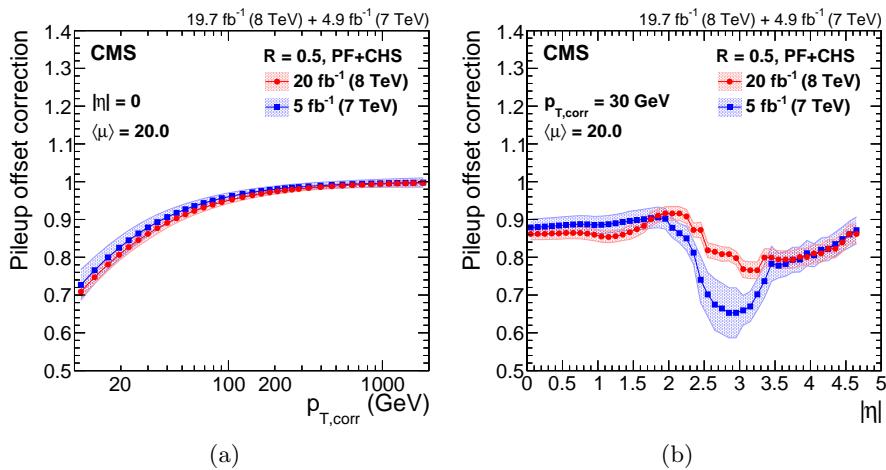


Figure 6.16: Pileup offset correction for AK5 CHS jets, with systematic uncertainty band, estimated for the typical 2012 condition of $\langle\mu\rangle = 20$. Corrections are shown for jets at $|\eta| = 0$ versus corrected jet p_T (a), and for jets with $p_T = 30 \text{ GeV}$ versus $|\eta|$ (b) [68].

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 η 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.

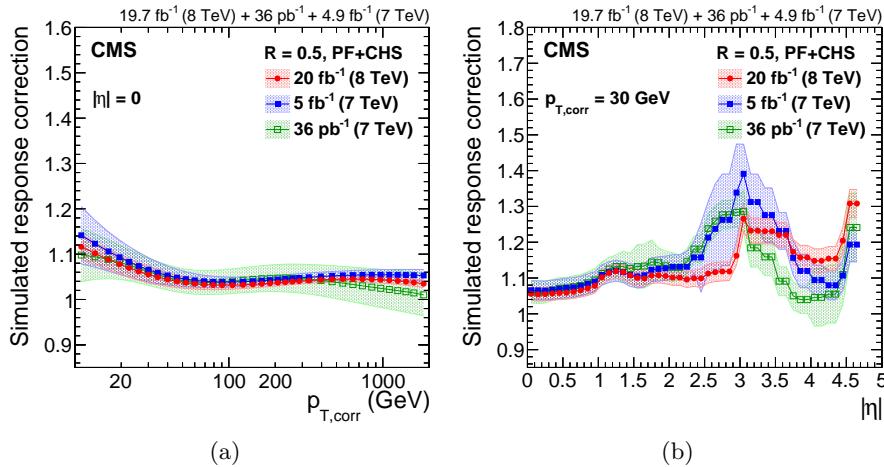


Figure 6.17: Response correction factors with their systematic uncertainty band from simulation for the 2012 data collected at 8 TeV for AK5 CHS jets. Corrections are shown for jets at $|\eta| = 0$ versus corrected jet p_T (a), and for jets with $p_T = 30$ GeV versus $|\eta|$ (b) [68].

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 η 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 γ). 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_{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_{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.

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:

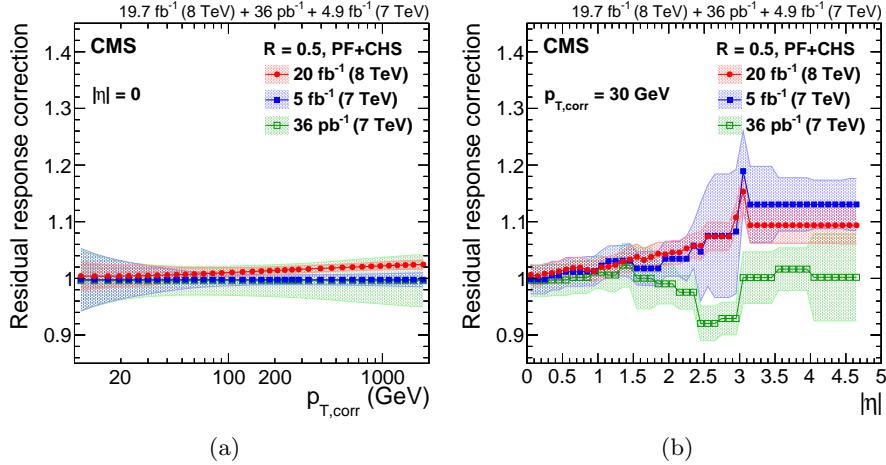


Figure 6.18: Residual data/simulation response correction factors for the 2012 data collected at 8 TeV for AK5 CHS jets. Corrections are shown for jets at $|\eta| = 0$ versus corrected jet p_T (a), and for jets with $p_T = 30$ GeV versus $|\eta|$ (b) [68].

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

where C_{pu} represents the pileup correction, C_{sim} is the simulated response correction and C_{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 η , 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 spurious jet-like features originating from isolated noise patterns in the calorimeters or the tracker. The efficiency of these requirements is above 99% for real jets.

All AK8 and AK5 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. Furthermore, the AK8 and AK5 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. This requirement is applied to clean the jet collection used in the analysis from leptons mis-identified as jets. Finally, AK4 jets are required to be separated from the AK8 jet representing the V jet candidate by $\Delta R > 0.8$ since an overlap is expected between the two reconstructions. The same selections are applied for CA8 and AK5 jets in the 8 TeV data analysis. For this case an additional selection is applied to the pseudorapidity of CA8 jets. In fact, CA8 jets are not used in

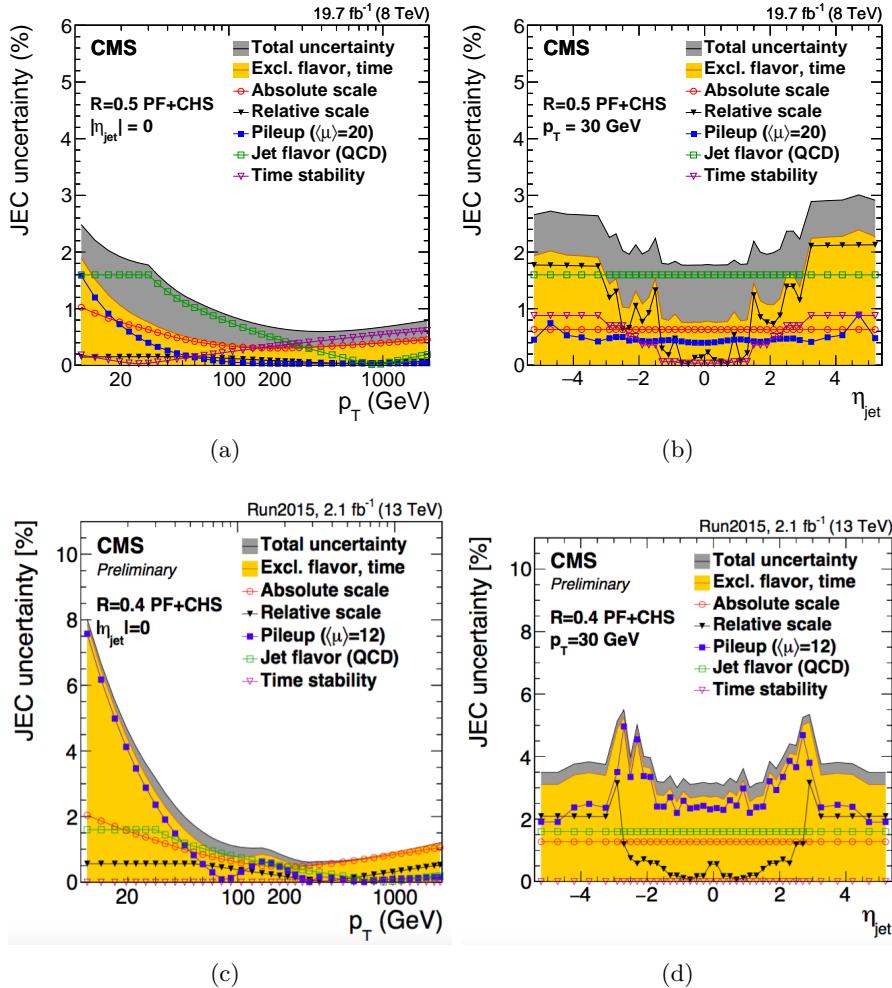


Figure 6.19: Summary of jet energy scale systematic uncertainties for the 2012 data collected at 8 TeV for AK5 CHS jets (upper plots) and for the 2015 data collected at 13 TeV for AK4 CHS jets (lower plots). Uncertainties are shown for jets at $|\eta| = 0$ versus corrected jet p_T (left), and for jets with $p_T = 30$ GeV versus $|\eta|$ (right) [68, 76].

the analysis if their pseudorapidity falls in the region $1.0 < |\eta| < 1.8$, thus overlapping the barrel-endcap transition region of the silicon tracker. In Run 1 it has been found that in this region, ‘noise’ can arise when the tracking algorithm reconstructs many fake displaced tracks associated with the jet. This issue in the reconstruction has been studied in details in the context of the analysis described in this work. The studies, presented and discussed in Appendix A, resulted in the choice of the η region to be excluded. In particular, the simulation does not sufficiently describe the full material budget of the tracking detector in that region, thus it does not accurately describe this effect. Without this requirement, a bias can be introduced in the b tagging, jet substructure and missing transverse momentum information, making this analysis systematically prone to that noise. As a consequence of these results, other analyses involving similar kinematic cuts and identification algorithms have been affected. However, this problem has been fixed for Run 2 and this additional fiducial cut does not have to be applied in the 13 TeV data analyses.

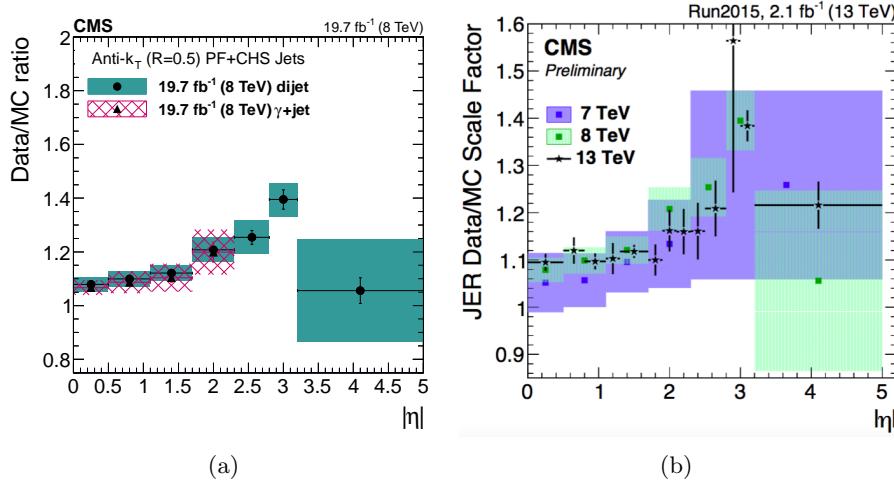


Figure 6.20: Data/MC scale factors for the jet p_T resolution as a function of $|\eta|$ determined from 2012 data collected at 8 TeV for AK5 CHS jets (a) and from 2015 data collected at 13 TeV for AK4 CHS jets (b) [68, 76].

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 ($c\tau \equiv 450\mu\text{m}$) corresponding to a flight distance that is observable with high resolution tracking detectors. A b hadron with $p_T = 50\text{ GeV}$ 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 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.

A variety of algorithms has been developed in CMS [77] to identify b jets based on the 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.

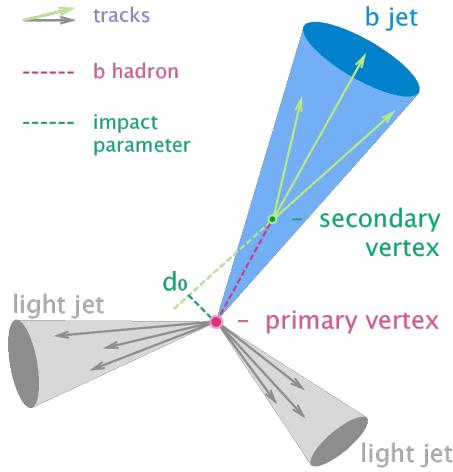


Figure 6.21: Representation of a b hadron decay and reconstructed b jet in the transverse plane.

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 $p_T > 1 \text{ GeV}$;
- 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 χ^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 $< 17 \text{ cm}$, 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 [78] 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 $< 6.5 \text{ GeV}$ and not compatible with the mass of the K_S^0 hadron in a window of 50 MeV;
- the angular distance ΔR 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) [79], 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 η 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.

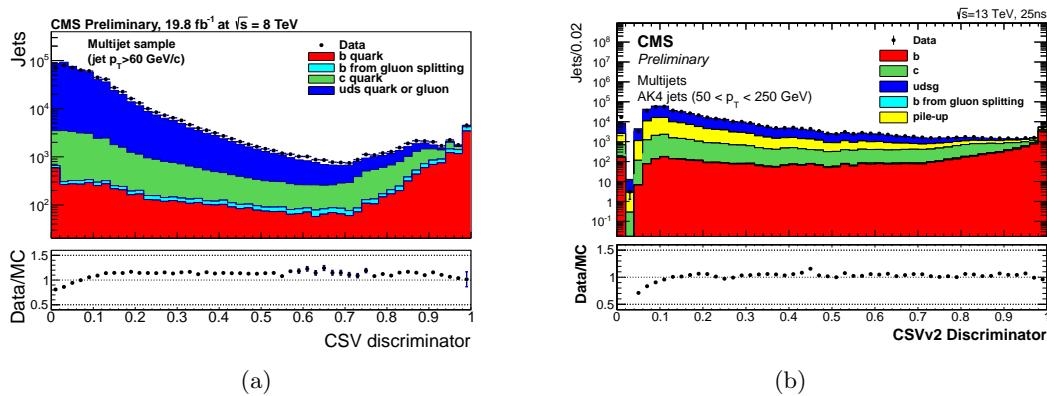


Figure 6.22: (a) Distribution of the CSV discriminator value in a multijet sample for data collected at 8 TeV and for simulation [80], for jets reconstructed with the AK5 algorithm. (b) Distribution of the CSVv2 discriminator value in a multijet sample for data collected at 13 TeV and for simulation, for jets reconstructed with the AK4 algorithm [81].

The CSV algorithm was further optimized for LHC Run 2 and the new version is referred to as CSV version 2 (CSVv2) [81]. The main differences with the Run 1 version of the CSV algorithm are the different vertex reconstruction algorithm used, the number of input 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 8 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.

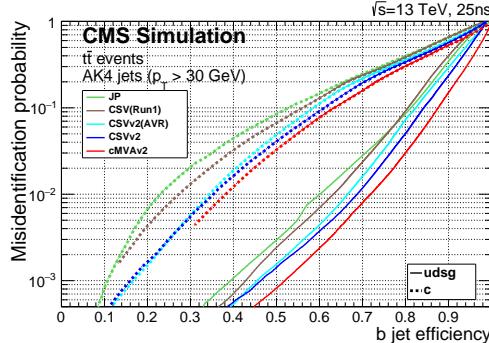


Figure 6.23: Performance of the CSVv2 algorithm demonstrating the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The improvement of this algorithm with respect to the Run 1 version of the algorithm is also shown [81].

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.

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
CSVv2 (Run 2)	CSVv2L	0.460	83
	CSVv2M	0.800	69
	CSVvsT	0.935	49

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

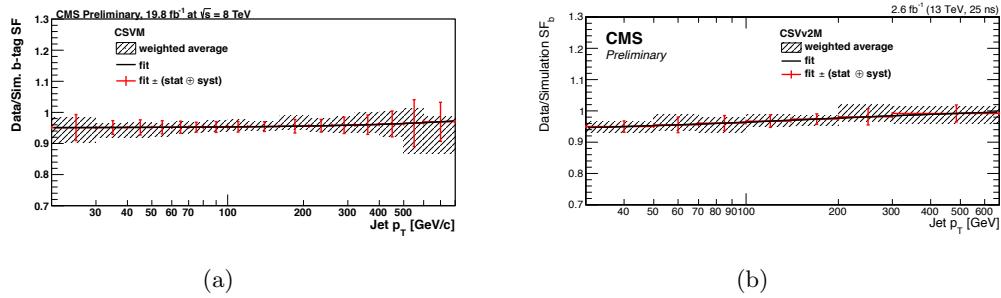


Figure 6.24: Data-to-simulation correction factors for the b tagging efficiency for the CSVM (a) and CSVv2M (b) algorithms as a function of the b jet p_T as measured in 8 and 13 TeV data, respectively [80, 81].

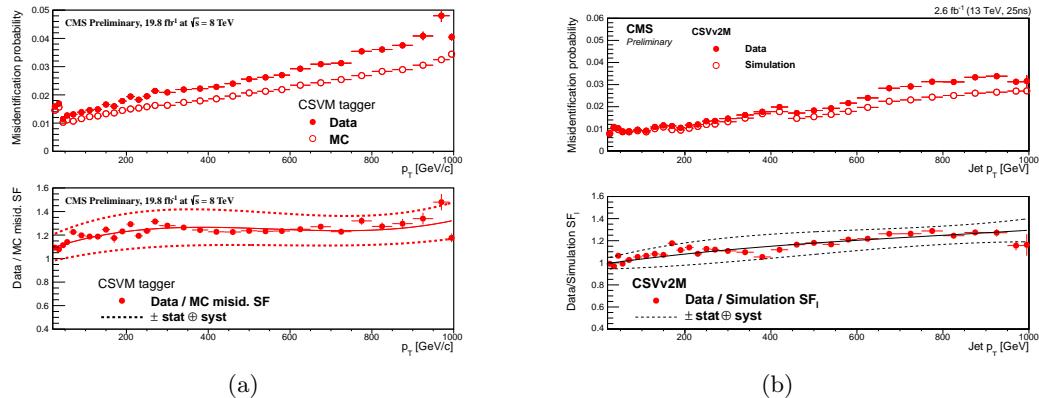


Figure 6.25: Data-to-simulation correction factors for the misidentification probability for the CSVM (a) and CSVv2M (b) algorithms as a function of the jet p_T as measured in 8 and 13 TeV data, respectively [80, 81].

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 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^{miss} .

In CMS several standard methods are available for the reconstruction of the \vec{E}_T^{miss} as for the jet reconstruction: using only calorimeter information, including also tracker information, or the PF algorithm [82]. In this analysis, the PF \vec{E}_T^{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^{miss} . The E_T^{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^{miss} plays a crucial role for measurements of standard model physics involving W bosons and top quarks.

The \vec{E}_T^{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^{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^{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^{miss} reconstruction in events with large numbers of pileup interactions. This is achieved as explained in the following.

The raw \vec{E}_T^{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 (PUch) particles so that the equation above can be expressed as

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

The contribution to the genuine \vec{E}_T^{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^{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

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

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 $f(\vec{v}) = c_1(1.0 + \text{erf}(-c_2|\vec{v}^{c_3}|))$, where the coefficients c_1 , c_2 , and c_3 are extracted from simulated minimum bias events. The corrected \vec{E}_T^{miss} is then obtained removing the additional contribution $\vec{\Delta}_{\text{PU}}$ from Eq. 6.6

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

Another type of correction is derived and applied to correct for a modulation in ϕ in the \vec{E}_T^{miss} present not only in data but also in simulation. The distribution of genuine \vec{E}_T^{miss} is instead independent of ϕ 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^{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

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

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 [83, 84] but will not be discussed in this Section since they are not used in this work.

The distributions of the PF E_T^{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^{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^{miss} measurement. The q_T can therefore be used as a reference to measure the scale and resolution of \vec{E}_T^{miss} . The hadronic recoil can be projected to the axis defined by q_T , yielding two signed components, parallel ($u_{||}$) 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 $-\langle u_{||} \rangle / \vec{q}_T$ is referred to as the \vec{E}_T^{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, $\sigma(u_{||})$ 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.

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^{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 [85]:

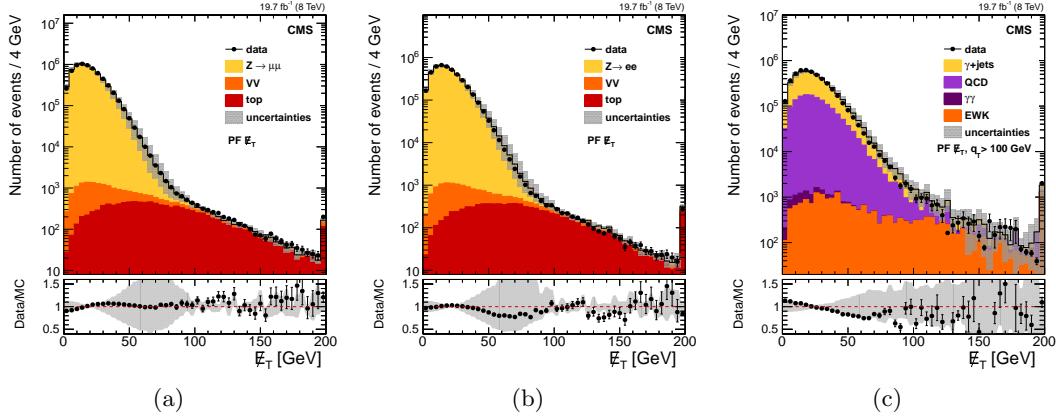


Figure 6.26: The PF E_T^{miss} distribution in $Z \rightarrow \mu^+\mu^-$ (a), $Z \rightarrow e^+e^-$ (b), and prompt photon (c) events for data collected at 8 TeV. The points in the lower panel of each plot show the data/MC ratio [83].

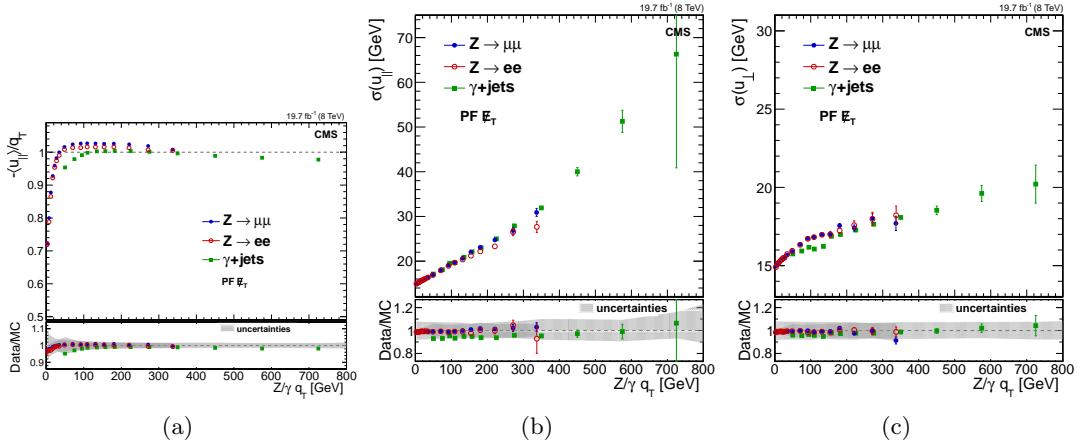


Figure 6.27: (a) Response curves for PF \vec{E}_T^{miss} in events with a Z-boson or prompt photon. Also shown are the resolution curves of the parallel (b) and perpendicular (b) recoil components as a function of the versus $Z/\gamma q_T$ for PF \vec{E}_T^{miss} in events with a Z-boson or prompt photon. 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. [83].

$$M_W^2 = m_\ell^2 + 2(E_\ell E_\nu - p_{x_\ell} p_{x_\nu} - p_{y_\ell} p_{y_\nu} - p_{z_\ell} p_{z_\nu}) = (80.4)^2 \quad (6.11)$$

In the case of two real solutions, the one with the smaller absolute value is chosen

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^{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 [86]. 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 invariant mass.

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 τ 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.

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 [65, 87]. 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 [65, 87]. In this analysis, the *jet pruning* algorithm [88, 89] is used, as it provides the best discrimination against QCD background as discussed in Ref. [65, 87].

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:

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

where m_j and p_T are the mass and transverse momentum of the originally-clustered jet, and z_{cut} and α 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_{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_{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_{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_{jet} for W, Z and H jet candidates from the decay of highly boosted W, Z and Higgs bosons, respectively. The distribution in m_{jet} for the W+jets background is also shown. Not-full-merged signal jets give rise to a peak at low masses.

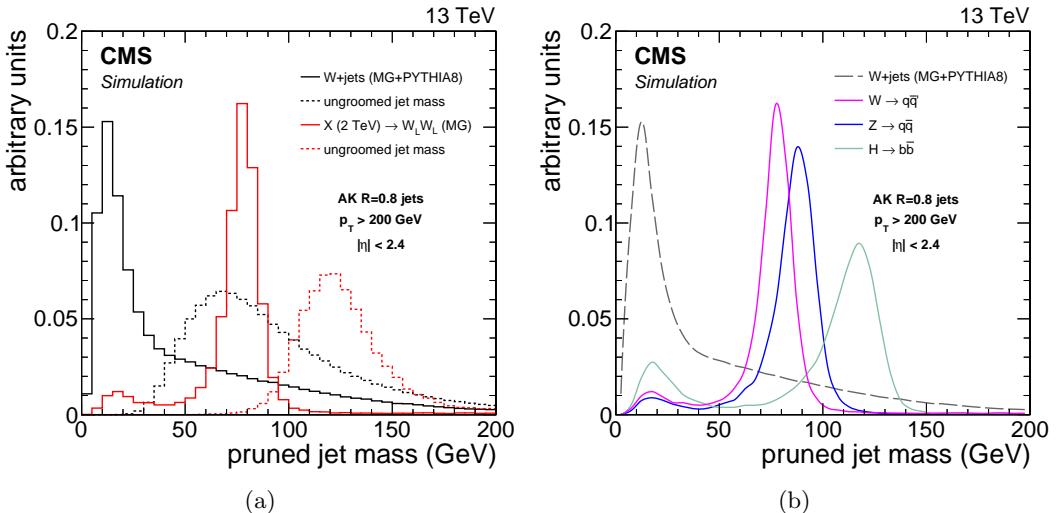


Figure 7.1: (a) Distribution in pruned jet mass m_{jet} for simulated events of highly boosted W bosons and inclusive QCD jets expected in the W+jet topology. The ungroomed jet mass is shown as dotted lines to illustrate the effect of pruning. MG denotes the MADGRAPH generator. (b) Comparison of the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

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* [90]. 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, τ_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 τ_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 [61], and reversing the last N clustering steps. The variable τ_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 τ_{21} relative to QCD background. Figure 7.2 shows the N -subjettiness ratio τ_{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.

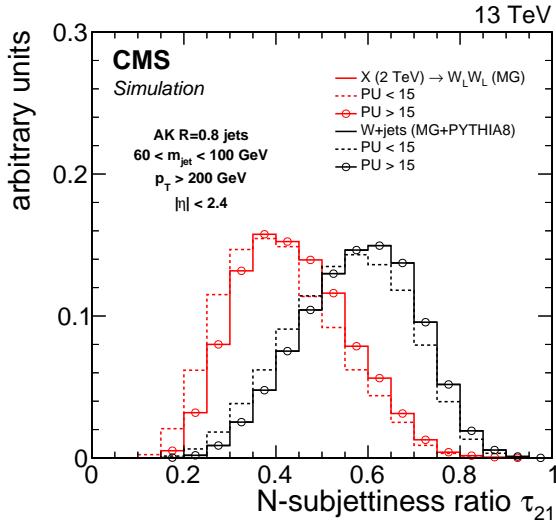


Figure 7.2: Distribution in N-subjettiness ratio τ_{21} for simulated events of highly boosted and longitudinally polarized W bosons and inclusive QCD jets expected in the W+jet topology. The distributions are shown after a selection on the pruned jet mass requiring $60 < m_{\text{jet}} < 100$ GeV. MG 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_{jet} and N-subjettiness ratio τ_{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 [91–93]. 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 τ_{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 τ_{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. [65]. From simulation studies it is observed that the efficiency of the m_{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 [72, 94].

The efficiency of the additional τ_{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 τ_{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 [72]. It is observed that the efficiency of the m_{jet} selection is constant as a function of PV, whereas the additional τ_{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 τ_{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 τ_{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 ΔR 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_{jet} and τ_{21} in simulation. The data/simulation discrepancies in m_{jet} and τ_{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 [72]. 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 τ_{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 τ_{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 obtained with 13 TeV data 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 ± 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++ [95] 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 Tables 7.3 and 7.2, as measured with 8 and 13 TeV data, respectively. The quoted uncertainties are statistical. The W jet mass scale in 13 TeV data is $\sim 1\%$ smaller than in simulation while its resolution is found to be larger by about 5%. In 8 TeV data both the W jet mass scale and resolution in data are larger than that in simulation. In the simulation m_{jet} must therefore be shifted by $1.7 \pm 0.6\%$ and σ be enlarged by $11 \pm 9\%$ to correct for the difference between data and simulation.

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, un-

derlying event, and initial-state radiation not completely removed in the jet pruning procedure.

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.

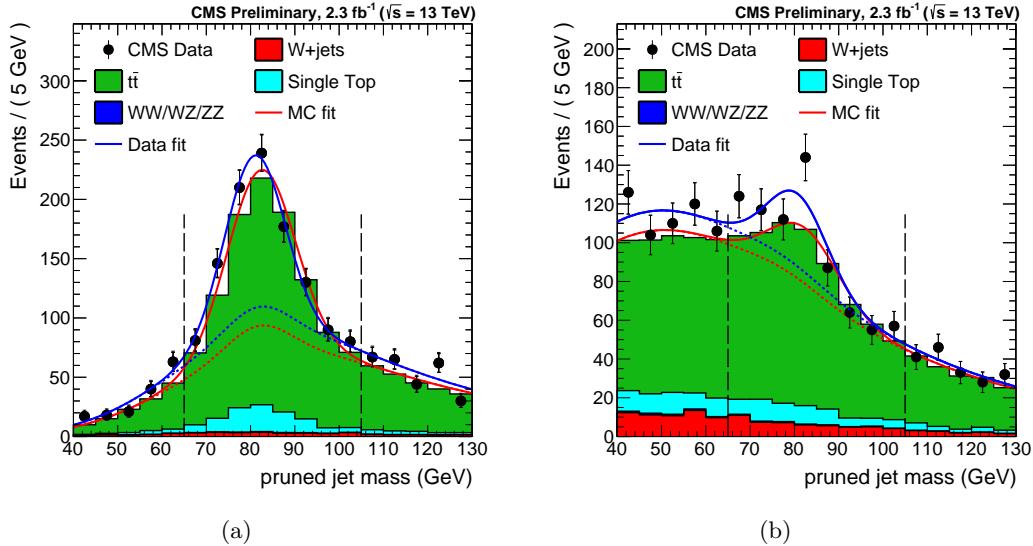


Figure 7.3: Distribution in pruned jet mass for events that (a) pass and (b) fail the $\tau_{21} < 0.45$ selection in the $t\bar{t}$ control sample. The result of the fit to data and simulation are shown, respectively, by the solid and long-dashed line and the background components of the fit are shown as dashed-dotted and short-dashed line [72].

Table 7.2: W jet mass peak position and resolution, as extracted from top quark enriched sample in 8 TeV data and from simulation, after applying the $\tau_{21} < 0.5$ selection [65].

$\tau_{21} < 0.45$	$m_{\text{jet}} [\text{GeV}]$	Standard deviation [GeV]
Data	84.1 ± 0.4	8.4 ± 0.6
Simulation	82.7 ± 0.3	7.6 ± 0.4

Table 7.3: W jet mass peak position and resolution, as extracted from top quark enriched sample in 13 TeV data and from simulation, after applying the $\tau_{21} < 0.45$ selection [72].

$\tau_{21} < 0.5$	$m_{\text{jet}} [\text{GeV}]$	Standard deviation [GeV]
Data	84.6 ± 0.7	8.2 ± 0.7
Simulation	85.1 ± 0.2	7.8 ± 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 [80]:

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

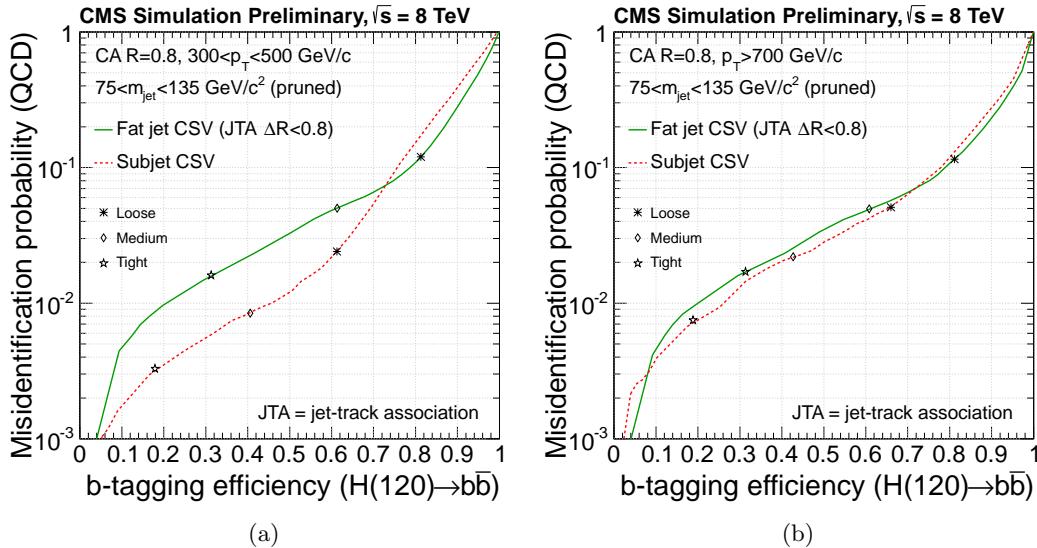


Figure 7.4: Misidentification probability as a function of b tagging efficiency for boosted $H \rightarrow b\bar{b}$ jets and inclusive QCD jets for the CSV algorithm applied to large-cone jets and pruned subjets for large-cone jets with (a) $300 < p_T < 500$ GeV and (b) $p_T > 700$ GeV. Loose, medium, and tight operating points of the CSV discriminator are indicated [80].

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 ΔR 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 ΔR 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 $\ell\nu + bb$ analysis presented in this work a requirement on the pruned jet mass of the reconstructed H jet candidate given by $110 < m_{jet} < 135$ GeV is used. The m_{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%.

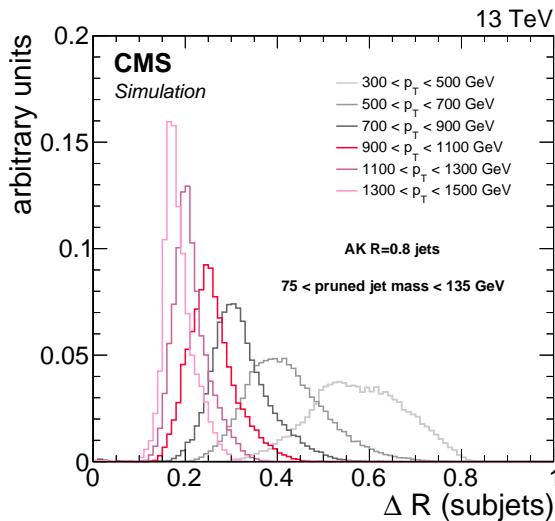


Figure 7.5: Distributions of the angular separation ΔR of the two subjets reconstructed within the fat jet for simulated events of highly boosted Higgs bosons decaying to $b\bar{b}$. The distributions are compared for different ranges of the H jet p_T .

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 [80]. 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.

Analysis strategy

8.1 Final event selection and categorization

Events are selected online with triggers requiring either one muon or electron as described in Sections 6.2.2 and 6.3.2. Several requirements are then applied offline to the selected events to enhance the analysis sensitivity as described in the following.

The two analyses described in this work feature the same selection strategy on the leptonic part of the final state. Both analyses require exactly one muon or one electron satisfying certain p_T and η requirements and passing the high- p_T lepton identification criteria described in Sections 6.3.3 and 6.2.3. As summarized in Tables 8.1 and 8.2, the only difference is in the p_T threshold of the lepton which is higher for the analysis at 13 TeV to match the increase in the trigger threshold.

The offline reconstructed p_T of the electron 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. Reconstructed electrons must have $|\eta| < 2.5$ and also be located outside of the overlap region between the ECAL barrel and endcaps, because the reconstruction of an electron object in this region is not optimal.

In a similar way, the offline reconstructed p_T of the muon must be greater than 50 (53) GeV for the 8 (13) TeV analysis, and within $|\eta| < 2.1$ as a consequence of the trigger criteria.

Events with additional well-identified muons and/or electrons are rejected to avoid contamination from events containing $Z \rightarrow \ell\ell$ decays.

The requirements $E_T^{\text{miss}} > 40$ and > 80 GeV are applied, respectively, in the muon and electron channels. The threshold is higher in the electron channel to further suppress the larger background from multijet processes expected in the low- E_T^{miss} range due to jets misidentified as electrons. The background is expected to be negligible in the muon channel, for which a lower E_T^{miss} threshold can be used to preserve a higher efficiency for a low-mass signal. The identified lepton and the E_T^{miss} are used to reconstruct the $W \rightarrow \ell\nu$ candidate as described in Section 6.6, which is required to have p_T above 200 GeV.

A different strategy is instead used in the two analyses, for the hadronic part of the final state. As described in Section 6.4, the CA8 and AK8 algorithms are used to reconstruct the H and V jet candidates in the 8 and 13 TeV analyses, respectively. In both cases the jet is required to have p_T above 200 GeV and $|\eta| < 2.4$. For CA8 jets, the pseudorapidity region $1.0 < |\eta| < 1.8$ is excluded corresponding to the barrel-endcap transition region of the silicon tracker where the reconstruction of tracks is not optimal (Section 6.4.2). The probability of signal events with jest outside this region is 80% (92%) for a resonance mass of 1.0 (2.5) TeV.

The 8 TeV analysis aims at isolating events with a high- p_T Higgs boson decaying to $b\bar{b}$ and the H tagging algorithm described in Section 7.3 is applied. The H tagging requires the selected CA8 jet to have pruned mass in the range $110 < m_{\text{jet}} < 135$ GeV. Furthermore, the subjets are required to be b tagged with the CSVL algorithm if their angular distance $\Delta R < 0.3$. Otherwise, b tagging is applied to the whole CA8 jet using the same algorithm.

Table 8.1: Summary of the final selection in the $\ell\nu+b\bar{b}$ analysis at 8 TeV.

Selection	Value
Lepton selection	
Electron	$p_T > 90 \text{ GeV}$ $ \eta < 2.5$ except [1.44, 1.57] range
Muon	$p_T > 50 \text{ GeV}$ $ \eta < 2.1$
AK5 jet selections	
Jet p_T	$p_T > 30 \text{ GeV}$
Jet η	$ \eta < 2.4$
E_T^{miss} selections	
E_T^{miss} (electron ch.)	$E_T^{\text{miss}} > 80 \text{ GeV}$
E_T^{miss} (muon ch.)	$E_T^{\text{miss}} > 40 \text{ GeV}$
Boson selections	
$W \rightarrow \ell\nu$	$p_T > 200 \text{ GeV}$
$H \rightarrow b\bar{b}$ (CA8 jet)	$p_T > 200 \text{ GeV}$
Back-to-back topology	$ \eta < 2.4$ except [1.0, 1.8] range $\Delta R(\ell, H_{b\bar{b}}) > 1.6$ $\Delta\phi(H_{b\bar{b}}, \vec{E}_T^{\text{miss}}) > 2$ $\Delta\phi(H_{b\bar{b}}, W_{\ell\nu}) > 2$
Diboson invariant mass	$m_{WH} > 0.7 \text{ TeV}$
H tagging selections	
Pruned jet mass	$110 < m_{\text{jet}} < 135 \text{ GeV}$
Combined b-tagging cut	2 CSVL b-tagged subjets if $\Delta R(\text{subjets}) > 0.3$ 1 CSVL b-tagged CA8 jet if $\Delta R(\text{subjets})$
$t\bar{t}$ rejection	
B-tag veto	no CSVM b-tagged AK5 jet within $\Delta R(H_{b\bar{b}}, AK5) = 0.8$
Top mass veto	$m_{\text{top}}^l < 120 \parallel m_{\text{top}}^l > 240$ $m_{\text{top}}^h < 160 \parallel m_{\text{top}}^h > 280$

The 13 TeV analysis is instead focused on events with a high- p_T V boson decaying to $q\bar{q}$ and the V tagging algorithm described in Section 7.3 is applied in this case. The pruned jet mass window is shifted down to the V boson mass, requiring the selected AK8 jet to have pruned jet mass in the range $65 < m_{\text{jet}} < 105 \text{ GeV}$. Furthermore, the V jet is required to have $\tau_{21} < 0.6$. Finally, the V jet is deemed a W-boson candidate if its pruned mass falls in the range 65–85 GeV, while it is deemed a Z-boson candidate if it falls in the range 85–105 GeV instead. This categorization has been added on the top of the V tagging requirements on the m_{jet} to enhance discrimination between resonances with different charge and spin. Indeed, the first category, referred to as ‘WW-enriched’, has higher sensitivity for resonances such the neutral spin-2 graviton or the neutral spin-1 Z' decaying to WW, where a W jet is expected. The second category, referred to as ‘WZ-enriched’, is instead optimized for resonances such as the charged spin-1 W' decaying to WZ, where a Z jet is expected.

In addition, there are specific topological selection criteria chosen for both the analyses. It is required that the two V bosons from the decay of a massive resonance are approximately back-to-back: the ΔR between the lepton and the V jet is greater than $\pi/2$; the $\Delta\phi$ between the vector \vec{E}_T^{miss} and the V jet, as well as between the $W \rightarrow \ell\nu$ and V jet candidates, are both greater than 2 radians.

To reduce the level of the $t\bar{t}$ background events with one or more reconstructed AK5 (or AK4) jets, not overlapping with the V jet candidate as described previously in Section 6.4.2, are analyzed. If one or more of the AK5 jets is b-tagged with the CSVM algorithm, the

Table 8.2: Summary of the final selection in the $\ell\nu q\bar{q}$ analysis at 13 TeV.

Selection	Value
Lepton selection	
Electron	$p_T > 120 \text{ GeV}$ $ \eta < 2.5$ except [1.44, 1.57] range
Muon	$p_T > 53 \text{ GeV}$ $ \eta < 2.1$
AK4 jet selections	
Jet p_T	$p_T > 30 \text{ GeV}$
Jet η	$ \eta < 2.4$
E_T^{miss} selections	
E_T^{miss} (electron ch.)	$E_T^{\text{miss}} > 80 \text{ GeV}$
E_T^{miss} (muon ch.)	$E_T^{\text{miss}} > 40 \text{ GeV}$
Boson selections	
$W \rightarrow \ell\nu$	$p_T > 200 \text{ GeV}$
$V \rightarrow q\bar{q}^{(\prime)}$ (AK8 jet)	$p_T > 200 \text{ GeV}$ $ \eta < 2.4$
Back-to-back topology	$\Delta R(\ell, V_{q\bar{q}^{(\prime)}}) > \pi/2$ $\Delta\phi(V_{q\bar{q}^{(\prime)}}, p_T) > 2$ $\Delta\phi(V_{q\bar{q}^{(\prime)}}, W_{\ell\nu}) > 2$ $m_{WV} > 0.7 \text{ TeV}$
Diboson invariant mass	
V tagging selections	
Pruned jet mass	$65 < m_{\text{jet}} < 105 \text{ GeV}$
2- to 1-subjettiness ratio	$\tau_{21} < 0.6$
m_{jet} categories	
WW-enriched	$65 < m_{\text{jet}} < 85 \text{ GeV}$
WZ-enriched	$85 < m_{\text{jet}} < 105 \text{ GeV}$
$t\bar{t}$ rejection	
B-tag veto	no CSV b-tagged AK5 jet within $\Delta R(V_{q\bar{q}^{(\prime)}}, \text{AK5}) = 0.8$

event is rejected. For the 8 TeV analysis an additional algorithm is applied to further reduce contamination from $t\bar{t}$ background, as follows. A leptonically decaying top quark candidate mass m_{top}^l is reconstructed from the lepton, E_T^{miss} , and the closest AK5 jet to the lepton using the method described in Section 6.6. A hadronically decaying top quark candidate mass m_{top}^h is also reconstructed, from the H jet candidate and the closest AK5 jet. Events with $120 < m_{\text{top}}^l < 240 \text{ GeV}$ or $160 < m_{\text{top}}^h < 280 \text{ GeV}$ are rejected. The chosen windows around the top quark mass are the result of an optimization carried out in this analysis, taking into account the asymmetric tails at larger values due to combinatorial background.

According to the above description of the final selections, the event categorization is based on 2 orthogonal classes of events for the 8 TeV analysis, depending on the lepton flavour (muon or electron), and on 4 orthogonal classes of events for the 13 TeV analysis, depending on the lepton flavour and on the pruned jet mass category (WW or WZ).

The two boson candidates are combined into a diboson candidate, with presence of signal then inferred from the observation of localized excesses in the $m_{\ell\nu j}$ distribution. When several diboson resonance candidates are present in the same event, only the one with the highest p_T V or H jet is kept for further analysis.

The reconstructed invariant mass of the resonance is required to be at least 0.7 TeV.

The distributions in p_T and N-subjettiness ratio τ_{21} distributions for the V jet candidate in the $\ell\nu q\bar{q}$ channel is shown in Fig. 8.1, after requiring $65 < m_{\text{jet}} < 105 \text{ GeV}$, for both simulation and 13 TeV data. Figure 8.2 shows the distribution in p_T for the H jet candidate

after requiring $40 < m_{\text{jet}} < 110 \text{ GeV}$, for both simulation and 8 TeV data.

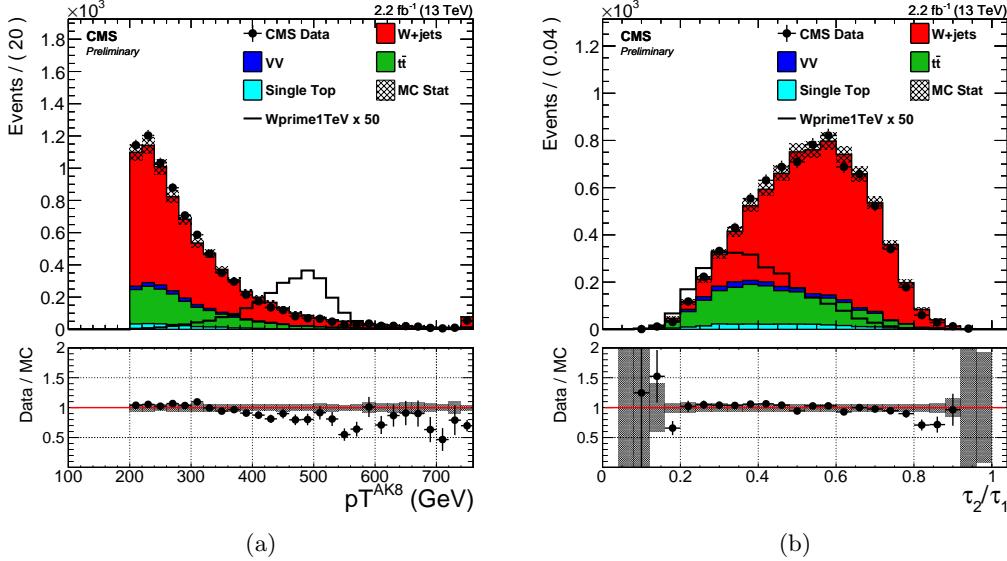


Figure 8.1: Distributions in p_T and N-subjettiness ratio τ_{21} for the V jet candidate obtained requiring $65 < m_{\text{jet}} < 105 \text{ GeV}$ after merging muon and electron channels. The SM diboson, $t\bar{t}$, and single top quark backgrounds are taken from simulation and are normalized to the integrated luminosity of the 13 TeV data sample. The $W+\text{jets}$ background is rescaled to match the number of events in data

FIXME: Redo these plots!

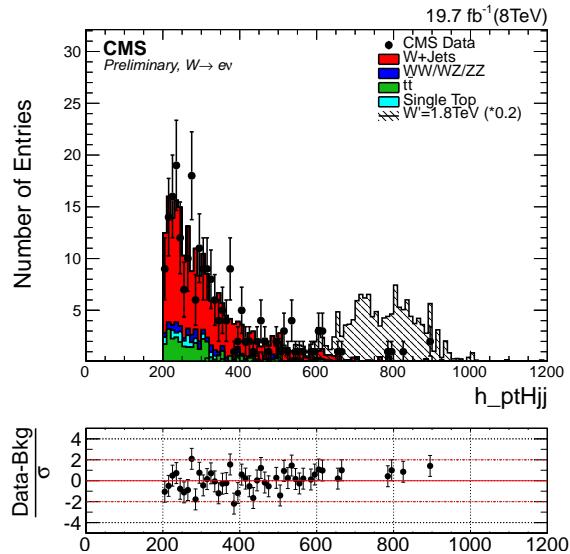


Figure 8.2: Distributions in p_T for the H jet candidate obtained requiring $40 < m_{\text{jet}} < 110 \text{ GeV}$ after merging muon and electron channels. The SM diboson, $t\bar{t}$, and single top quark backgrounds are taken from simulation and are normalized to the integrated luminosity of the 8 TeV data sample. The $W+\text{jets}$ background is rescaled to match the number of events in data

FIXME: Redo these plots!

8.2 Modelling of top quark production

The backgrounds from $t\bar{t}$ and single top quark production in both analysis channels are estimated from data-based correction factors in the normalization of the simulation. A top quark enriched control sample is selected by applying all the analysis requirements except that the b jet veto is inverted by requiring, instead, at least one b-tagged AK4 (or AK5) jet in the event. For the $\ell\nu+V$ -jet channel, the comparison between data and simulation yields normalization correction factors for $t\bar{t}$ and single top quark background processes evaluated in the pruned jet mass signal region $65 < m_{\text{jet}} < 105$ GeV. The measured correction factors are 0.87 ± 0.04 and 0.83 ± 0.07 for the muon and electron channel, respectively, where the quoted uncertainty is only statistical. The disagreement is consistent with the difference between NLO and NNLO shape prediction for large top quark p_T [96].

For the $\ell\nu+H$ -jet channel, a unique correction factor is calculated with a simultaneous fit to number of data events in the muon and electron channels in the pruned jet mass region $40 < m_{\text{jet}} < 150$ GeV. The difference in normalization between data and simulation is found to be $4.6 \pm 5.6\%$, where the quoted uncertainty is only statistical.

These scale factors include both the W boson signal and the combinatorial components mainly due to events where the extra b jet from the top quark decay is in the proximity of the W, and are used to correct the normalization of the $t\bar{t}$ and single top quark simulated background predictions in the signal regions. The relative uncertainties are used to quantify the uncertainty in the $t\bar{t}$ and single top quark background normalization. The m_{jet} distribution in the top quark enriched sample obtained with the $\mu\nu+V$ -jet selections is shown in Fig. 8.3(a), while Fig. 8.3(b) shows the τ_{21} distribution. The pruned jet mass distribution in the top quark enriched sample is also shown for the $\ell\nu+H$ -jet analysis channel in Fig. 8.4. In all cases, the m_{jet} spectrum shows a clear peak for events with a W boson decaying to hadrons, including the combinatorial background, while a reasonable agreement between the shapes in data and simulation is observed. The distributions of the reconstructed m_{top}^l and m_{top}^h for the $\mu\nu+H$ -jet channel are also shown in Fig. 8.5, where the peak at the top quark mass is clearly visible.

8.3 W+jets background estimate with α ratio method

The $m_{\ell\nu j}$ distribution observed in data is dominated by SM background processes where single quark or gluon jets are falsely identified as V jets. The dominant process is inclusive W boson production. Since both normalization and shape discrepancies are visible between data and simulation [87], a data driven method has been developed to estimate this background component, as described in the following. Sub-dominant backgrounds include $t\bar{t}$, single top quark, and non resonant diboson SM production, which are estimated from MC, after applying correction factors for residual data-to-simulation disagreement measured in control samples selected in data.

8.3.1 Description

The W+jets background is estimated through the so called α ratio method. This method assumes that the correlation between m_{jet} and $m_{\ell\nu j}$ for the dominant W+jets background can be adequately modelled by simulation. A signal-depleted control region (sideband) is defined by requiring the mass of the V or H jet to lie below or above the nominal selection; the $m_{\ell\nu j}$ distribution observed in this region is then extrapolated to the nominal region through a transfer function estimated from simulation. Other minor sources of background, such as

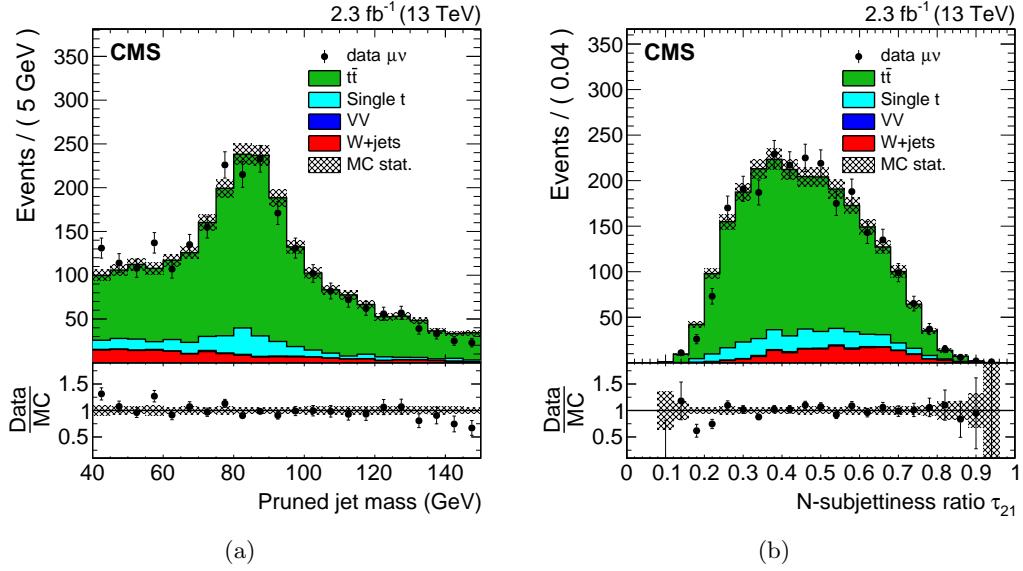


Figure 8.3: Distributions in the N-subjettiness ratio τ_{21} (a) and pruned jet mass m_{jet} (b) from the top quark enriched control sample in the $\mu\nu + V\text{-jet}$ channel. The $t\bar{t}$ background is rescaled such that the total number of background events matches the number of events in data. **FIXME: Redo these plots!**

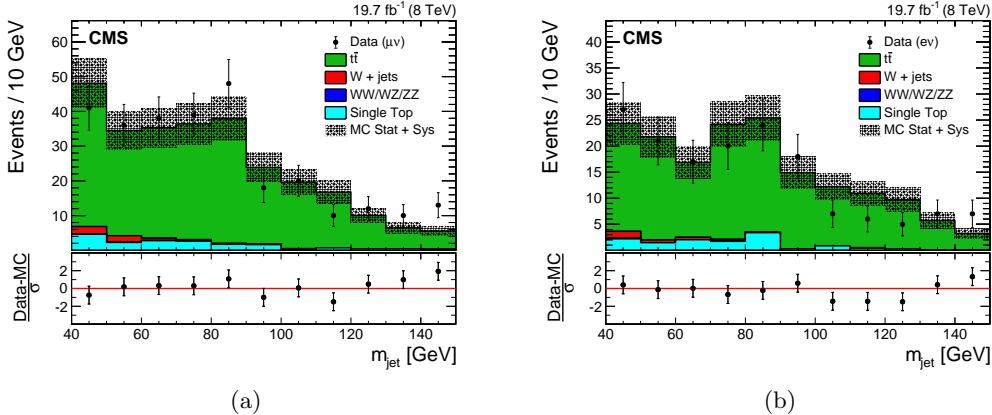
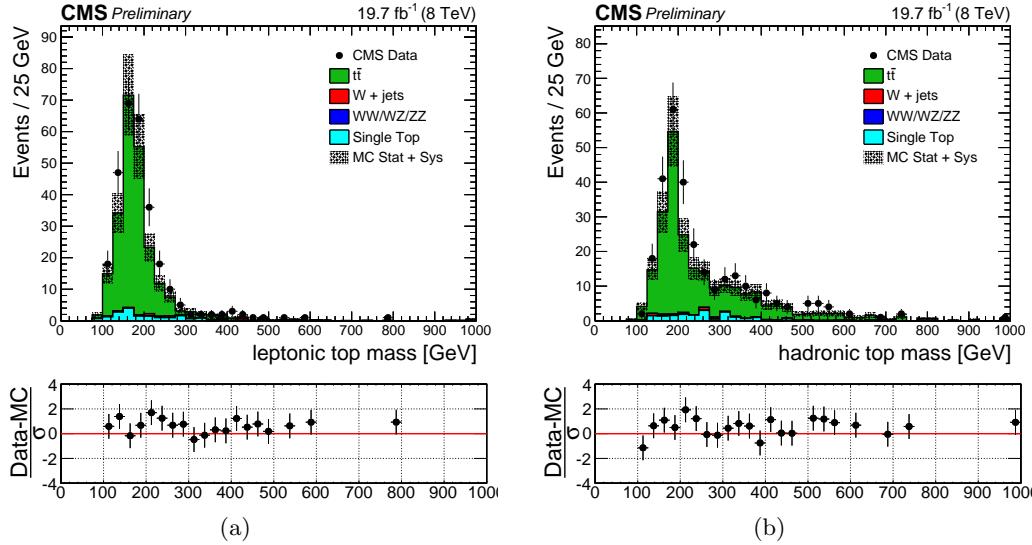


Figure 8.4: Distributions in pruned jet mass m_{jet} in the top quark enriched control sample in the electron (a) and muon (b) channels of the $\ell\nu + H\text{-jet}$ analysis. The hatched region indicates the overall uncertainty in the background. In the lower panels, the bin-by-bin residuals, $(\text{Data} - \text{MC})/\sigma$ are shown, where σ is the sum in quadrature of the statistical uncertainty of the data, the simulation, and the systematic uncertainty in the $t\bar{t}$ background.

$t\bar{t}$, single top quark, and SM diboson production, are estimated using simulated events after applying correction factors based on control regions in data, as described in Sections 7.2 and 8.2. The sideband region is defined around the jet mass window described in Section 8.1. The lower and upper sidebands for the two analyses are summarized in Table 8.3. For the 13 TeV analysis a “gap” is introduced between the signal region and the upper sideband, since the range defined by $105 < m_{jet} < 135$ might include contribution from signals with highly Lorentz-boosted Higgs bosons in the final state. Since these types of searches at 13 TeV [97] have been performed simultaneously with the one described in this work, this region has been discarded to avoid introducing a bias in the shape and normalization extrapolation due

Figure 8.5: **FIXME:** Redo these plots!

to a possible signal. On the other hand, the lower sideband of the 8 TeV $\ell\nu+H$ -jet analysis includes the region where signals from highly Lorentz-boosted V bosons might occur. In fact, this analysis has been performed after the search for $X \rightarrow WV$ resonances at 8 TeV discovered the signal region, where no deviation from the predicted SM background have been observed [91].

Table 8.3: Sidebands

m_{jet} sideband	final state	
	$\ell\nu+H$ -jet (8 TeV)	$\ell\nu+V$ -jet (13 TeV)
Low sideband (LSB)	40–110 GeV	40–65 GeV
High sideband (HSB)	135–150 GeV	135–150 GeV

8.3.2 Extraction of the W+jets normalization

The overall normalization of the W+jets background in the signal region is determined from a fit to the m_{jet} distribution in the lower and upper sidebands of the data. The analytical form of the fitting function is chosen from simulation studies, as are the contributions from minor backgrounds. A summary of the empirical functional forms used to parametrize each background contribution are listed in Table 8.4, and defined as follows:

$$\begin{aligned}
 F_{\text{Exp}}(x) &= e^{cx} \\
 F_{\text{ErfExp}}(x) &= e^{cx} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2} \\
 F_{\text{ExpGaus}}(x) &= c_0 \cdot e^{cx} + c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) \\
 F_{\text{4Gaus}}(x) &= c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) + c_2 \cdot \text{Gaus}(x, x_2, \sigma_2) + c_3 \cdot \text{Gaus}(x, x_3, \sigma_3) + c_4 \cdot \text{Gaus}(x, x_4, \sigma_4) \\
 F_{\text{ErfExp2Gaus}}(x) &= e^{cx} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2} + c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) + c_2 \cdot \text{Gaus}(x, x_2, \sigma_2)
 \end{aligned} \tag{8.1}$$

Figure 8.6 shows the functional forms listed in Table 8.4 for the $\ell\nu+V$ -jet channel, after fitting the simulation data of each background component, demonstrating that the chosen functions well reproduce the expected m_{jet} spectra.

Table 8.4: Summary of the functional forms used for fit the m_{jet} spectra of each background component in each final state.

Final state	W+jets	$t\bar{t}$	single top quark	diboson
$\ell\nu + \text{H-jet}$	$F_{\text{ErfExp}}(x)$	$F_{\text{Exp}}(x)$	$F_{\text{Exp}}(x)$	$F_{\text{ExpGaus}}(x)$
$\ell\nu + \text{V-jet}$	$F_{\text{ErfExp}}(x)$	$F_{\text{ErfExp2Gaus}}(x)$	$F_{\text{ExpGaus}}(x)$	$F_{\text{4Gaus}}(x)$

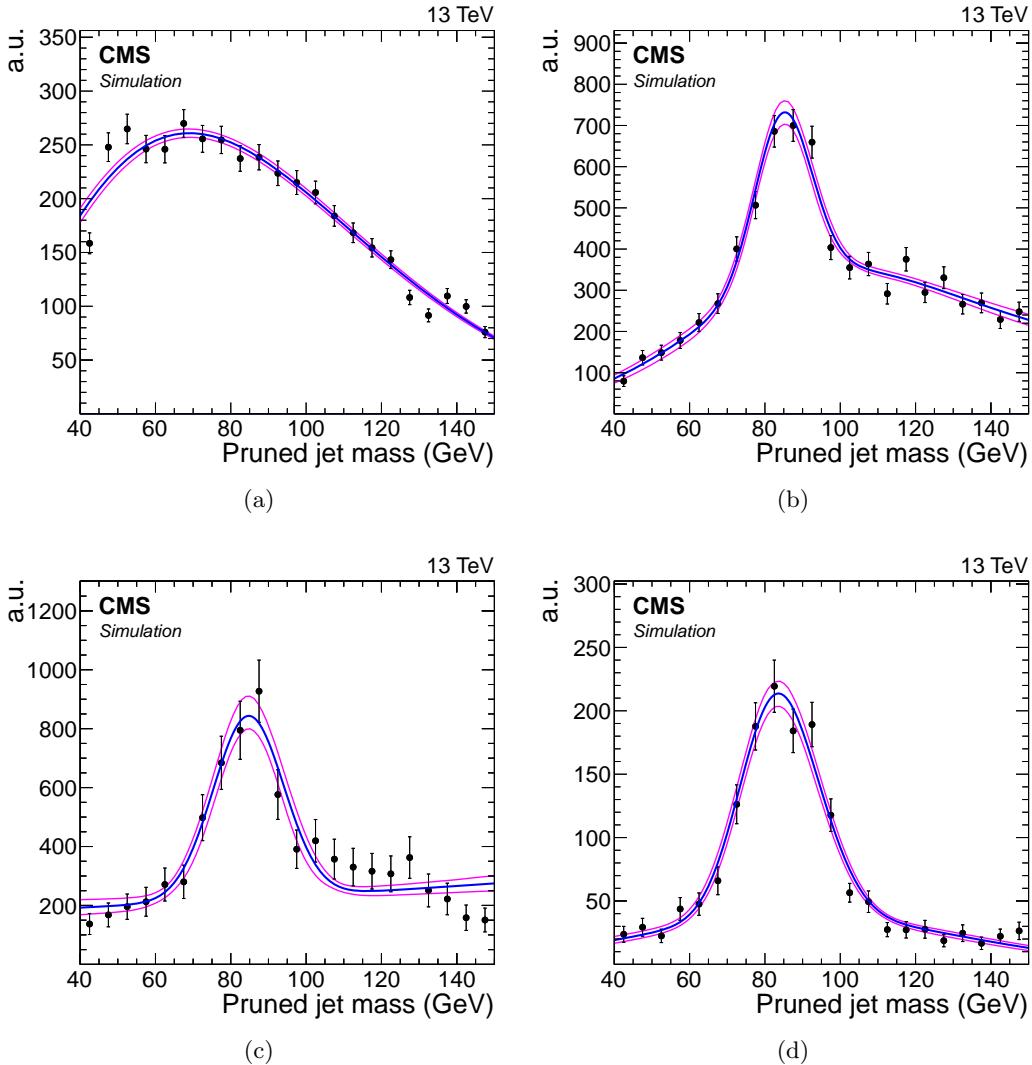


Figure 8.6: Functional forms describing the m_{jet} spectra for each background contribution after fitting the simulation data: (a) $W + \text{jets}$. (b) $t\bar{t}$. (c) Single top quark. (d) Diboson.

The results of this fit procedure to extract the $W + \text{jets}$ normalization are shown in Figures 8.7 and 8.8 for the $\ell\nu + \text{H-jet}$ and the $\ell\nu + \text{V-jet}$ channels, respectively. The factors for correcting the simulated W -peak position and resolution to represent the observed data, taken from the top quark enriched control sample as described in Section 7.2, are included in the m_{jet} spectra of Fig. 8.8.

8.3.3 Extraction of the $W + \text{jets}$ shape

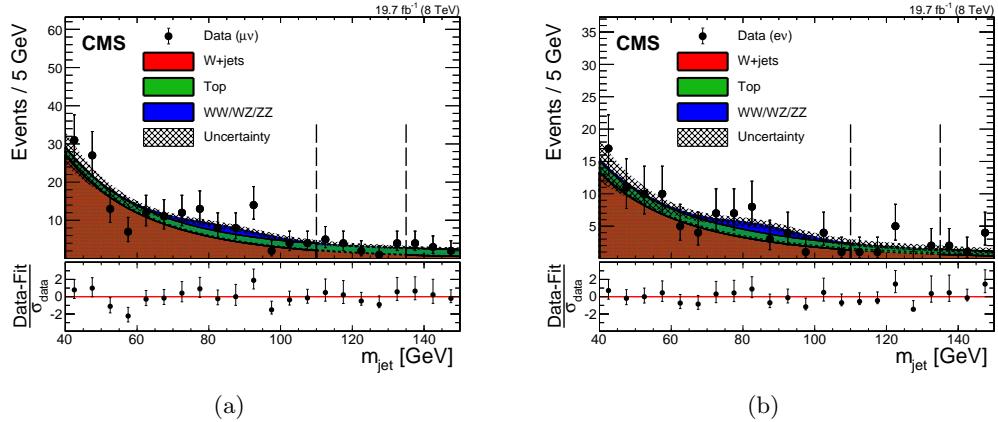


Figure 8.7: Distributions of the pruned jet mass m_{jet} in the muon (a) and electron (b) channels for the $\ell\nu+H$ -jet analysis at 8 TeV. All selections are applied except the requirement on the m_{jet} signal window. The signal region lies between the dashed vertical lines. The hatched region indicates the statistical uncertainty of the fit. At the bottom of each plot, the bin-by-bin fit residuals, $(N_{\text{data}} - N_{\text{fit}})/\sigma_{\text{data}}$, are shown.

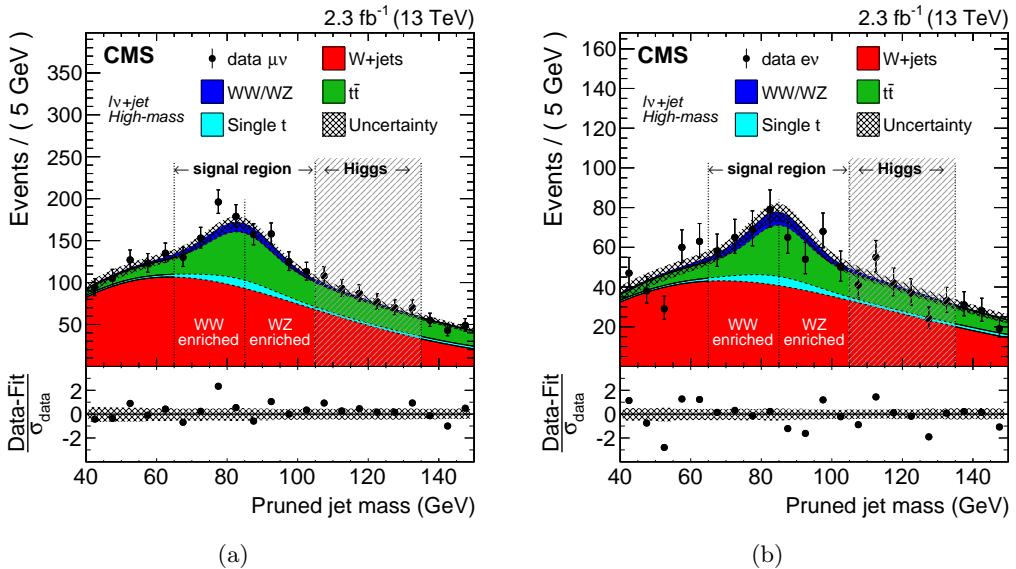


Figure 8.8: Distributions of the pruned jet mass m_{jet} in the muon (a) and electron (b) channels for the $\ell\nu+V$ -jet analysis at 13 TeV. All selections are applied except the requirement on the m_{jet} signal window. The signal regions and m_{jet} categories of the analysis are indicated by the vertical dotted lines. The shaded m_{jet} region 105–135 GeV is not used in the analysis. At the bottom of each plot, the bin-by-bin fit residuals, $(N_{\text{data}} - N_{\text{fit}})/\sigma_{\text{data}}$, are shown together with the uncertainty band of the fit normalized by the statistical uncertainty of data, σ_{data} .

The form of the $m_{\ell\nu j}$ distribution for the W+jets background in the signal region (SR) is determined from the lower m_{jet} sideband, through the transfer function $\alpha_{\text{MC}}(m_{\ell\nu j})$ obtained from the W+jets simulation, and defined as:

$$\alpha_{\text{MC}}(m_{\ell\nu j}) = \frac{F_{\text{MC},\text{SR}}^{\text{W+jets}}(m_{\ell\nu j})}{F_{\text{MC},\text{SB}}^{\text{W+jets}}(m_{\ell\nu j})}, \quad (8.2)$$

where $F_{\text{MC,SB}}^{\text{W+jets}}$ and $F_{\text{MC,SR}}^{\text{W+jets}}$ are the probability density functions used to describe the simulated $m_{\ell\nu j}$ spectrum in the lower m_{jet} sideband and signal region, respectively. The upper m_{jet} sideband is not considered since the W+jets shape is different here compared to what expected in the lower sideband. Furthermore, the upper sideband suffers from a larger $t\bar{t}$ background contamination.

Since the lower sideband region does not represent a perfectly pure sample of W+jets events in data, the presence of minor backgrounds is subtracted from the observed diboson invariant mass distribution to obtain an estimation of the W+jets contribution in the sideband control region of the data, $F_{\text{data,SB}}^{\text{W+jets}}(m_{\ell\nu j})$.

The $m_{\ell\nu j}$ range used in the estimate of the background distribution determines the region of masses probed by these searches. This range is chosen to ensure a smoothly falling background spectrum, and therefore far enough from the kinematic turn-on at low masses generated by the acceptance selections, allowing for a good stability and a robust control of the background estimation. For this reason the low edge of the range is chosen at 0.7 TeV while the high edge is chosen such that it is not too far from the last value where data are still present. Therefore, the fits are performed in the range $0.7 < m_{\ell\nu j} < 4$ TeV for the 13 TeV analysis, while at 8 TeV no data are present above $m_{\ell\nu j} = 3$ TeV and the chosen range is $0.7 < m_{\ell\nu j} < 3$ TeV.

To describe the smoothly falling W+jets background distribution, a parametrization of the form of a leveled exponential is adopted, defined as

$$F_{\text{ExpTail}}(x) = e^{-\frac{x}{a+bx}}. \quad (8.3)$$

This functional form is found to adequately describe the simulation in both the signal region and the low sideband as shown in Figure 8.9. Tests are performed with alternative functional forms, and the background prediction is found to agree with the one of the default function within the uncertainties. The minor background contributions are parametrized with a simple exponential functional form, except for the diboson contribution for which the $F_{\text{ExpTail}}(x)$ defined above is used.

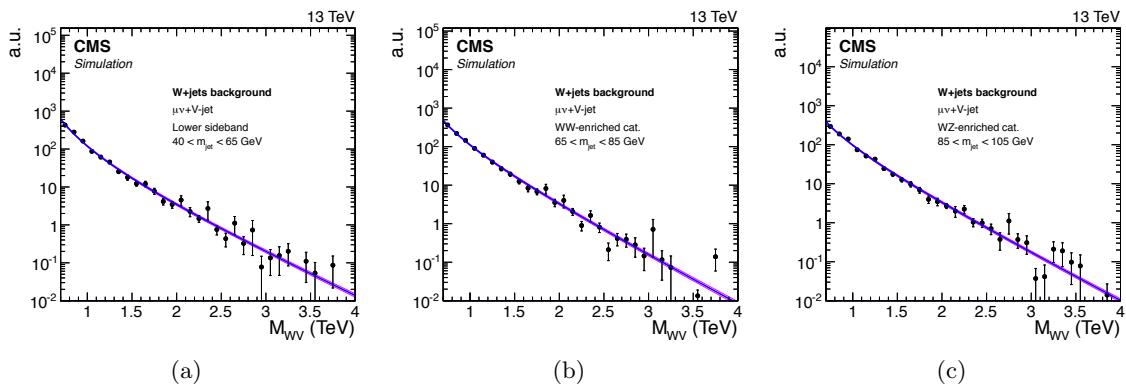


Figure 8.9: Functional form describing the m_{WW} spectrum of the W+jets background after fitting the simulation data, for the $\mu\nu+V\text{-jet}$ channel. The distributions for the lower m_{jet} sideband (a), and the WW-enriched (b) and WZ-enriched (c) signal regions are shown.

Figure 8.10 shows the transfer functions $\alpha_{\text{MC}}(m_{\text{WW}})$ for the WW-enriched and WZ-enriched categories in the $\mu\nu+V\text{-jet}$ channel obtained from a simultaneous fit of W+jets simulated events in the lower sideband and in the signal region using the parametrization defined in Equation 8.3. The blue and the red lines represent the probability density functions describing the W+jets background with m_{jet} in the lower sideband and signal

region, respectively, and given by the leveled-exponential function of Eq. 8.3. A simultaneous fit is performed of the two distributions, where the parameters used to model the shape in the signal region are correlated with the ones used to model the shape in the sideband. The transfer function $\alpha_{\text{MC}}(m_{\text{WV}})$ is shown as a solid black line, while the dark (light) shaded region corresponds to the 1σ (2σ) statistical uncertainty of the fit. These uncertainties only represent the uncertainty in the modelling of the W+jets shape. The bands have a size ≈ 0 around 2 TeV as the $\alpha_{\text{MC}}(m_{\text{WV}})$ is the ratio of two probability density functions which have in order to cross to conserve the total probability. Similar results are obtained for the $\ell\nu+\text{H-jet}$ analysis.

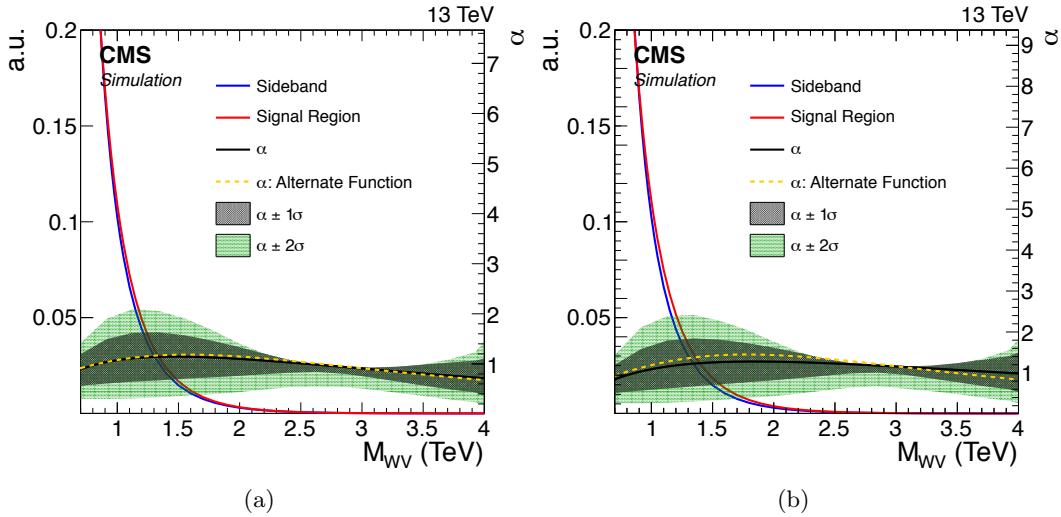


Figure 8.10: The transfer functions $\alpha_{\text{MC}}(m_{\text{WV}})$ from the lower m_{jet} sideband to the signal region for the WW-enriched (a) and WZ-enriched (b) categories in the $\mu\nu+\text{V-jet}$ channel. The grey area represents the statistical uncertainty of the fit. The blue and the red lines represent the probability density functions describing the W+jets background with m_{jet} in the lower sideband and signal region, respectively. The $\alpha_{\text{MC}}(m_{\text{WV}})$ obtained fitting the W+jets with and alternative function is shown as yellow line.

In Figure 8.11, the result of the fit to the m_{WV} distribution of the data with m_{jet} in the lower sideband is shown for the electron and muon channels of the $\ell\nu+\text{V-jet}$ analysis. From this fit, an estimation of $F_{\text{data},\text{SB}}^{\text{W+jets}}(m_{\ell\nu j})$ is obtained. Finally, the W+jets background distribution in the signal region is then extrapolated by rescaling $F_{\text{data},\text{SB}}^{\text{W+jets}}$ by $\alpha_{\text{MC}}(m_{\ell\nu j})$. The minor backgrounds are then added to the W+jets background to obtain the total SM prediction in the signal region, which is given by

$$N_{\text{SR}}^{\text{bkg}}(m_{\ell\nu j}) = N_{\text{SR}}^{\text{W+jets}} \times \alpha_{\text{MC}}(m_{\ell\nu j}) \times F_{\text{data},\text{SB}}^{\text{W+jets}}(m_{\ell\nu j}) + \sum_k N_{\text{SR}}^k \times F_{\text{MC},\text{SR}}^k(m_{\ell\nu j}). \quad (8.4)$$

where the sum runs over the products of the normalization $N_{\text{MC},\text{SR}}^k$ and shape $F_{\text{MC},\text{SR}}^k(m_{\ell\nu j})$ of each minor background contribution k , while $N_{\text{SR}}^{\text{W+jets}}$ and $F_{\text{data},\text{SB}}^{\text{W+jets}}(m_{\ell\nu j})$ represent the normalization and shape of the W+jets background derived from data as described previously in this chapter. The transfer function $\alpha_{\text{MC}}(m_{\ell\nu j})$ accounts for small kinematic differences between the signal region and the sideband one.

Results of the final background extraction in the signal region will be presented in Sections 9 and 10 or the $\ell\nu+\text{H-jet}$ and $\ell\nu+\text{V-jet}$ analysis, respectively.

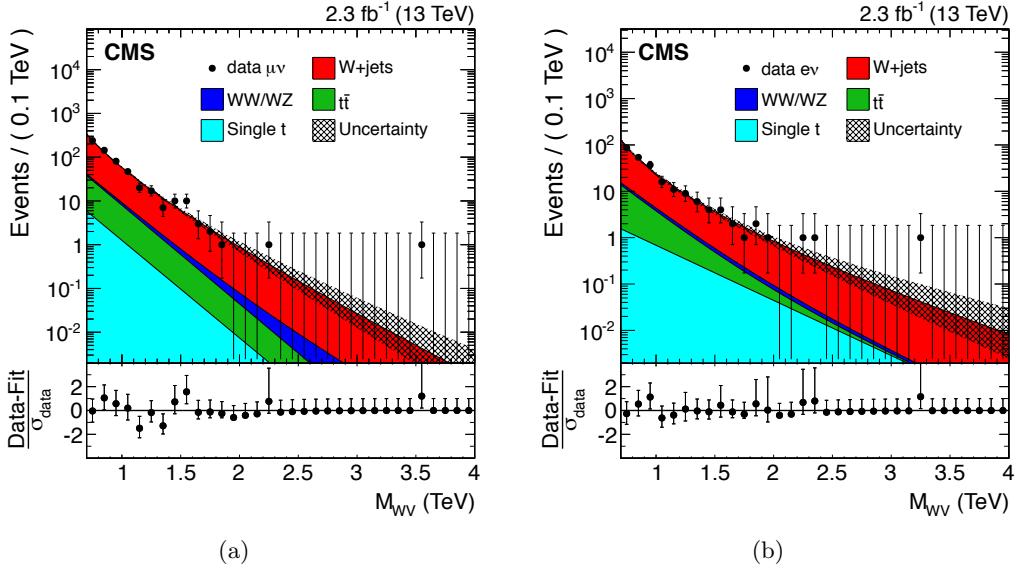


Figure 8.11: Results of the fit to the m_{WW} distribution of the data with m_{jet} in the lower sideband to estimate $F_{\text{data,SB}}^{\text{W+jets}}$ for both muon (a) and electron (b) channels of the $\ell\nu+\text{V-jet}$ analysis. Minor backgrounds are estimated from simulation, while the W+jets contribution is the result of the fit to the data.

8.3.4 Validation of the α method

To test the validity and the robustness of the data driven method for the W+jets described previously in this section, a closure test is performed. In this test, the background is extracted to a signal free control region that allows to check the compatibility with data for both the distribution and normalization. In order to achieve this, the low mass sideband defined in Table 8.3 is divided into two regions: $40 < m_{jet} < 55 \text{ GeV}$, referred to as "region A" is used as sideband, while $55 < m_{jet} < 65 \text{ GeV}$, referred to as "region B" is used as signal region. The W+jets background normalization is then predicted in region B by performing a fit to the m_{jet} distribution of the data in region A and in the upper sideband (Table 8.3), while its distribution in $m_{\ell\nu j}$ is extrapolated in region B with a fit of the data in region A and a suitable transfer function $\alpha_{\text{MC}}(m_{\ell\nu j})$. In this test, the $\alpha_{\text{MC}}(m_{\ell\nu j})$ is defined as the ratio between the simulated W+jets background distributions in $m_{\ell\nu j}$ in region B and A.

Figure 8.12 shows the results of the fit to the m_{jet} distribution of the data for $\mu\nu+\text{V-jet}$ events inside the region A and the HSB, performed to extract the W+jets prediction inside the region B.

Figure 8.13(a) shows the transfer function $\alpha_{\text{MC}}(m_{\ell\nu j})$ obtained from a simultaneous fit of W+jets simulated events in the region A and in the region B, using the leveled-exponential parametrization defined in Equation 8.3. In Fig. 8.13(b), the result of the fit to the $m_{\ell\nu j}$ distribution of the data with m_{jet} in the lower sideband is shown for $\mu\nu+\text{V-jet}$ events, where the W+jets shape is modelled through the same leveled-exponential function.

Finally, Fig. 8.14 shows a comparison between the total predicted background, obtained through Equation 8.4, and the data inside the signal free region B. A good agreement is found over the whole $m_{\ell\nu j}$ range. The test has been performed for both lepton flavours for the $\ell\nu+\text{V-jet}$ analysis, as well as for the $\ell\nu+\text{H-jet}$ analysis where slightly different definitions for region A and B are used. In all the cases, consistency between the predicted background and the data is observed, thus validating the proposed strategy for the W+jets background

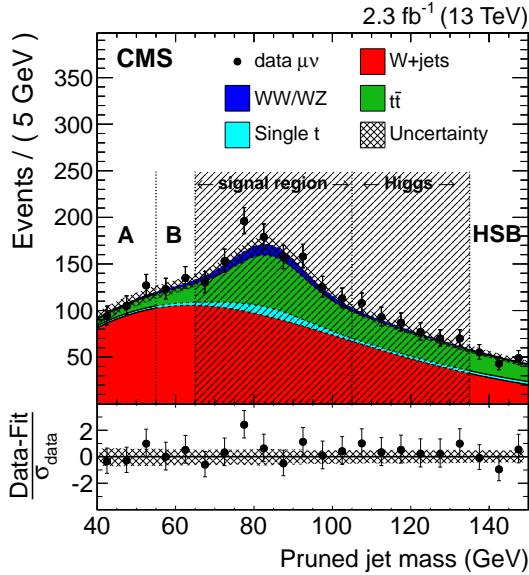


Figure 8.12: blabla

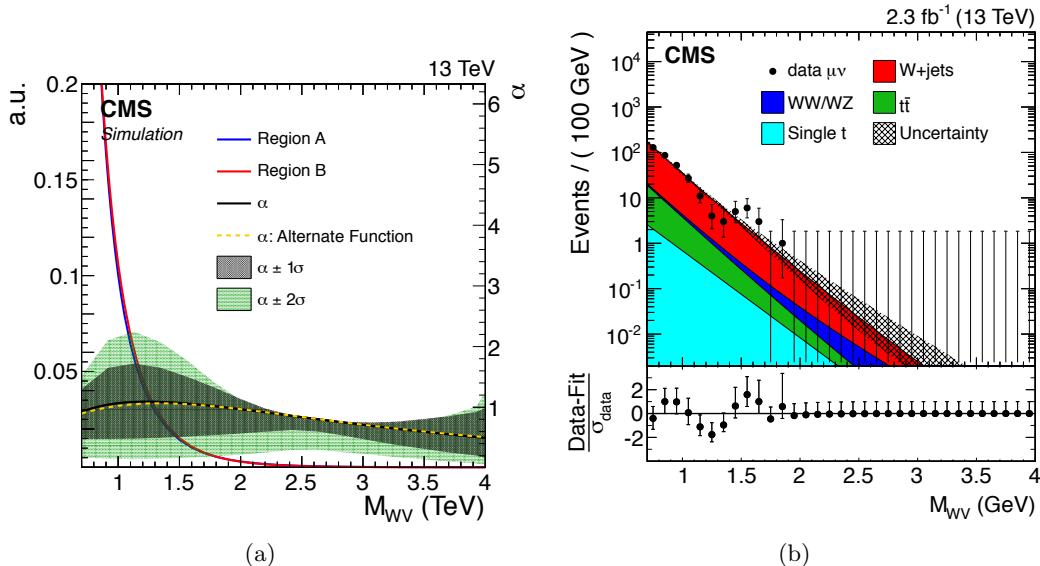


Figure 8.13: blabla

estimation.

FIXME: consider adding results with dijet function in the appendix

8.4 Signal modeling

The potential discovery and exclusion power of these analyses rely on the ability of finding a local enhancement on the top of a smoothly falling background. This is ultimately achieved through an unbinned likelihood fit of the signal + background model to the reconstructed diboson invariant mass, which depends on the accurate description of the signal shape.

An analytical parametrization of the signal shape is chosen such that it well reproduces

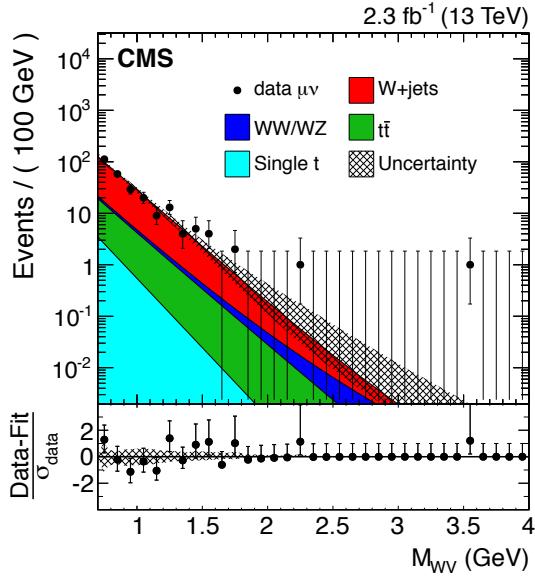


Figure 8.14: blabla

the simulated resonance distributions. As stated in Section 5.2.1, simulated signal events are generated with a resonance natural width sufficiently small compared to the detector resolution. This makes the model used for generating the events independent from the detector effects on the signal shape, allowing a model independent search for narrow resonances where only the detector resolution has to be described. A double-sided Crystal-Ball (CB) function [98] (i.e. a Gaussian core with power law tails on both sides) is found to well serve this purpose. To take into account differences between muon and electron momentum resolutions, the signal invariant mass distribution is parametrized separately in the two lepton flavor categories.

Figure 8.15 shows an example for the fitted signal distribution through a CB function for a W' mass of 1.5 TeV in the muon channel of the $\ell\nu+H$ -jet analysis. Other examples are shown in Fig. 8.16, for the $\ell\nu+V$ -jet analysis in the muon channel, for several signal hypotheses and for both m_{jet} categories.

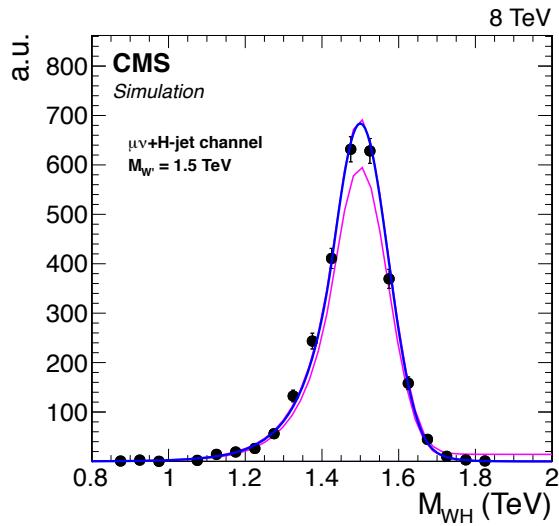


Figure 8.15: blabla

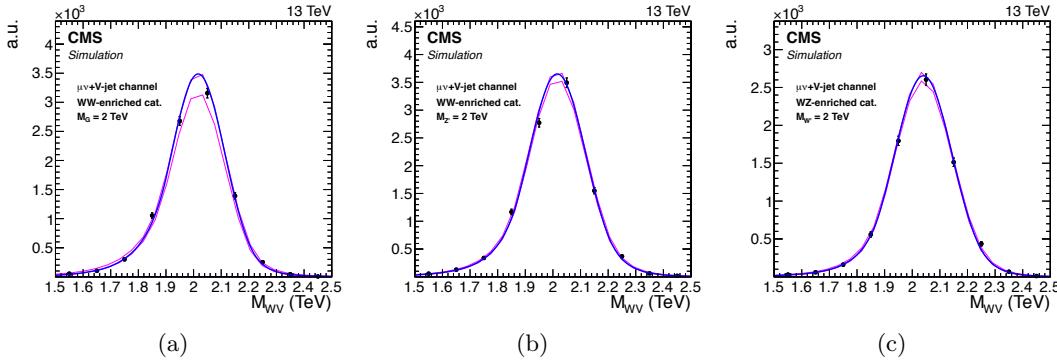


Figure 8.16: blabla

Because of the limited number of available simulated samples, a linear interpolation is performed for each parameter of the CB function between the shapes obtained for some reference mass points, in order to extrapolate the distribution for intermediate values of the resonance mass. The resolution of the reconstructed diboson invariant mass is given by the width of the Gaussian core and it ranges between 7 and 4% depending on the resonance mass, as shown in Fig. 8.17. The resolution is dominated by the jet and E_T^{miss} contributions.

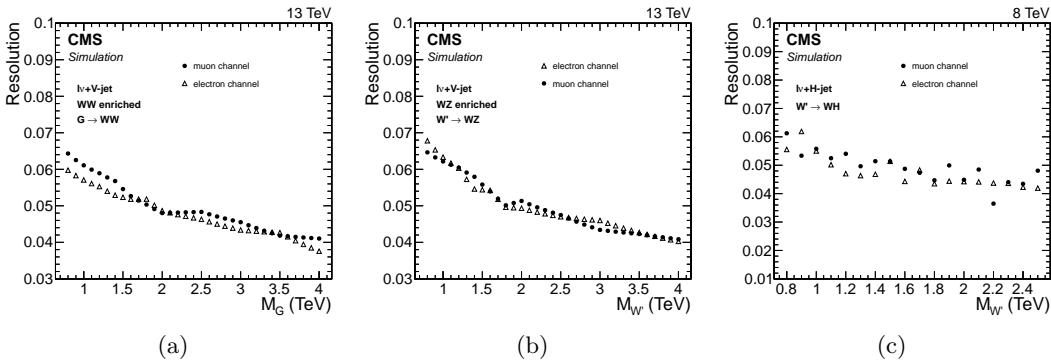


Figure 8.17: blabla

The signal selection efficiency, evaluated for each category, is defined as the number of selected signal events over the number of generated ones, which include all the possible lepton flavours (e , μ and τ). As shown in Figure 8.18 the efficiency for a Z' or bulk graviton signal in the WW-enriched category is ≈ 2 times larger compared to a W' signal. On the other hand, the efficiency for a W' signal in the WZ-enriched category is ≈ 4 times larger compared to a Z' or bulk graviton signal. For both categories and for each signal hypothesis the efficiency is smaller compared to the large m_{jet} window used for V tagging. However, the resulting loss in sensitivity in each of the category is recovered with a combination of the two m_{jet} categories which allows the use of all the available data. With this solution the discrimination between the two type of signals is maximized together with a gain in sensitivity of 10–20% depending on the resonance mass.

A linear interpolation of the signal efficiency is performed between the values obtained for some reference mass points in order to extrapolate the efficiency for intermediate resonance masses for which a simulated sample is not available. The efficiency for the electron channel is lower compared to the muon channel over most of the phase space due to the tighter requirements on the electron p_T and E_T^{miss} . This effect is less visible in the $\ell\nu+H$ -jet channel

(see Fig. 8.19) where the electron selections are less strict. For all cases, at low masses the efficiency increases with the resonance mass because of the increase in the acceptance of the lepton, E_T^{miss} and $m_{\text{WV}/\text{WH}}$ selections together with the inefficiency of the jet algorithms in reconstructing the merged jet for a low boosted V boson (**FIXME: point to figure**). At larger resonance masses the efficiency slightly decrease due to τ_{21} selection inefficiency for high p_T V jets, as described in Section 7.2. For the electron channel this effect is compensated by a larger increase in the lepton selection acceptance, resulting in a nearly flat efficiency at high resonance masses. Similar considerations hold for the efficiency in the $\ell\nu+\text{H-jet}$ channel shown in Fig. 8.19.

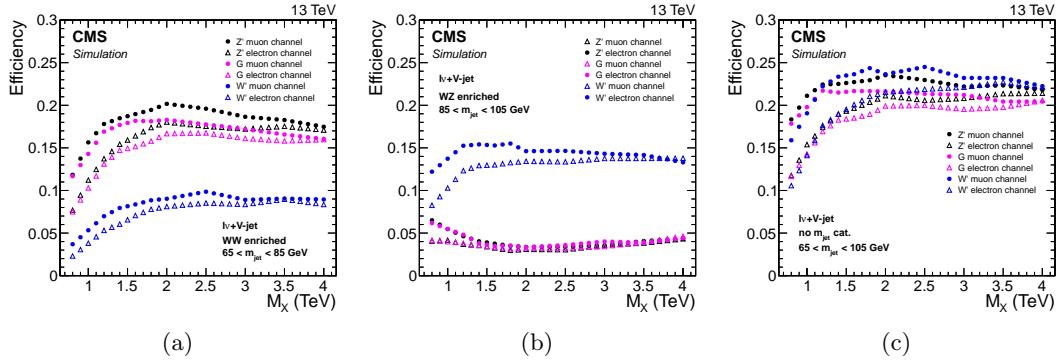


Figure 8.18: blabla

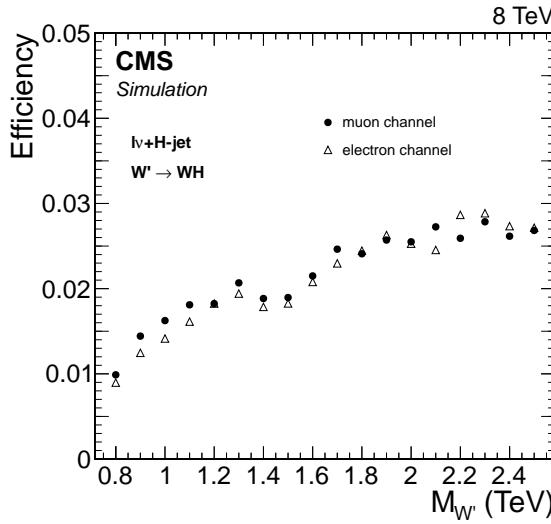


Figure 8.19: blabla

8.5 Systematic uncertainties

This section describes the systematic uncertainties in the signal ad background predictions affecting both the normalizations and the $m_{\ell\nu j}$ distributions. The uncertainties described below are include as nuisance parameters in the calculation of the limits on the cross section as well as of the p-values of potential excesses of events observed in the data.

8.5.1 Systematic uncertainties in the background estimation

The uncertainty on the W+jets background normalization is mainly due to the uncertainties on the parameters extracted from the fit of the data in the pruned jet mass sideband. This contribution is statistical in nature since it depends on the amount of data in the m_{jet} sideband regions, and it is evaluated varying the fit parameters from the final fit values by random amounts sampled from the covariance matrix. An additional effect due to the difference arising from alternative parametrization of the W+jets m_{jet} distribution is taken into account and added in quadrature to the pure statistical contribution. This contribution is found to constitute up to 15% of the total uncertainty. The total uncertainty on the W+jets yields remains below 10% in the $\ell\nu$ +V-jet channel, while uncertainties above 40% are obtained for the $\ell\nu$ +H-jet channel where the amount of data in the sidebands is largely reduced by the tight b tagging requirements.

As described in Section 8.3.3 the extrapolated background shape in the signal region is computed from the product of $F_{\text{data,SB}}^{\text{W+jets}}(m_{\ell\nu j})$ and $\alpha_{\text{MC}}(m_{\ell\nu j})$. Thus, the shape uncertainty comes from both uncertainties in the W+jets $m_{\ell\nu j}$ shape obtained from the fit of the data in the lower m_{jet} sideband region and in the modelling of the transfer function $\alpha_{\text{MC}}(m_{\ell\nu j})$. Both contributions are mainly statistical in nature, as they are driven by the available amount of data in the sideband and by the number of simulated W+jets events passing the analysis requirements, respectively. These effects are estimated from the covariance matrix of the fit and included in the final limit and p-value calculations after a procedure which diagonalizes the matrix to decorrelate the fitted parameters. In this procedure, the new parameters are defined in such a way to be centered at zero and with error equal to unity. The background fit parameterization is then redefined as a function of these new, uncorrelated parameters. This new fit function together with the uncertainties on the fitted parameter is used to describe the background distribution in the limit and p-value calculations explained in Section 8.6. As shown in Fig. 8.10, the $\alpha_{\text{MC}}(m_{\ell\nu j})$ presents variations due to the choice of the parametrization used to model the W+jets $m_{\ell\nu j}$ distribution. Previous studies showed that additional variations of about the same size are due to the use of different parton showering algorithms [91]. This effect has been evaluated comparing the α obtained with simulated samples with parton showering implemented through HERWIG++ and PYTHIA. These variations are found to be equal or slightly smaller than the statistical uncertainties on the α , and hence the associated systematic effect is taken into account by enlarging the errors on the decorrelated fit parameters by a factor $\sqrt{2}$. This is sufficiently conservative to cover all the shape variations.

In a similar way, variations in the $F_{\text{data,SB}}^{\text{W+jets}}(m_{\ell\nu j})$ due to the same effects, are as well taken into account.

The uncertainties in the W+jets normalization are treated as uncorrelated among the different lepton flavor channels and m_{jet} mass categories while the uncertainties on the W+jets $m_{\ell\nu j}$ distribution are partially correlated according to the following scheme:

- uncertainties on the $F_{\text{data,SB}}^{\text{W+jets}}(m_{\ell\nu j})$ parameters are correlated;
- uncertainties on the $\alpha_{\text{MC}}(m_{\ell\nu j})$ parameters are uncorrelated.

This solution takes into account the fact that in the different m_{jet} categories the same data are used to estimate the W+jets $m_{\ell\nu j}$ distribution in the sideband, while the transfer function is used to predict the shape in the two orthogonal signal regions given defined by the categories.

The systematic uncertainty in the normalization of the $t\bar{t}$ /single top quark backgrounds is driven by the uncertainties in the data-to-simulation scale factors estimated in the top

quark enriched control sample (see Section 8.2). In the $\ell\nu+V$ -jet analysis these uncertainties are measured to be 4.6% and 8.4% in the muon and electron channel, respectively. For the $\ell\nu+H$ -jet analysis, this uncertainty amounts to 5.6%. For the single top quark background an additional systematic uncertainty related to the cross section calculations is assigned to be 15% and 5%, for the 8 and 13 TeV analysis respectively [99, 100].

The $t\bar{t}$ background distribution in $m_{\ell\nu j}$ is taken from simulation and this choice is found to be reasonable given the agreement between data and simulation in the top quark enriched control sample as shown in Section 8.2. However, previous studies [91] showed that variations in the shape occur due to the choices of regularization or factorization scales (varied up and down by a factor of 2), to the matching scales in the MADGRAPH simulation, and to different generators (MADGRAPH or POWHEG). In order to cover all these effects, the errors on the decorrelated fit parameters for the $t\bar{t}$ background distribution in $m_{\ell\nu j}$ is enlarged by a factor of 2.

The systematic uncertainties in the diboson normalization is due to the uncertainty in the inclusive cross sections, which are assigned to be 10% [101] and 3% [102] for the 8 and 13 TeV analysis, respectively. For the $\ell\nu+V$ -jet analysis, the uncertainty in the diboson background normalization is as well due to the uncertainty of 3% in the measured data-to-simulation scale factors for the V tagging efficiency derived in the top quark enriched control sample as described in Section 7.2.

Additional source of systematic uncertainties in the background normalization are due to the uncertainty in the integrated luminosity, and in the measured data-to-simulation scale factors for the efficiency of lepton trigger and identification, described in the following section. A summary of the systematic uncertainties in the normalization of the predicted background is provided in Tables 8.5 and 8.6 for the 8 and 13 TeV analysis, respectively.

Source	W+jets	$t\bar{t}$	single top quark	diboson
Integrated luminosity		2.6%	2.6%	2.6%
Cross section	-	-	15%	10%
Data-driven prediction	42% (μ) / 59% (e)	5.6%	5.6%	-
Lepton trigger (μ/e)	-	1% / 1%	1% / 1%	1% / 1%
Lepton identification (μ/e)	-	1% / 3%	1% / 3%	1% / 3%

Table 8.5: Summary of the systematic uncertainties in the normalization of the predicted background in the $\ell\nu+H$ -jet analysis at 8 TeV.

Source	W+jets	$t\bar{t}$	single top quark	diboson
Integrated luminosity		2.7%	2.7%	2.7%
Cross section	-	-	5%	3%
V-tagging efficiency	-	-	-	3%
Data-driven prediction	5–9%	5–8%	5–8%	-
Lepton trigger (μ/e)	-	1% / 1%	1% / 1%	1% / 1%
Lepton identification (μ/e)	-	1% / 3%	1% / 3%	1% / 3%

Table 8.6: Summary of the systematic uncertainties in the normalization of the predicted background in the $\ell\nu+V$ -jet analysis at 13 TeV.

8.5.2 Systematic uncertainties in the signal prediction

Systematic uncertainties affecting the predicted signal efficiency (or normalization) and $m_{\ell\nu j}$ distribution arise from several sources as described in the following and summarized in Tables 8.7 and 8.8. The effect of each source is evaluated for each considered simulated signal hypothesis as a function of the resonance mass.

One the primary sources affecting the signal normalization for the $\ell\nu+V$ -jet channel is due to uncertainties in data-to-simulation scale factors for the V tagging efficiency, derived from top quark enriched control sample as described in Section 7.2. These uncertainties include separately the uncertainty of 3% on the scale factor measured in $t\bar{t}$ events with an average $p_T \approx 200$ GeV, and the uncertainty due to the extrapolation of the scale factor to higher momenta, which is assigned to be 6–10% depending on the signal mass. Additional uncertainties are assigned due to the pruned jet mass scale and resolution measured in $t\bar{t}$ events (Table 7.2). These are computed by rescaling or smearing the m_{jet} value according to the uncertainties in the respective m_{jet} scale or resolution. The selection efficiencies are recalculated on these modified events, with the resulting changes taken as systematic uncertainties that depend on the resonance mass.

In a similar way, systematic uncertainties are assigned in the $\ell\nu+H$ -jet due to the uncertainty in the H tagging efficiency. This contribution arises from both uncertainties in the data-to-simulation scale factors for the pruned jet mass scale and resolution, derived from the top quark enriched control sample with 8 TeV data (Table 7.3), and for b-tagged jet identification efficiencies (Section 6.4.3). These sources introduce a systematic uncertainty in the mass tagging and b tagging of the Higgs boson of 2–10% and 2–8%, respectively, depending on the signal mass.

The accuracy on energy and momentum measurements for leptons and jets represents an important source of systematic uncertainties in the signal efficiency. In particular, the muon momentum scale and resolution, the electron energy scale and resolution, and the jet energy scale and resolution are considered. The event selection is applied to the signal samples after varying the lepton four-momenta within one standard deviation of the corresponding uncertainty in the muon momentum scale [54] or electron energy scale [103], or applying an appropriate Gaussian momentum/energy smearing in case of resolution uncertainties. The same procedure is also applied for the jet four-momenta using the corresponding energy scale and resolution uncertainties (Section 6.4.2). In this process, variations in the lepton and jet four-momenta are propagated consistently to the \vec{E}_T^{miss} vector. The signal efficiency is then recalculated using modified lepton and jet four-momenta separately for each source of systematic uncertainties. The largest relative change in the signal efficiency compared to the default value is taken as the systematic uncertainty for that specific source. The induced relative migration among V jet mass categories is evaluated for the $\ell\nu+V$ -jet channel, but do not affect the overall signal efficiency. The muon, electron, and jet uncertainties are assumed to be uncorrelated. Finally, the resulting changes on the reconstructed resonances are propagated on the reconstructed $m_{\ell\nu j}$ signal shape, resulting in a small effect on both peak position and width of the Gaussian core.

The systematic uncertainties in the lepton trigger, identification, and isolation efficiencies are derived using a dedicated T&P analysis in $Z \rightarrow \ell^+\ell^-$ events. For both analysis channels, an uncertainty of 1% is assigned to the trigger efficiency for both lepton flavors, while for lepton identification and isolation efficiency, the systematic uncertainty is estimated to be 1% for the muon and 3% for electron flavors.

The 2.7% and 2.6% uncertainty in the integrated luminosity affects to the normalization of both signal and backgrounds in the $\ell\nu+V$ -jet and $\ell\nu+H$ -jet, respectively, as obtained in measurements performed for the 2015 and 2012 data taking periods [104, 105]. For the

$\ell\nu + V$ -jet channel, uncertainties on the signal yield due to variations in the parton distribution function and the choice of factorization (μ_f) and renormalization (μ_r) scales are also taken into account. The PDF uncertainties are evaluated using the NNPDF 3.0 [21] PDFs. The uncertainty related to the choice of μ_f and μ_r scales is evaluated following the proposal in Refs. [106, 107] by varying the default choice of scales in the following 6 combinations of factors: $(\mu_f, \mu_r) \times (1/2, 1/2), (1/2, 1), (1, 1/2), (2, 2), (2, 1)$, and $(1, 2)$. The uncertainty in the signal cross section from the choice of PDFs and of factorization and renormalization scales ranges from 4 to 77%, and from 1 to 22%, respectively, depending on the resonance mass, particle type and its production mechanism. For the $\ell\nu + H$ -jet channel, only the impact of the proton PDF uncertainties on the signal efficiency is evaluated with the PDF4LHC prescription [108, 109], using the MSTW2008 [110] and NNPDF2.1 [111] PDF sets. This effect is found to be $< 0.5\%$.

Finally, the systematic uncertainty due to the modelling of pileup is estimated by reweighting the signal simulation samples such that the distribution of the number of interactions per bunch crossing is shifted according to the uncertainty in the inelastic proton-proton cross section compared with that found in data. This contribution is found to be 0.5% in both channels.

Table 8.7: Summary of the signal systematic uncertainties for the $\ell\nu + H$ -jet analysis and their impact on the event yield in the signal region and on the reconstructed m_{WH} shape (mean and width) for both muon and electron channels.

Source	Relevant quantity	Uncertainty (%)
Lepton trigger (μ/e)	Signal yield	1 / 1
Lepton identification (μ/e)	Signal yield	1 / 3
Lepton p_T scale (μ/e)	Signal yield	1 / 0.5
Lepton p_T resolution (μ/e)	Signal yield	0.1 / 0.1
Jet energy scale	Signal yield	1–3
Jet energy resolution	Signal yield	0.5
Integrated luminosity	Signal yield	2.6
Pileup	Signal yield	0.5
PDFs	Signal yield	< 0.5
H jet mass tagging efficiency	Signal yield	2–10
H jet b tagging efficiency	Signal yield	2–8
Jet energy scale	Resonance shape (mean)	0.5
Jet energy scale	Resonance shape (width)	4
Jet energy resolution	Resonance shape (mean)	0.2
Jet energy resolution	Resonance shape (width)	4
Lepton p_T resolution	Resonance shape (mean)	0.1
Lepton p_T resolution	Resonance shape (width)	1.2
Lepton p_T scale	Resonance shape (mean)	0.7
Lepton p_T scale	Resonance shape (width)	2.5

8.6 Testing new resonance hypothesis

The purpose of this analysis is to infer a constraint on the existence of a new resonance decaying into diboson for a set of different signal mass hypotheses. The comparison between the diboson invariant mass distribution observed in data and the SM background prediction is used to check for the presence of the new resonance. A hypothesis test is built to decide

Table 8.8: Summary of the signal systematic uncertainties for the $\ell\nu + V$ -jet analysis and their impact on the event yield in the signal region and on the reconstructed m_{WV} shape (mean and width) for both muon and electron channels. The last uncertainty results in migrations between event categories, but does not affect the overall signal efficiency.

Source	Relevant quantity	Uncertainty (%)
Lepton trigger (μ/e)	Signal yield	1 / 1
Lepton identification (μ/e)	Signal yield	1 / 3
Lepton p_T scale (μ/e)	Signal yield	0.7 / 0.2
Lepton p_T resolution (μ/e)	Signal yield	0.1 / 0.1
Jet energy and m_{jet} scale	Signal yield	0.2–4
Jet energy and m_{jet} resolution	Signal yield	0.1–2
Integrated luminosity	Signal yield	2.7
Pileup	Signal yield	0.5
PDFs (W')	Signal yield	4–19
PDFs (Z')	Signal yield	4–13
PDFs (G_{bulk})	Signal yield	9–77
Scales (W')	Signal yield	1–14
Scales (Z')	Signal yield	1–13
Scales (G_{bulk})	Signal yield	8–22
V tagging efficiency	Signal yield	3
V tagging p_T -dependence	Signal yield	6–10
Jet energy scale	Resonance shape (mean)	1.3
Jet energy scale	Resonance shape (width)	3
Jet energy resolution	Resonance shape (mean)	0.1
Jet energy resolution	Resonance shape (width)	3
Lepton p_T resolution	Resonance shape (mean)	0.1
Lepton p_T resolution	Resonance shape (width)	0.1
Lepton p_T scale	Resonance shape (mean)	0.1
Lepton p_T scale	Resonance shape (width)	0.5
Jet energy and m_{jet} scale	Migration	2–24

between a null hypothesis given by the predicted SM background only against an alternative hypothesis which includes both background as well as the sought after signal. In principle one can either test the background-only hypothesis and exclude it if there is a large deviation of the data from the SM background prediction, or test the signal hypothesis and exclude it if there is a large deviation of the data from the expected signal model. In particular, if no significant deviation from the SM background prediction is observed in data, compatible with the signal hypothesis, an upper limit on production cross section of such resonances is usually set, up to a certain degree of belief. The CMS community has agreed upon a procedure for computing upper limits, which is based on the modified frequentist method, often referred to as CL_s . While a detailed description of such method can be found in Refs. [112, 113], the basic ingredients will be summarized Section 8.6.1. A description of the procedure followed to quantify an excess of events is provided in Section 8.6.3. A summary of the final results will be given in the next chapter.

8.6.1 Limit setting procedure

The procedure to establish the exclusion of a given signal hypothesis is based on a

frequentist significance test which uses a log-likelihood ratio as a test statistic. In order to construct the test statistic a likelihood function is defined as

$$\mathcal{L}(\text{data}|\mu, \theta) = \text{Poisson}(\text{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \quad (8.5)$$

In this definition, s and b denote the expected signal and background event yields, respectively, which prior to the scrutiny of the observed data entering the statistical analysis, are subject to multiple uncertainties that are handled by introducing nuisance parameters θ , so that signal and background expectations become functions of the nuisance parameters: $s(\theta)$ and $b(\theta)$. The exclusion of a signal hypothesis is generally expressed as an upper limit on the *signal strength modifier* μ on the cross section used as input in the evaluation of the expected signal yields. With this definition, the likelihood represents the Poisson probability of observing a certain amount of data when the expected yield is $\mu \cdot s(\theta) + b(\theta)$ and given the probability $p(\tilde{\theta}|\theta)$ of measuring a value $\tilde{\theta}$ for the nominal nuisance parameter θ . Note that, in this likelihood definition, “data” stands for a generic dataset, either experimental or a pseudo-data generated randomly.

The likelihood can be either binned or unbinned. In the first case the $\text{Poisson}(\text{data}|\mu \cdot s + b)$ in Equation 8.5 is the product of Poisson probabilities for observing n_i events in each bin i of the signal+background model

$$\prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}. \quad (8.6)$$

For the unbinned case each even enters the calculation

$$k^{-1} \prod_i (\mu S f_s(x_i) + B f_b(x_i)) e^{-(\mu S + B)}, \quad (8.7)$$

where $f_s(x)$ and $f_b(x)$ are the probability density functions of signal and background of the observable x , while S and B are total event rates expected for signal and backgrounds. In this analysis the unbinned likelihood is used, where the observable x coincides with the reconstructed diboson invariant mass.

To compare the compatibility of the data with the background-only and signal+background hypotheses, where the prediction for the signal is allowed to be scaled by some factor μ , the test statistic \tilde{q}_μ is constructed based on the profile likelihood ratio as

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \quad \text{with} \quad 0 \leq \hat{\mu} \leq \mu \quad (8.8)$$

Here $\hat{\theta}_\mu$ denotes the value of θ that maximizes the likelihood for the hypothesized μ , i.e. it is the conditional maximum-likelihood (ML) estimator of θ (and thus is a function of μ). The procedure of refitting the nuisance parameters to maximize the likelihood for each possible value of the parameter of interest μ , is usually referred to as “profiling”. The denominator is the maximized (unconditional) likelihood function, i.e., $\hat{\mu}$ and $\hat{\theta}$ are the global maximum of the likelihood. The presence of the nuisance parameters broadens the profile likelihood as a function of μ relative to what one would have if their values were fixed. This reflects the loss of information about μ due to the systematic uncertainties. Higher values of \tilde{q}_μ correspond to increasing incompatibility between the data and the hypothesized signal of strength μ . The lower constraint for $\hat{\mu}$ in the denominator excludes the possibility of negative signal yields. The upper constraint is introduced to avoid that data with $\hat{\mu} > \mu$ (upward fluctuations) are considered as representing less compatibility with μ than what obtained with data.

The observed value of the test statistic, $\tilde{q}_\mu^{\text{obs}}$ for the given signal strength modifier μ under test is computed, as well as the nuisance parameters $\hat{\theta}_0^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$ maximising the likelihood (Equation 8.5) under the background-only and signal+background hypothesis, respectively. Furthermore, the p.d.f.'s of the chosen test statistic \tilde{q}_μ under the signal+background hypothesis, $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$, and the background-only hypothesis, and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$, are constructed by means of ensembles of toy MC pseudo-experiments generated according to the same Poisson probabilities used to build the likelihood. In this process the nuisance parameters are fixed to the values $\hat{\theta}_0^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$ obtained by fitting the observed data.

Using the $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$ and $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$ distributions, two p-values are computed

$$\begin{aligned} p_\mu &\equiv \text{CL}_{s+b} = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | \mu s(\hat{\theta}_\mu^{\text{obs}}) + b(\hat{\theta}_\mu^{\text{obs}})) = \int_{\tilde{q}_\mu^{\text{obs}}}^{+\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}}) d\tilde{q}_\mu \\ p_0 &\equiv \text{CL}_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | b(\hat{\theta}_0^{\text{obs}})) = \int_{\tilde{q}_\mu^{\text{obs}}}^{+\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}}) d\tilde{q}_\mu. \end{aligned} \quad (8.9)$$

The two probabilities are shown in the example in Fig. 8.20(a). In the classical frequentist approach, the level of agreement between the data and hypothesized μ is evaluated by using the CL_{s+b} probability only, and one says that the hypothesized signal μ is excluded at 95% C.L. if $\text{CL}_{s+b} \leq 0.05$.

However, such a definition as a caveat. If the distributions of the test statistic for the signal+background and background-only hypotheses have a not negligible overlap as in the plot (c) of Fig. 8.20(b), the experiment would tend to exclude the hypothesized signal μ even if the experiment in this case has little sensitivity to discriminate it against the background. In fact, in this case the experimental data are highly contaminated with background and a statement about the signal would be a mistake of interpretation. To prevent the inference of a signal in such cases, the so-called modified frequentist approach has been introduced at the time of LEP [112, 113]. In this approach, the level of agreement between the data and hypothesized μ is evaluated by using instead the quantity

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b}, \quad (8.10)$$

and the hypothesized signal μ is excluded at 95% confidence level if $\text{CL}_s \leq 0.05$. It is straightforward to see from plot (a) of Fig. 8.20(b) that, if the distribution of the test statistic for the signal+background hypothesis is well separated from the background-only distribution, then $\text{CL}_s \sim \text{CL}_{s+b}$ and there is no risk of misinterpretation.

In order to quote, as conventionally done, 95% C.L. observed upper limits, the full procedure is iterated for different values of μ , until $\text{CL}_s = 0.05$ is found. This value of μ is denoted as $\mu_{95\%}$, and one can infer that the hypothesized resonance $X \rightarrow WV/VH$ with a cross section μ -times larger than the one predicted by some specific theoretical model σ_{th} used as input to the statistical analysis, is excluded at 95% C.L.. In this analysis, model-independent limits on the cross section are set by rescaling the $\mu^{95\%} = \sigma_{95\%}/\sigma_{th}$ by the input cross section in order to obtain $\sigma_{95\%}$.

In addition to the observed upper limit derived from the actual data distribution, it is important to study also the expected limit given the observed data. In fact, the expected limit quantifies the sensitivity of the experiment independent from statistical fluctuations in the data. In order to compute the median-expected upper limit, and the associated $\pm 1\sigma$ and $\pm 2\sigma$ bands, a large set of background-only pseudo-experiments is generated and, for each of them, the $\mu_{95\%}$ is calculated. From the cumulative distribution of $\mu_{95\%}$, the median value is

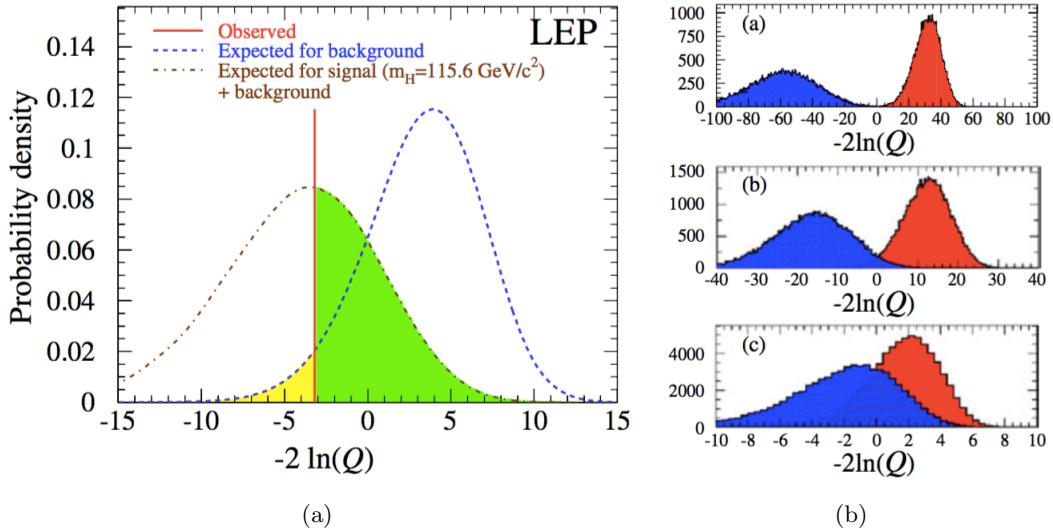


Figure 8.20: blabla

taken as the expected limit, while the $\pm 1(2)\sigma$ uncertainty bands on the expected limits are extracted from the values of the 16% (2.5%) and 84% (97.5%) quantiles.

8.6.2 The asymptotic approximation

In order to compute the CL_s the sampling distributions of the test statistics are required. In particular, one needs the p.d.f.'s $f(\tilde{q}_\mu | \mu')$, where $\mu' = 0$ or $\mu' = \mu$, which are obtained from MC toys which requiring very expensive computational resources. An approximation for the CL_s method, valid in the large sample limit, also referred to as “asymptotic approximation” has been proposed in Ref. [114] and it is briefly described in the following.

By using the Wald approximation [115] the desired distribution $f(\tilde{q}_\mu | \mu')$ can be obtained by expressing the test statistic given by the log-likelihood ratio as

$$\tilde{q}_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}), \quad (8.11)$$

where $\hat{m}\hat{\mu}$ follows a Gaussian distribution with a mean μ' and standard deviation σ , and N represents the data sample size. For large data samples ($N \rightarrow \infty$), the $\mathcal{O}(1/\sqrt{N})$ can be neglected and it can be shown [116] that the distribution $f(\tilde{q}_\mu | \mu')$ of the test statistic \tilde{q}_μ follows a *noncentral chi-square* distribution for one degree of freedom with noncentrality parameter

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}. \quad (8.12)$$

For the special case $\mu' = \mu$ one has $\Lambda = 0$ and the test statistic is distributed as a chi-square for one degree of freedom. For the general case in which $\mu' \neq \mu$, the standard deviation σ of $\hat{\mu}$ has to be evaluated, which depends on the MLE estimator of the nominal nuisance parameters. The evaluation of σ is greatly simplified considering a special, artificial data set that we call the “Asimov data set” where all statistical fluctuations are suppressed and the estimators for all parameters are replaced by their expectation values:

$$\hat{\mu} = \mu' \quad \text{and} \quad \hat{\theta} = \theta. \quad (8.13)$$

With these assumptions the test statistic $\tilde{q}_{\mu,A}$ for the Asimov dataset is given by

$$\tilde{q}_{\mu,A} \approx \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda \quad (8.14)$$

That is, from the Asimov data set one obtains an estimate of the noncentrality parameter Λ that characterizes the distribution $f(\tilde{q}_\mu|\mu')$. Equivalently, the above equation can be used to obtain the variance σ^2 which characterizes the distribution of $\hat{m}u$, namely,

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{\tilde{q}_{\mu,A}}, \quad (8.15)$$

and the distribution obtained by using σ_A^2 has a median given by the corresponding Asimov value $\tilde{q}_{\mu,A}$. Using these formulae asymptotic relations are derived which are easily solved for the observed upper limits with the CL_s method, as well as for the expected median and error bands.

8.6.3 Quantifying an excess of events

The presence of the signal is quantified by the background-only p-value, i.e. the probability for the background to fluctuate and give an excess of events as large or larger than the observed one. As for the upper limits, this evaluation requires defining a test statistic and the construction of its sampling distribution. For a given resonance mass hypothesis M_X , the test statistic used in this case is \tilde{q}_0 :

$$\tilde{q}_0 = -2 \ln \frac{\mathcal{L}(data|0, \hat{\theta}_0)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}, \quad \text{with} \quad \hat{\mu} \geq 0. \quad (8.16)$$

The sampling distribution $f(\tilde{q}_0|0, \hat{\theta}_0^{obs})$ is built by generating toy MC pseudo-data under the assumption of the background-only hypothesis. From this distribution, the p-value corresponding to a given experimental observation \tilde{q}_0^{obs} is evaluated:

$$p_0 = P(\tilde{q}_0 \geq \tilde{q}_0^{obs} | b(\hat{\theta}_0^{obs})) = \int_{\tilde{q}_0^{obs}}^{+\infty} f(\tilde{q}_0|0, \hat{\theta}_0^{obs}) d\tilde{q}_0. \quad (8.17)$$

This probability is converted into a *significance*, also referred to as *Z value*, as follows

$$Z = \Phi^{-1}(1 - p_0) \quad (8.18)$$

A significance of 5σ , corresponding to a p-value of 2.87×10^{-7} is conventionally used in high energy physics to claim a discovery, and 3σ for an evidence.

It can be demonstrated that in the asymptotic approximation (Section 8.6.2), the likelihood ratio test statistic \tilde{q}_0 follows a chi-square distribution for one degree of freedom, and a fair estimate of the p-value and of the significance can be obtained from the observed value \tilde{q}_0^{obs} itself, without the need for generating pseudo-data

$$p_0 = \frac{1}{2} [1 - \text{erf}(\sqrt{\tilde{q}_0^{obs}/2})] \quad (8.19)$$

$$Z = \sqrt{\tilde{q}_0^{obs}}$$

The p-value discussed above is evaluated at a fixed resonance mass M_X and can be referred to as a *local p-value*. In this search, a scan is performed over a wide range of resonance mass hypotheses with the aim of finding the minimum local p-value, which describes the probability of a background fluctuation for that particular resonance mass hypothesis. However, it is important to distinguish the probability of finding a fluctuation in some particular location from the probability of finding such a fluctuation anywhere else. The former is associated to the so called *local significance*, whereas the latter is referred to as the *global significance*. The fact that the global significance is usually smaller than the largest local one is often referred to as the “look-elsewhere effect” (LEE). As demonstrated in Ref. [117], the global and local p-values are related to each other by a multiplicative factor, usually referred to as “trial factor”, proportional to the number of independent search regions. In the asymptotic approximation the trial factor grows linearly with the local significance, through a proportional constant that is related to the ratio between the mass range under consideration divided by its resolution. In particular, it can be shown that

$$\text{trial\#} = \frac{p_{\text{global}}}{p_{\text{local}}} \approx \frac{1}{3 \text{ mass resolution}} Z_{\text{local}} \quad (8.20)$$

The trial factor is best estimated through MC methods as it will be shown in Section 9.2. However, a good agreement with the equation above is obtained.

Results with 8 TeV data

The final results of the analysis performed with 8 TeV data and focused on the search for a heavy charged resonance decaying into W and Higgs bosons in the $\ell\nu b\bar{b}$ final state, are presented and discussed in this Chapter. In particular, the final observed m_{WH} spectrum is used to check for the presence of a new resonance. Firstly, a search is conducted for local enhancement in the m_{WH} distribution, which might be due to a signal. As described in the following, since no significant excesses are found, upper limits are set on the production cross section of the new resonance.

9.1 Final m_{WH} distribution

The predicted number of background events in the signal region after the inclusion of all backgrounds is summarized in Table 9.1 and compared with observations. The yields are quoted in the range $0.7 < m_{\text{WH}} < 3\text{TeV}$. The expected background is derived with the sideband procedure described in Section 8.3. The uncertainties in the background prediction from data are statistical in nature, as they depend on the number of events in the sideband region. The muon channel has more expected background events than the electron channel owing to the lower $E_{\text{T}}^{\text{miss}}$ requirement and its worse mass resolution at high p_{T} .

Table 9.1: Observed and expected yields in the signal region together with statistical uncertainties.

	$e\nu + \text{H-jet}$	$\mu\nu + \text{H-jet}$
Observed yield	9	16
Expected total background	11.3 ± 3.1	14.9 ± 3.1
W+jets	4.7 ± 2.9	7.0 ± 3.1
Top	6.3 ± 1.1	7.3 ± 0.4
VV	0.4 ± 0.1	0.6 ± 0.2

Figure 9.1 shows the final observed m_{WH} spectra after all selection criteria have been applied. The highest mass event is in the electron category and has $m_{\text{WH}} \approx 1.9\text{TeV}$. The observed data and the predicted background in the muon channel agree. In the electron channel, an excess of three events is observed with $m_{\text{WH}} > 1.8\text{TeV}$, where about 0.3 events are expected, while in the muon channel no events with $m_{\text{WH}} > 1.8\text{TeV}$ are observed, where about 0.3 events are expected.

9.2 Significance of the data

A comparison between the m_{WH} distribution observed in data and the largely data-driven background prediction is used to test for the presence of a resonance decaying into WH. As described in Section 8.6, the statistical test is performed based on a profile likelihood discriminant for an unbinned shape analysis. Systematic uncertainties in the signal and background yields are treated as nuisance parameters and profiled in the statistical interpretation

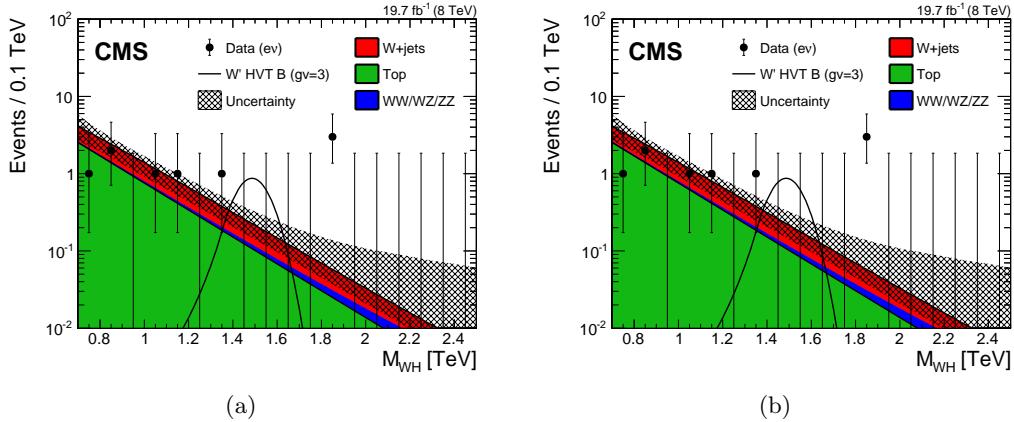


Figure 9.1: Final distributions in m_{WH} for data and expected backgrounds for electron (a) and muon (b) categories. The 68% error bars for Poisson event counts are obtained from the Neyman construction [118]. The hatched region indicates the statistical uncertainty of the fit combined with the systematical uncertainty in the shape. This figure also shows a hypothetical 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.

using log-normal priors, while Gaussian priors are used for shape parameters only. The local significance of the observations is evaluated in the context of the described statistical test, under the assumptions of a narrow resonance decaying into the WH final state and lepton universality for the W boson decay, by combining the two event categories. Correlations arising from the uncertainties common to both channels are taken into account. The result is shown in Fig. 9.2. The highest local significance of 2.2 standard deviations is found for a resonance mass of 1.8 TeV, driven by the excess in the electron channel described in the previous section. The corresponding local significance for a resonance of 1.8 TeV in the electron channel is 2.9 standard deviations, while in the muon channel there is no significance.

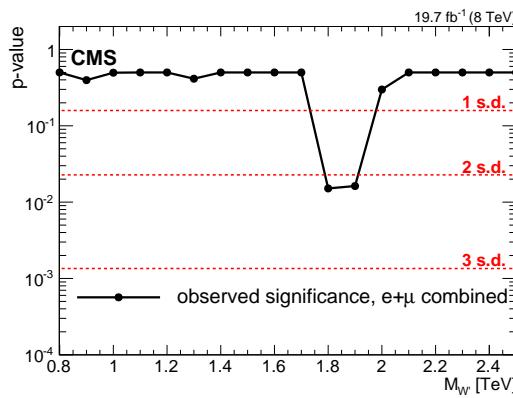


Figure 9.2: Local p-value of the combined electron and muon data as a function of the W' boson mass, probing a narrow WH resonance.

Taking into account the look-elsewhere effect (Section 8.6.3), the local significance of 2.9 standard deviations can be translated into a global significance value by computing the trial factor as given in Equation 8.20. Considering the mass range 0.8–2.5 TeV and the step of 0.1 TeV used for the search, a trial factor of ≈ 16.4 is obtained. The factor, when multiplied by the local p-value, gives a global significance of 1.9 standard deviations when searching

for resonances over the full mass range and across two channels. In order to cross check this final value, the LEE is also estimated by means of background-only pseudo-experiments. The relation between the global and local significances obtained with this method is shown in Fig. 9.3, and it agrees with the calculation performed with the trial factor. It can be concluded that the results are thus statistically compatible with the SM expectation within 2 standard deviations.

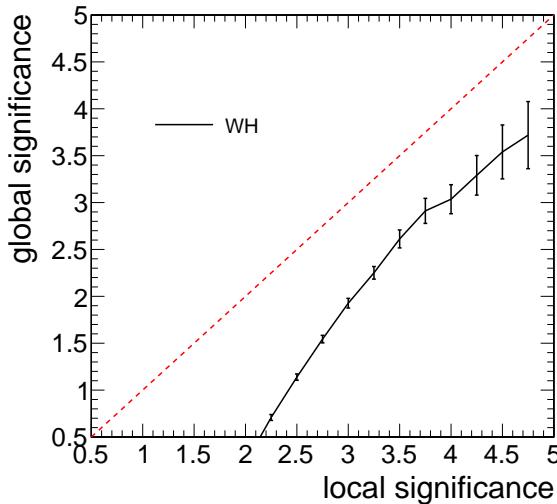


Figure 9.3: Global significance as a function of the local significance which corresponds to the maximal significance in the m_{WH} range 0.8–2.5 TeV in the two categories. The global significance is estimated with a frequentist approach using background-only pseudo-experiments and corresponds to the fraction of toys (translated from a p-value to significance) with at least a certain local significance in the m_{WH} range in the two categories.

9.3 Cross section limits

FIXME: point to theory with widths, xsec, model parameters etc. Since no excesses with significance larger than three standard deviations are observed, upper limits are set on the production cross section of the new resonance following the modified-frequentist CL_s method described in Section 8.6. Exclusion limits can be set as a function of the W' resonance mass, under the narrow-width approximation. The results are interpreted in the HVT model B which mimics the properties of composite Higgs scenarios, and in the context of the little Higgs model.

Figure 9.4 shows the expected and observed exclusion limits at 95% CL on the product of the W' production cross section and the branching fraction of $W' \rightarrow \text{WH}$ for the electron and muon channels separately, and for the combination of the two. For the combined channels, the observed and expected lower limits on the W' mass are 1.4 TeV in the LH model and 1.5 TeV in the HVT model B. For the electron (muon) channel, the observed and expected lower limits on the W' mass are 1.2 (1.3) TeV in the LH model and 1.3 (1.3) TeV in the HVT model B.

These results are finally combined with other searches for heavy resonances decaying into diboson with 8 and 13 TeV data as described in Chapter 11.

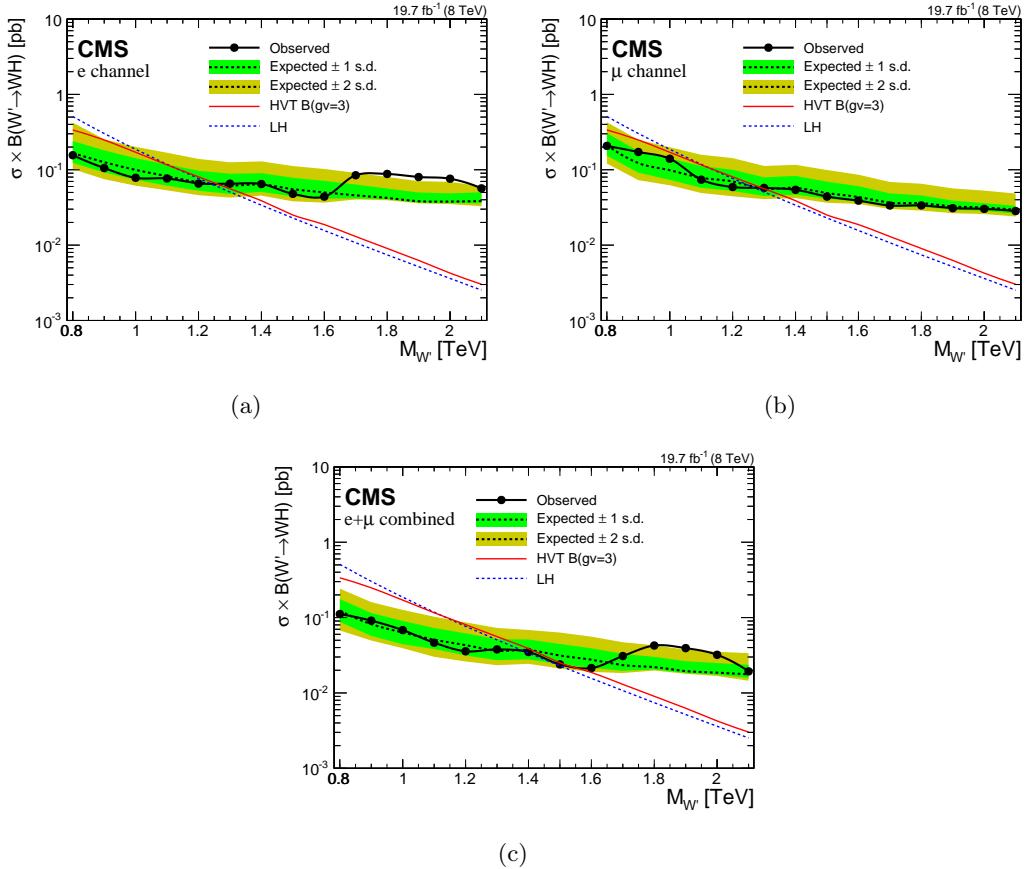


Figure 9.4: Observed (solid) and expected (dashed) upper limits at 95% CL on the product of the W' production cross section and the branching fraction of $W' \rightarrow WH$ for electron (a) and muon (b) channels, and the combination of the two channels (c). The products of cross sections and branching fractions for W' production in the LH and HVT models are overlaid.

Results with 13 TeV data

In this chapter, the final results of the analysis performed with 13 TeV data and focused on the search for a heavy resonances decaying into a pair of vector bosons (WW/WZ) in the $\ell\nu q\bar{q}^{(\prime)}$ final state, are presented and discussed. As for the analysis conducted with 8 TeV data described in the previous chapter, the final m_{WV} spectrum observed in data is used to check for the presence of a new resonance. No significant excesses are observed and upper limits are set on the production cross section of such resonances under a variety signal hypotheses by combining all the event categories.

10.1 Final m_{WV} distribution

The final m_{WV} spectra observed in data and for the background predicted with the α ratio method (Section 8.3) for all event categories are shown in Fig. 10.1. The observed data and the predicted background are found to well agree. The highest mass events are at $m_{WV} = 2.95$ and 3.15 TeV for the muon and electron category, respectively.

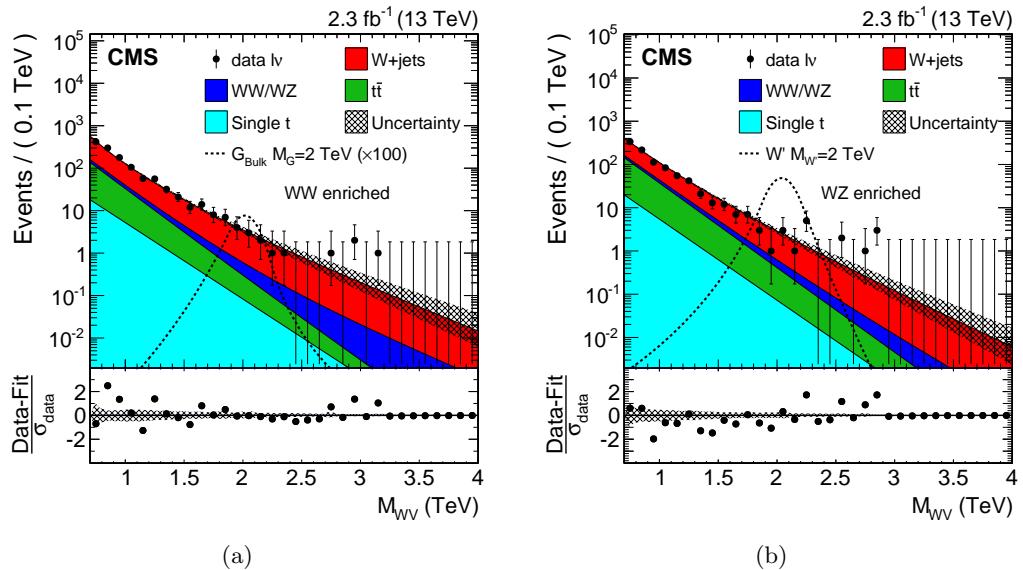


Figure 10.1: Final m_{WV} distributions for data and expected backgrounds obtained combining muon and electron channels in the WW-enriched (a) and WZ-enriched (b) signal regions. In each plot the solid curve represents the background estimation provided by the α ratio method. The hatched band includes both statistical and systematic uncertainties. The data are shown as black points. At the bottom of each plot are the bin-by-bin fit residuals, $(N_{\text{data}} - N_{\text{fit}})/\sigma_{\text{data}}$, shown together with the uncertainty band of the fit normalized by the statistical uncertainty of data, σ_{data} . Several signal benchmarks are also shown with black dashed lines.

10.2 Cross section limits

Since no excesses with significance larger than three standard deviations are observed, upper limits are set on the production cross section of the new resonance by combining all event categories. The asymptotic approximation of the CL_s criterion described in Section 8.6 is followed. The exclusion limits computed with this approach are found to agree with the results obtained using the modified frequentist prescription. Systematic uncertainties are treated as nuisance parameters in the statistical interpretation using log-normal, and they are profiled following the frequentist convention as discussed in Section 8.6.

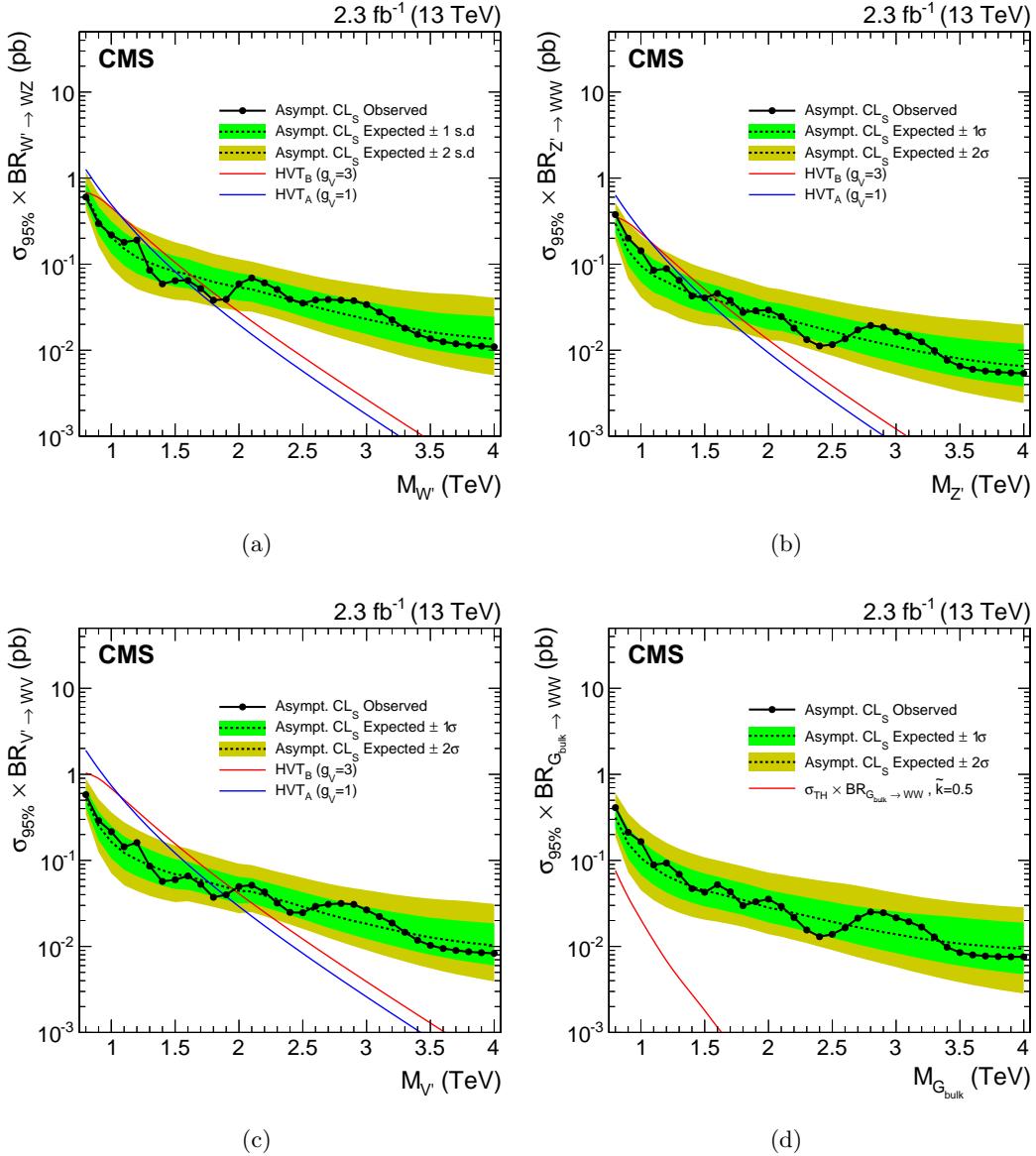


Figure 10.2: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. In the upper plots, limits are set in the context of a spin-1 charged W' (a) and neutral Z' (b) resonances, and compared with the prediction of the HVT Models A and B. (c) Limits are set in the same model under the triplet hypothesis (W' and Z'). (d) Limits are set in the context of a bulk graviton with $k/\bar{M}_{\text{Pl}} = 0.5$ and compared with the prediction.

Exclusion limits are set in the context of the bulk graviton model and of the HVT Models A and B, under the assumption of a natural width negligible compared to the experimental resolution. Figure 10.2 shows the resulting 95% CL expected and observed exclusion limits on the signal cross section as a function of the resonance mass for all signal hypotheses. The limits are compared with the product of cross section and branching fraction ($\sigma \times \mathcal{B}$) to WW for a bulk graviton with $k/\bar{M}_{\text{Pl}} = 0.5$, and with $\sigma \times \mathcal{B}$ for WZ and WW for spin-1 particles predicted by the HVT Models A and B. In this context, a scenario is considered, where the W' and Z' bosons are expected to be degenerate in mass (triplet hypothesis). In addition, the statistical interpretation is provided in a scenario where only a charged (W') or a neutral (Z') resonance is expected at a given mass (singlet hypothesis).

In the narrow-width bulk graviton model, the sensitivity of the search is not large enough to set mass limits, however, cross sections are excluded in the range 0.007–0.4 pb. For HVT Models A and B, the data exclude singlet W' , and Z' resonances with masses below 1.6 and < 1.5 , and < 1.9 and < 1.6 TeV, respectively. Under the triplet hypothesis, spin-1 resonances with masses below 1.9 and < 2 TeV are excluded for HVT Models A and B, respectively.

These results supersede the ones obtained analyzing 8 TeV data, where the lower mass limit of 1.5 TeV for a W' in the context of the HVT model B is reached (Fig. 9.4). However, the most stringent limits are obtained in the final combination of these results with other searches for heavy resonances decaying into diboson with 8 and 13 TeV data, as described in Chapter 11.

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

In addition to the analyses described in this work, several similar searches for heavy resonances decaying to pairs of W, Z, and Higgs bosons in various final states have been performed with the CMS experiments in both LHC Run 1 and Run 2 [91–93, 97, 119–122]. As these searches have individually very similar sensitivity to benchmark physics scenarios of interest, a statistical combination to maximize the overall sensitivity is performed and presented in this Chapter. Furthermore, the combination of these analyses is fundamental to fully understand the compatibility of the excess observed in the $\ell\nu b\bar{b}$ final state at $m_{WH} = 1.8$ TeV as discussed in Chapter 9. The interest in this excess has been further enhanced by the observation of an excess at the same diboson invariant mass values by ATLAS in the all-hadronic VV final state [123].

The analyses taken into account in the statistical combination are based on pp collision data collected by the CMS experiment during 2012 and 2015 at $\sqrt{s} = 8$ TeV and 13 TeV, corresponding to an integrated luminosity of 19.7 fb^{-1} and $2.3\text{--}2.7 \text{ fb}^{-1}$, respectively. Analyses with all-leptonic, semi-leptonic, and all-jets final states are considered. This includes the decay into charged leptons (ℓ) and neutrinos (ν) of W and Z bosons, as well as reconstructed jets containing the decay products of hadronically decaying W or Z bosons. The latter are labeled as q \bar{q} final states that include $W \rightarrow q\bar{q}' \rightarrow \text{jet}$ and $Z \rightarrow q\bar{q} \rightarrow \text{jet}$. For Higgs bosons, hadronic decays labeled as b \bar{b} or q $\bar{q}q\bar{q}$ final states referring to $H \rightarrow b\bar{b}$ or $H \rightarrow q\bar{q}'q\bar{q}'$ are considered.

Altogether, results are combined corresponding to the following final states: $\ell\nu q\bar{q}$ (13 TeV) [93], q $\bar{q}q\bar{q}$ (13 TeV) [93], $\ell\ell b\bar{b}/\ell\nu b\bar{b}/\nu\nu b\bar{b}$ (13 TeV) [97], $3\ell\nu$ (8 TeV) [120], $\ell\nu q\bar{q}$ (8 TeV) [91], $\ell\ell q\bar{q}$ (8 TeV) [91], q $\bar{q}q\bar{q}$ (8 TeV) [92], $\ell\nu b\bar{b}$ (8 TeV) [121], q $\bar{q}b\bar{b}/6q$ (8 TeV) [119], q $\bar{q}\tau\tau$ (8 TeV) [122].

The results are interpreted in the context of the BSM models described in Section 2.3, namely, heavy vector triplet and singlet models predicting W' and Z' bosons, and the bulk graviton model. Combined cross section limits as a function of resonance mass are obtained. This is the first combined search for high mass resonances with both WW/WZ and WH/ZH signatures.

A summary of the analyses entering the combination is given in Section 11.1. The combination procedure is described in Section 11.2, and finally the results are presented and discussed in Section 11.3.

11.1 Inputs to the combination

A statistical combination is carried out of searches for new heavy resonances that are performed on top of the steeply falling invariant mass distribution of two reconstructed W, Z or Higgs bosons. Various decay modes of these bosons are considered. The Z $\rightarrow \ell\ell$ candidates are

reconstructed from electron and muon candidates, while $W \rightarrow \ell\nu$ candidates are reconstructed from identified muons or electrons with the method described in Section 6.6, which makes use of the missing transverse momentum under the constraint that the $\ell\nu$ invariant mass is equal to the known W -boson mass. The $H \rightarrow \tau\tau$ candidates are reconstructed from electron, muon and hadronically-decaying tau candidates in combination with missing transverse momentum. The $W \rightarrow q\bar{q}'$, $Z \rightarrow q\bar{q}$, $H \rightarrow b\bar{b}$ and $H \rightarrow q\bar{q}'q\bar{q}'$ candidates are reconstructed with jet algorithms with a distance parameter of 0.8 (CA for the 8 TeV data analyses, AK for the 13 TeV analyses).

All analyses are focused on high mass resonances which decay in highly boosted $W/Z/H$ bosons. Hence, their decay products are reconstructed close-by in angle, requiring special reconstruction techniques. For highly boosted $W/Z/H$ bosons decaying to electron, muon and tau candidates, identification and isolation requirements are adapted such that the nearby reconstructed leptons do not reduce the identification efficiency.

For highly boosted V bosons decaying to quark anti-quark pairs, the V algorithm described in Section 7.2 is applied. In the 8 TeV data analyses, a V jet candidate is identified if its pruned mass, m_{jet} , falls in a range around the W or Z mass. In the 13 TeV data analyses, two distinct categories enriched in W or Z bosons are defined by two exclusive ranges in m_{jet} as described in Section 8.1. In the 8 TeV data analyses the sensitivity is further enhanced by distinguishing two categories, a low purity (LP) and a high purity (HP) one based on the τ_{21} variable. This same strategy is followed in the dijet 13 TeV analysis. Although the HP category dominates the total sensitivity of the analyses, the LP category is retained, since for large masses of a new resonance it provides improved signal efficiency with only moderate background contamination.

Higgs-boson identification is similarly performed using a pruned jet mass window around the Higgs mass together with b-tagging algorithms applied to the H jet or to its subjets as described in Section 7.3. To distinguish $H \rightarrow WW \rightarrow q\bar{q}'q\bar{q}'$ jets from background, a similar technique as V tagging is applied using the τ_{42} ratio. The selection efficiencies for each signal and channel are summarized in Table 11.1.

Table 11.1: Summary of the signal efficiencies of all analysis channels for all signal models for a 2 TeV resonance. For analyses with categorization in high-purity (HP) and low-purity (LP) categories, both efficiencies are quoted in the form HP/LP. The signal efficiencies are in percent and include the SM branching ratios of the bosons to the final state of the analysis channel, effects from detector acceptance, as well as reconstruction and selection efficiencies.

Channel	HVT				RS bulk	
	W'		Z'		G_{bulk}	
	WZ	WH	WW	ZH	WW	ZZ
$3\ell\nu$ (8 TeV)	0.6	-	-	-	-	-
$\ell\ell q\bar{q}$ (8 TeV)	1.1/-	-	-	0.2/-	-	3.0/1.0
$\ell\nu q\bar{q}$ (8 TeV)	4.8/-	-	9.4/-	-	10.6/7.1	-
$q\bar{q}q\bar{q}$ (8 TeV)	5.9/5.5	0.8/0.7	5.7/5.3	0.8/0.7	3.8/3.1	5.7/4.2
$\ell\nu b\bar{b}$ (8 TeV)	-	0.9	-	-	-	-
$q\bar{q}\tau\tau$ (8 TeV)	-	1.2	-	1.3	-	-
$q\bar{q}b\bar{b}/6q$ (8 TeV)	-	3.0/1.8	-	1.7/1.1	-	-
$\ell\nu q\bar{q}$ (13 TeV)	10.2/-	1.7/-	19.4/-	-	18.1/-	-
$q\bar{q}q\bar{q}$ (13 TeV)	9.7/12.3	1.8/2.5	8.2/10.6	1.9/2.6	8.7/12.4	11.0/13.5
$\ell\ell b\bar{b}$ (13 TeV)	-	-	-	1.5	-	-
$\ell\nu b\bar{b}$ (13 TeV)	-	4.0	-	-	-	-
$\nu\nu b\bar{b}$ (13 TeV)	-	-	-	4.2	-	-

In all-jets final states, the background dominated by QCD multijets production is estimated with a fit of signal+background to the data, where the background is described by a smooth functional form. In semi-leptonic final states, the dominant backgrounds from V+jets production are estimated using data in m_{jet} sidebands with the method described in Section 8.3. In all-leptonic final states, the dominant background from standard model diboson production is estimated using simulated events.

More details are given in the following for the analyses where not all signal models presented in the combination were originally considered.

11.1.1 Reinterpretations

In the searches for new heavy resonances decaying into a pair of vector bosons (WW, ZZ or WZ) in the semi-leptonic ($\ell\nu q\bar{q}$ and $\ell\ell q\bar{q}$) final states [91] with pp collision data collected at $\sqrt{s} = 8$ TeV, exclusion limits at 95% CL have been set on the production cross section of a bulk graviton. The results were published with a parametrization for the reconstruction efficiency as a function of W and Z boson kinematics, enabling a reinterpretation in the context of the HVT model B described in Section 2.3.3, which predicts the production of charged and neutral spin-1 resonances decaying preferably to WW and WZ. The reinterpretation in the context of this model is obtained by rescaling the bulk graviton signal efficiencies by scale factors taking into account the different kinematics of W and the Z bosons from W' and Z' production compared to the graviton production. The scale factors have been derived for each mass point by means of the tables published in Ref. [91]. Since the efficiency parametrization is restricted to the HP category of the analyses, the LP category is not used for the HVT W' and Z' interpretations of these channels. The m_{jet} window that defines the signal regions of the analysis channels is such that the $\ell\nu q\bar{q}$ channel is sensitive to both the charged and neutral resonance predicted by HVT models. This is taken into account in the combination presented in Section 11.3.2.

The searches for new heavy resonances decaying into a pair of vector bosons (WW, ZZ or WZ) in the semi-leptonic ($\ell\nu q\bar{q}$ and $\ell\ell q\bar{q}$) [91, 91, 93], and all-hadronic ($q\bar{q}q\bar{q}$) final states [92, 93] at 8 and 13 TeV, are also sensitive to WH and ZH signatures, since a small fraction of jets initiated by Higgs bosons have a pruned jet mass in the range considered to identify W or Z bosons. These searches were therefore re-interpreted with WH and ZH signals to profit from this additional signal sensitivity. The additional signal efficiencies for those signals are indicated in Table 11.1.

The search for resonances in the $q\bar{q}\tau\tau$ final state [119] was optimized for a resonance Z' decaying into Z and a Higgs boson. However, given the large m_{jet} window ($65 \text{ GeV} < m_{\text{jet}} < 105 \text{ GeV}$) used to tag the hadronically decaying Z boson, this analysis channel is also sensitive to the production of the charged spin-1 W' resonance decaying into W and Higgs bosons as predicted in HVT models. This overlap is taken into account in the statistical combination described in Section 11.3.2.

11.2 Combination procedure

In all the analysis channels a search is performed for a peak on top of the falling background distribution in the diboson invariant mass by means of a maximum likelihood fit to the data. As done for the main analyses described in this work (Section 8.6), the likelihood function is maximized to obtain the best fit of the signal strength modified μ for each signal and resonance mass hypothesis. The function is constructed from the reconstructed diboson invariant mass distribution observed in data, the background prediction, and the signal

resonance shape to test for the presence of a new resonance decaying to two bosons. For the $3\ell\nu$, $q\bar{q}q\bar{q}$, $q\bar{q}b\bar{b}/6q$, and $q\bar{q}\tau\tau$ analyses, the likelihood function is computed using events binned as a function of reconstructed diboson invariant mass as in Equation 8.6. For the remaining analyses ($\ell\nu q\bar{q}$, $\ell\ell q\bar{q}$, $\ell\nu b\bar{b}$), the functional form for an unbinned likelihood is similarly defined using functional forms that describe the shape of the reconstructed diboson invariant mass for background and signal resonance as given by Equation 8.7.

The treatment of the background in the maximum likelihood fit depends on the analysis channel. In the $q\bar{q}q\bar{q}$ and $q\bar{q}b\bar{b}/6q$ analyses, the background fit function parameters are left floating in the maximum likelihood fit, such that the background prediction is simultaneously obtained with the signal μ for every hypothesis. The remaining analyses ($\ell\nu q\bar{q}$, $\ell\ell q\bar{q}$, $\ell\ell b\bar{b}$, $\ell\nu b\bar{b}$, $\nu\nu b\bar{b}$) follow the same procedure as for the analyses described in this work: the background is estimated using data sidebands and uncertainties related to its parametrized shape are treated as nuisance parameters constrained with Gaussian probability density functions in the maximum likelihood fit. Except for the cases described in Section 11.1, which have been found to be negligible, selection are exclusive. The combined likelihood is then obtained from the product of the likelihoods of each individual analysis channel.

The asymptotic approximation of the CL_s criterion (Section 8.6.2) is used with the test statistic given by Equation 8.8 to set upper limits on the cross section for resonance production. When combining 8 and 13 TeV analyses, limits are set on the signal scale factor μ taking into account the production cross section ratio evaluated from theory between 8 TeV and 13 TeV.

The dominant sources of systematic uncertainties are treated as nuisance parameters constrained with a log-normal probability density function. All nuisance parameters are profiled following the frequentist convention discussed in Section 8.6. When the likelihoods of multiple analyses channels are combined, the correlation of systematic effects across analysis channels is taken into account by categorizing the uncertainties into fully correlated (associate to same nuisance parameter) and fully uncorrelated (associate to different nuisance parameters). Table 11.2 summarizes which uncertainties are treated as correlated among 8 and 13 TeV analyses, electron and muon channels, HP and LP categories and W, Z and Higgs enriched categories in the combination. Further categorisation within individual analyses are described therein.

Table 11.2: Correlation of systematic uncertainties.

Systematic uncertainty	Type	8+13 TeV	e+ μ	HP+LP	W+Z+Higgs
Lepton trigger	yield	no	no	yes	yes
Lepton identification	yield	no	no	yes	yes
Lepton momentum scale	yield, shape	no	no	yes	yes
Jet energy scale	yield, shape	no	yes	yes	yes
Jet energy resolution	yield, shape	no	yes	yes	yes
Jet mass scale	yield	no	yes	yes	yes
Jet mass resolution	yield	no	yes	yes	yes
b tagging	yield	no	yes	yes	yes
W tagging τ_{21} (HP/LP)	yield	no	yes	yes	yes
Integrated luminosity	yield	no	yes	yes	yes
Pileup	yield	no	yes	yes	yes
PDF	yield	yes	yes	yes	yes
μ_f, μ_r scales	yield	yes	yes	yes	yes

The most important and only nuisance parameters treated as correlated between 8 and 13 TeV analyses are those related to the PDFs and the choice of factorization (μ_f) and

renormalization (μ_r) scales used to estimate the signal cross sections. They have been re-evaluated for this combination for both 8 and 13 TeV analyses, estimating the full impact on the expected signal yield rather than the impact on only the signal acceptance. The PDF uncertainties are evaluated using the NNPDF 3.0 [21] PDFs. The uncertainty related to the choice of μ_f and μ_r scales is evaluated following the proposal in Refs. [106, 107] by varying the default choice of scales in the following 6 combinations of factors: $(\mu_f, \mu_r) \times (1/2, 1/2)$, $(1/2, 1)$, $(1, 1/2)$, $(2, 2)$, $(2, 1)$, and $(1, 2)$. The experimental uncertainties are all treated as uncorrelated between 8 and 13 TeV. At 13 TeV the systematic uncertainties are dominated by the statistical uncertainty of the datasets used to evaluate scale factors applied to the signal simulation to reproduce data.

11.3 Results

In this section the combination of the individual analysis channels described in Section 11.1 is presented, for each of the signal hypothesis described in Section 2.3. For each channel the 95% CL exclusion limits on the signal strength modifier $\mu = \sigma_{95\%}/\sigma_{\text{theory}}$ are presented.

11.3.1 Limits on W' and Z' singlets

Figures 11.1(a) and 11.1(b) show the comparison and combination of the results obtained in the 8 and 13 TeV searches for a W' singlet resonance for model A and model B, respectively. The 95% CL exclusion limits on the signal strength in the resonance mass range $0.6 < m_{W'} < 4$ TeV for model A and $0.8 < m_{W'} < 4$ TeV for model B are shown. Table 11.3 summarizes the resulting resonance mass exclusion limits. Below resonance mass values of about 1.4 TeV, the $3\ell\nu$ channel at 8 TeV is most sensitive. At higher masses, the $q\bar{q}q\bar{q}$ search at 13 TeV dominates the sensitivity. The overall sensitivity benefits from the combination up to resonance masses of about 2 TeV, lowering the cross section exclusion limit by up to a factor 1/3 when comparing to the most sensitive single channel. Above masses of 2 TeV the 8 TeV channels do not add any significant contribution compared to the $q\bar{q}q\bar{q}$ search at 13 TeV. The observed mass limit is not affected by the combination compared to that obtained from the 13 TeV searches. However, the expected mass limit is slightly improved from 2.3 to 2.4 TeV.

Table 11.3: Resonance mass 95% CL exclusion limits in HVT model scenarios.

Model	Observed limit (TeV)	Expected limit (TeV)
W' (model A)	2.3	2.1
Z' (model A)	2.2	2.0
HVT ($W' + Z'$) (model A)	2.4	2.4
W' (model B)	2.3	2.4
Z' (model B)	2.3	2.1
HVT ($W' + Z'$) (model B)	2.4	2.6

Figures 11.1(c) and 11.1(d) show the comparison and combination of the results obtained in the 8 and 13 TeV searches for a Z' singlet resonance for model A and model B, respectively. The $\ell\nu q\bar{q}$ channel at 8 TeV and the $q\bar{q}q\bar{q}$, $\ell\nu q\bar{q}$, $\ell\ell b\bar{b}/\nu\nu b\bar{b}$ channels at 13 TeV dominate the sensitivity over the whole range, with 8 and 13 TeV analyses giving almost equal contributions for masses below 2 TeV. Above this value, the sensitivity is mainly driven by the 13 TeV analyses. Under this signal hypothesis the sensitivities reached by the 8 and 13 TeV channels are similar at low resonance masses. As for the W' case, the mass limit is not affected by the combination compared to what is obtained from the 13 TeV searches.

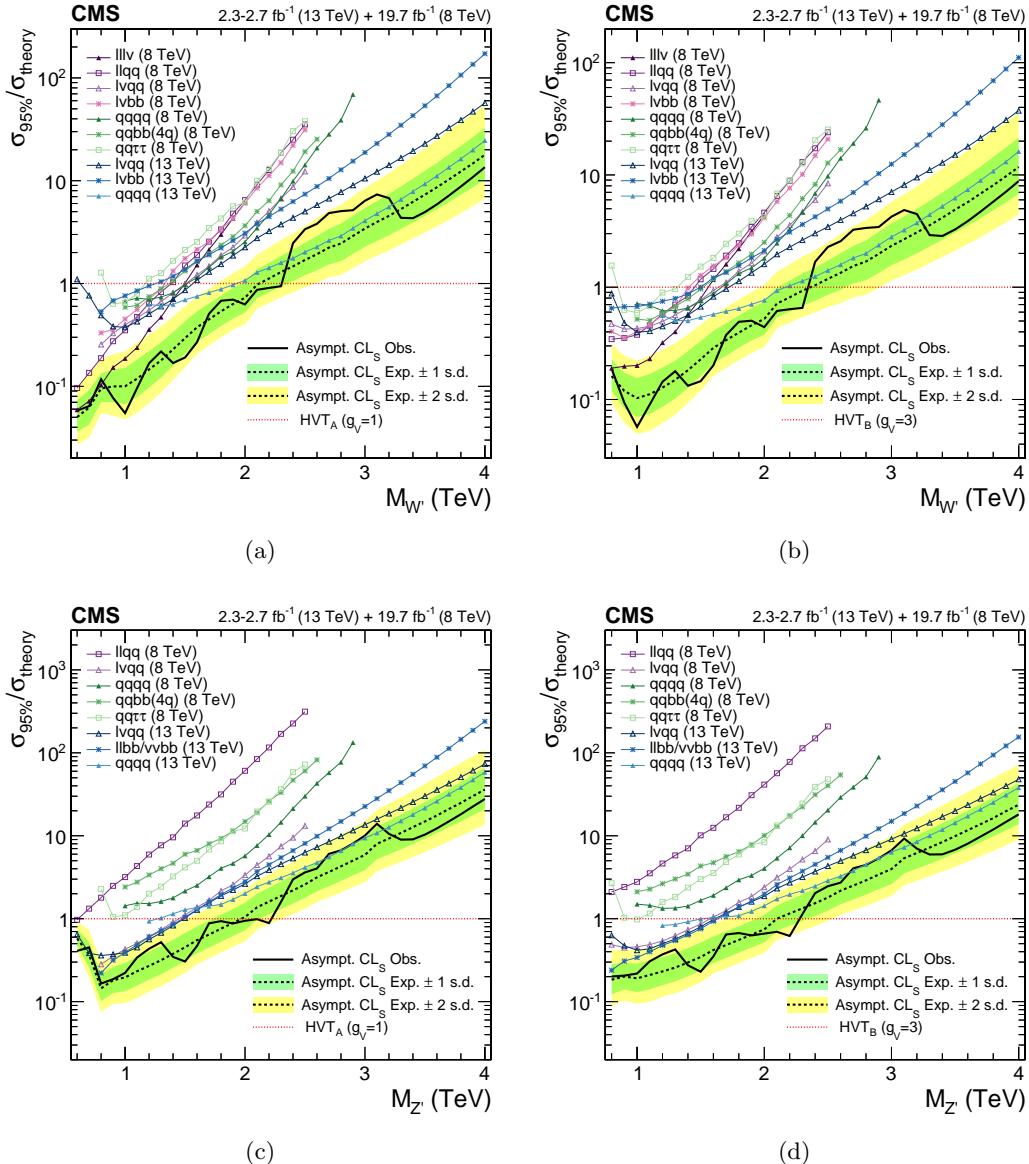


Figure 11.1: Exclusion limits at 95% CL on the signal strength for (top) $W' \rightarrow WZ/WH$ and (bottom) $Z' \rightarrow WW/ZH$ in (left) HVT model A and (right) HVT model B as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In each of the plots the different colored lines correspond to the searches entering the combination.

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

Figures 11.2(a) and 11.2(b) shows the comparison and combination of the results obtained in the 8 and 13 TeV searches for a heavy vector triplet scenario. As for the W' and Z' cases, the observed mass limit of 2.4 TeV obtained combining 8 and 13 TeV searches is determined by the 13 TeV channels.

In Fig. 11.2(c), a scan of the coupling parameters and the corresponding observed 95% CL exclusion contours in the HVT model from the combination of the 8 and 13 TeV analyses are shown. The parameters are defined as g_{VCH} and $g^2 c_F/g_V$, in terms of the coupling strengths (Section 2.3.3) of the new resonance to the Higgs boson and to fermions. The range of the scan is limited by the assumption that the new resonance is narrow. A contour is overlaid,

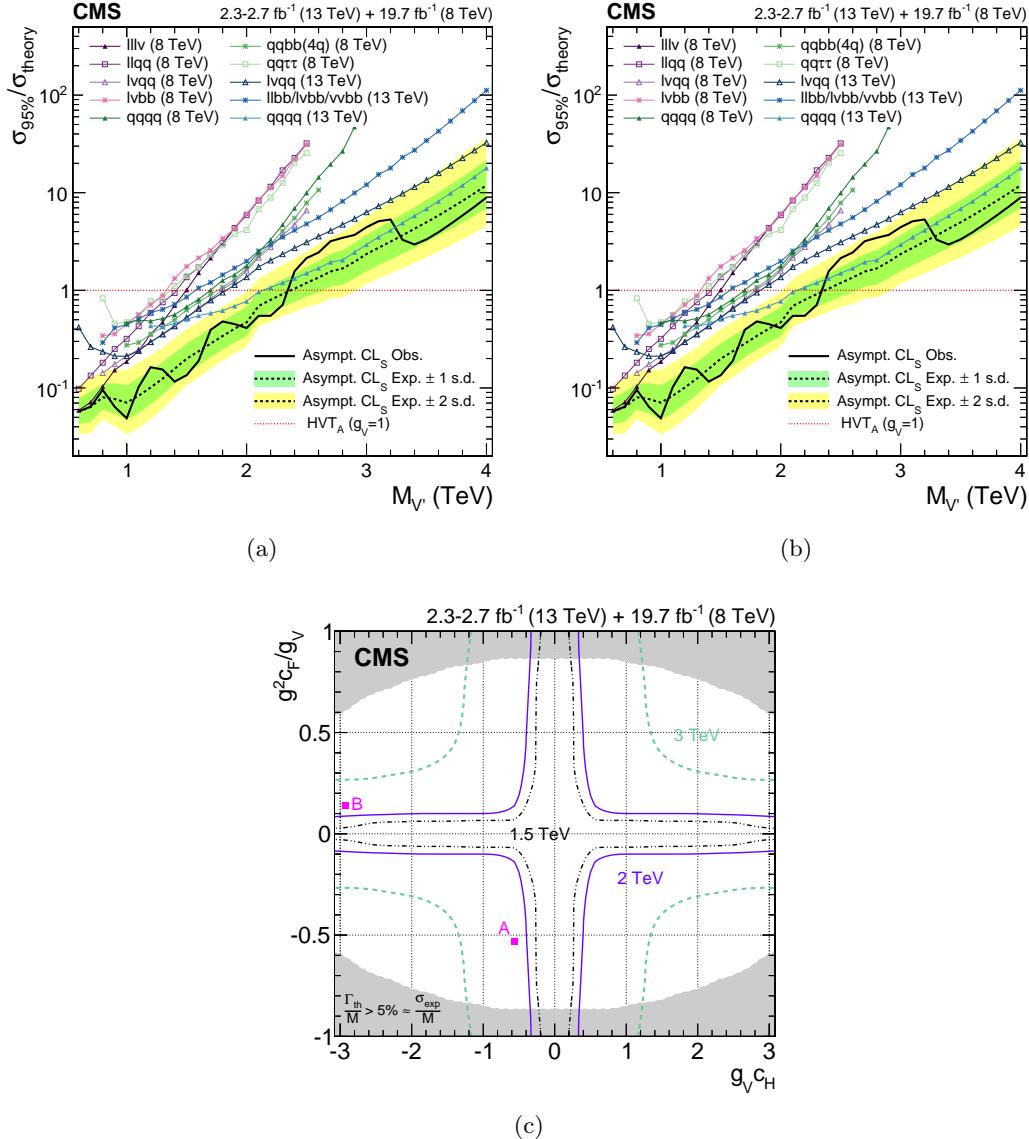


Figure 11.2: Exclusion limits at 95% CL on the signal strength in (a) HVT model A and (b) HVT model B as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In both plots the different colored lines correspond to the searches entering the combination. (c) Exclusion regions in the plane of the HVT-model couplings ($g_V c_H$, $g^2 c_F/g_V$) for three resonance masses, 1.5, 2, and 3 TeV, where g denotes the weak gauge coupling. The points A and B of the benchmark models used in the analysis are also shown. The boundaries of the regions outside these lines that are excluded by this search are indicated by the solid and dashed lines. The areas indicated by the solid shading correspond to regions where the resonance width is predicted to be more than 7% of the resonance mass and the narrow-resonance assumption is not satisfied.

representing the region where the theoretical width is larger than the experimental resolution of the searches, and hence where the narrow-resonance assumption is not satisfied. This contour is defined by a predicted resonance width of 5%, corresponding to the narrowest resonance mass resolution of the considered searches.

11.3.3 Limits on bulk graviton

Figure 11.3 shows the comparison and combination of the results obtained in the 8 and 13 TeV VV searches in the bulk graviton scenario with $k/\bar{M}_{Pl} = 0.5$. The sensitivity is mainly driven by the 13 TeV $q\bar{q}q\bar{q}$ and $\ell\nu q\bar{q}$ channels. Under this signal hypothesis, the sensitivity reached by the 13 TeV searches supersedes the 8 TeV combination down to very low resonance masses (0.7 TeV), since this signal is produced via gluon-fusion in contrast to the HVT resonances produced via $q\bar{q}$ annihilation. Hence, the contribution given by 8 TeV channels is less significant with respect to the spin-1 resonance hypotheses. The combination yields the most stringent signal strength limits on narrow bulk graviton resonances ($k/\bar{M}_{Pl} = 0.5$) to date in the mass range from 0.6 to 4 TeV.

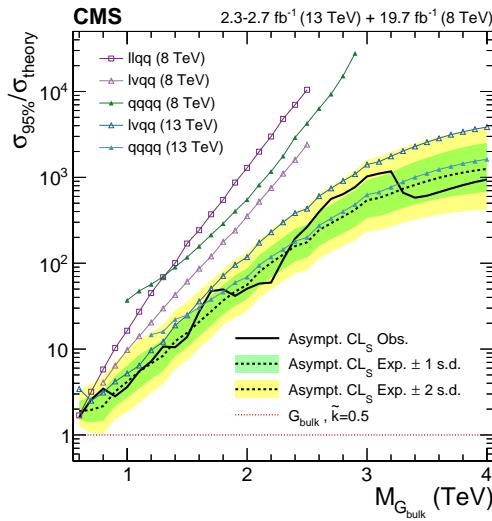


Figure 11.3: Exclusion limits at 95% CL on the signal strength as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches in the bulk graviton scenario with $k/\bar{M}_{Pl} = 0.5$. The different colored lines correspond to the searches entering the combination.

11.3.4 Significance at 2 TeV

ATLAS reported an excess in the all-hadronic $VV \rightarrow q\bar{q}q\bar{q}$ search corresponding to a local significance of 3.4σ for a W' resonance with a mass of 2 TeV [123]. For CMS, the largest deviation of 2.2σ has been observed in the semi-leptonic $WH \rightarrow \ell\nu b\bar{b}$ search described in this work (Chapter 9). The combined significance of the 8 and 13 TeV CMS searches in the range 1.8–2.0 TeV is here evaluated and showed in Figure 11.4 for a W' hypothesis.

Combining all 8 TeV VH searches in the W' hypothesis, the local significance of the excess at 1.8 TeV is slightly reduced to 2.1σ . Combining all 8 TeV VV and VH searches, it is increased back to 2.2σ , since the VV searches observed a small deviation in the same resonance mass range. However, in combination with the 13 TeV VV and VH searches, the overall significance at 1.8 TeV is reduced to 0.8σ . This remains the largest significance for the overall combination of 8+13 TeV searches considering all signal hypothesis over the mass range 1.8–2.0 TeV, thus not supporting the excesses observed in the two individual channels in 8 TeV data.

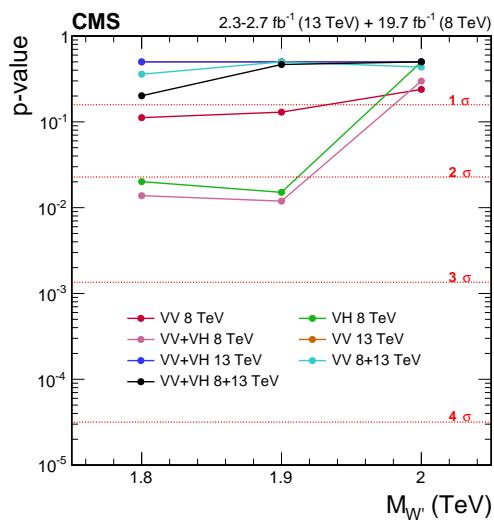


Figure 11.4: Local p-values of the excesses observed in the resonance mass range 1.8–2 TeV in the various combinations of searches for a W' hypothesis.

CHAPTER 12

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

14.1 Design of the CMS Pixel Barrel Detector

14.2 Detector modules

14.2.1 Sensor

14.2.2 Readout Chip

14.2.3 Token Bit Manager

14.3 Readout and control system

14.3.1 Analog readout chain

14.3.2 Front End Driver

14.3.3 Supply Tube

14.3.4 Communication and Control Unit

14.3.5 Front End Controller

14.4 Pixel Online Software

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

Optimization and commissioning for LHC Run II

- 15.1 Radiation damage after LHC Run I
- 15.2 Optimization for LHC Run II
 - 15.2.1 Overview of pixel calibrations
 - 15.2.2 Temperature dependence
- 15.3 Commissioning for LHC Run II
 - 15.3.1 Installation into CMS
 - 15.3.2 Check out of optical connections
 - 15.3.3 Adjustment of readout chain settings
 - 15.3.4 Optimisation of signal performance

Phase I Upgrade of the CMS Pixel Barrel Detector

16.1 Motivations

16.2 Summary of changes

16.3 The digital readout chain

16.4 The Phase I supply tubes

16.5 The test stand

16.6 Supply tubes assembly and commissioning

16.7 Detector commissioning

CHAPTER 17

Conclusions

Part III

Summary

APPENDIX A

Studies on track reconstruction problems

A scan of the displays of all the events in 8 TeV data with $m_{\text{WH}} > 1.6$ TeV passing all the selection criteria for the $\ell\nu b\bar{b}$ final state (Table 8.1), reported that presence of two events characterized by a rare specific kind of noise. This noise arises from an anomalous behaviour of the tracking algorithm in the transition region between the barrel (TOB) and endcap (TEC) regions of the silicon tracker, namely, in the pseudorapidity range $1 < |\eta| < 1.5$ (Fig. 3.4). As a consequence, many fake (displaced) tracks are associated to the selected H jet candidate. Figure A.1 shows the event display of one of the two events affected by this problem, while Figure A.2 shows the same feature in simulation.

In order to reject this type of noise, it is common in CMS analyses to apply a standard filtering algorithm that discards the event if there is an anomalous amount of tracks that have been seeded in the TOB-TEC transition region. The efficiency of this filter on signal events is about 97% independently on the H jet p_T .

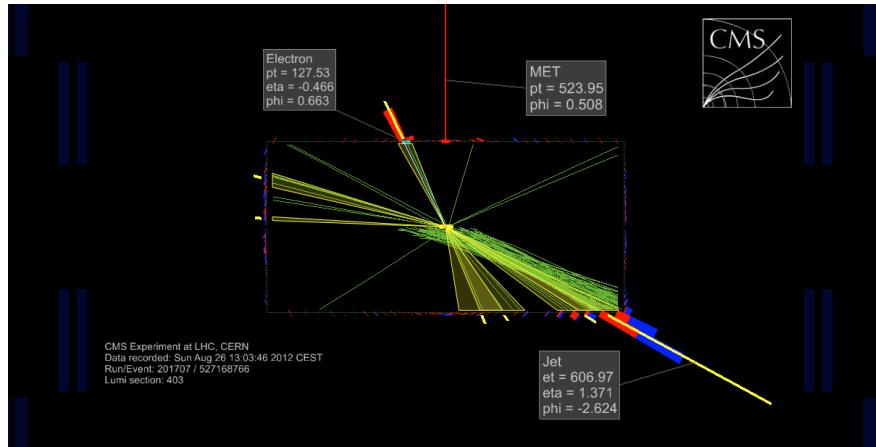


Figure A.1: Display of one typical anomalous event found in data recorded by the CMS experiment. Many fake and displaced tracks are reconstructed creating a bias in the jet reconstruction. Only tracks with p_T larger than 2 GeV are shown.

However, further checks performed on the anomalous events showed that after applying the standard filter residual noise can still be identified in the problematic η region. Therefore, it has been decided for the analysis described in this work to apply an additional requirement on the η of the selected H jet candidate (Section 6.4.2). In particular, CA8 jets are rejected if their pseudorapidity falls in the problematic region $1 < |\eta| < 1.8$. As described in the following, the choice for this fiducial cut is motivated by the disagreement between data and simulation in the rate at which the noise occur.

The efficiency of the standard filter is studied as a function of the H jet p_T and η in a dijet sample with high statistics in both data and simulation. The sample is selected requiring at least two jets, with $p_T > 400$ GeV for the leading jet and $p_T > 80$ GeV for the sub-leading

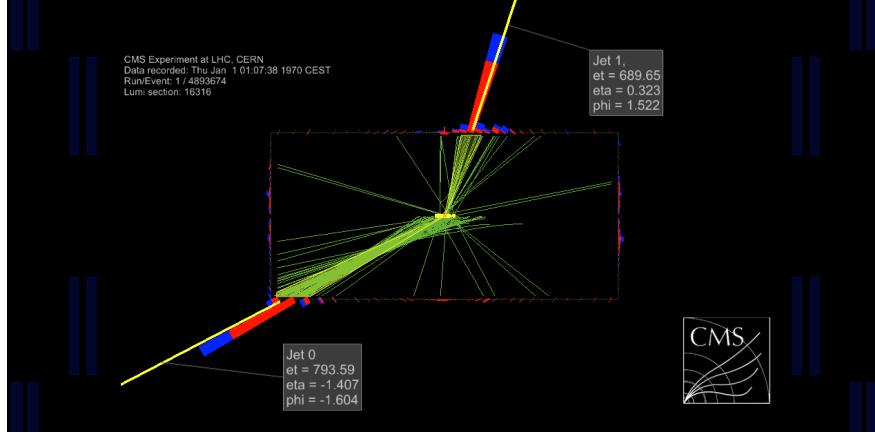


Figure A.2: Display of one typical anomalous event in simulation. Only tracks with p_T larger than 2 GeV are shown.

one. At least one of the jet has to be b-tagged using the same combined b tagging algorithm as for the main analysis selection, representing thus the H jet candidate. The jet that fails the b tagging is required to have low pruned mass ($m_{jet} < 40$ GeV).

Figure A.3 shows the effect of the filter on the jet η distribution comparing data, simulated signal and QCD background: the signal distribution is rather unaffected while data and QCD background distributions show a reduction of events in the problematic η region after applying the filter, as expected.

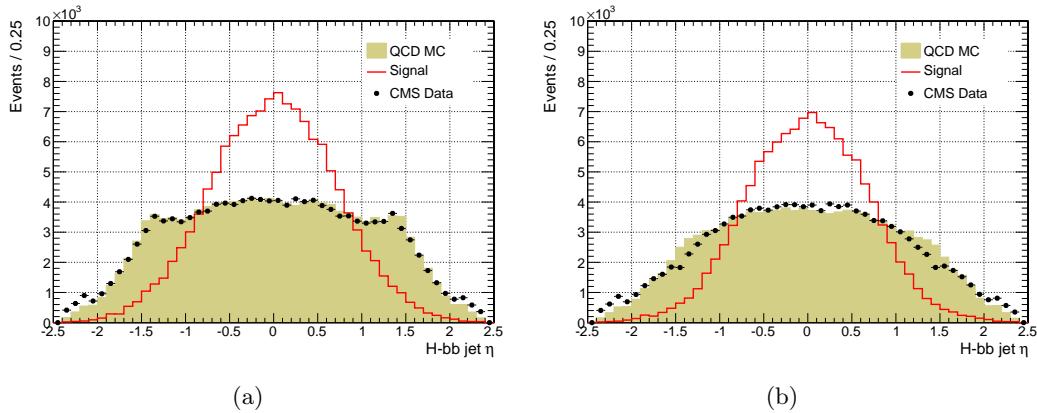


Figure A.3: Comparison of the H jet η distributions for data, and simulated signal and QCD background before (a) and after (b) applying the tracking noise filter. Signal jets are mostly central in the detector.

Figure A.4 shows the filter efficiency on data and simulated signal and QCD background as a function of the H jet candidate p_T for different jet η regions. A little dependence of the filter efficiency with the jet p_T is observed in the regions $0 < |\eta| < 1$ and $1.0 < |\eta| < 1.5$, while in the forward region $1.5 < |\eta| < 2.4$ the efficiency decreases with the jet p_T . The performance of the filter in the different η regions is summarized in Figure A.5(a). A large discrepancy between data and simulation is found in the pseudorapidity region $1.0 < |\eta| < 1.8$, where the simulation does not sufficiently well describe the full material budget of the tracking detector. The same studies are also performed removing the b-tagging requirement. The filter efficiency as a function of the leading jet η for this case is shown in Fig. A.5(b), for both data and

simulation. The increase in efficiency compared to what obtained in b-tagging shows that the b-tagging requirement enriches the samples with events characterized by this noise up to 30%, making this analysis systematically prone to it.

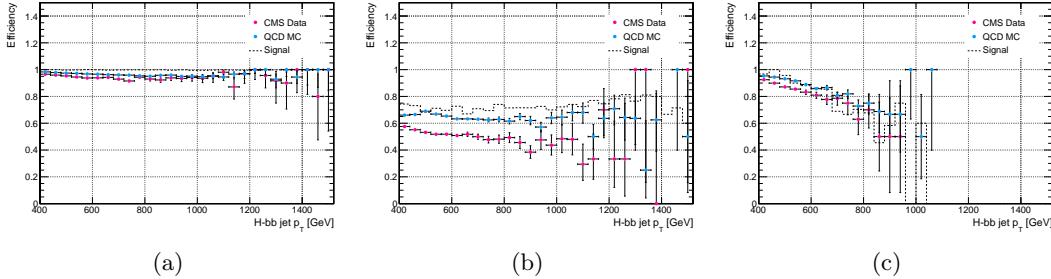


Figure A.4: Efficiency of the tracking noise filter as a function of the H jet p_T for data, and simulated signal and QCD background for jets reconstructed in the pseudorapidity regions $0 < |\eta| < 1$ (a), $1.0 < |\eta| < 1.5$ (b), and $1.5 < |\eta| < 2.4$ (c).

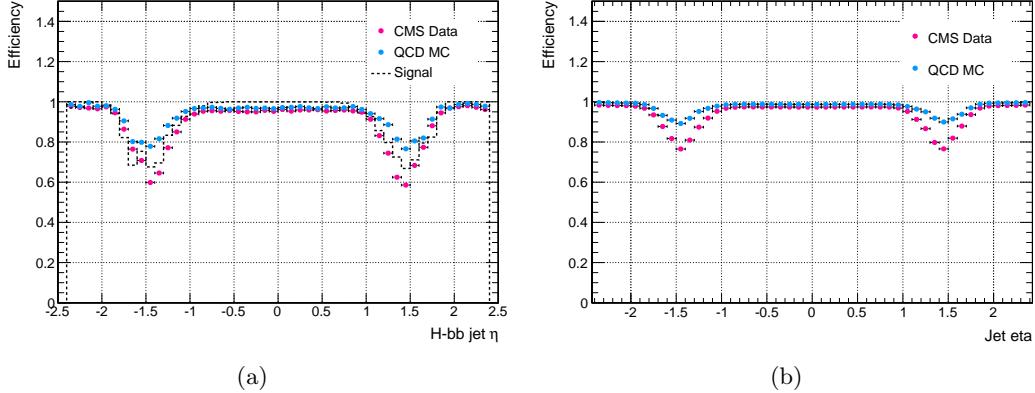


Figure A.5: Efficiency of the tracking noise filter as a function of the leading jet p_T for data, and simulated signal and QCD background. (a) The leading jet is required to be b-tagged with the combined b tagging algorithm used in the main analysis. (b) The b tagging requirements for the leading jet are removed.

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] M. A. Dobbs et al., “Les Houches guidebook to Monte Carlo generators for hadron collider physics”, in *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003*, pp. 411–459. 2004. arXiv:hep-ph/0403045.
- [15] J. C. Collins and D. E. Soper, “The Theorems of Perturbative QCD”, *Ann. Rev. Nucl. Part. Sci.* **37** (1987) 383–409, doi:10.1146/annurev.ns.37.120187.002123.
- [16] ZEUS Collaboration, “The ZEUS detector: Status report 1993”.,
- [17] H1 Collaboration, “The H1 detector at HERA”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **386** (1997), no. 2, 310 – 347, doi:[http://dx.doi.org/10.1016/S0168-9002\(96\)00893-5](http://dx.doi.org/10.1016/S0168-9002(96)00893-5).

- [18] G. Altarelli and G. Parisi, “Asymptotic freedom in parton language”, *Nuclear Physics B* **126** (1977), no. 2, 298 – 318,
`doi: http://dx.doi.org/10.1016/0550-3213(77)90384-4.`
- [19] J. Pumplin et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **07** (2002) 012,
`doi: 10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.`
- [20] H.-L. Lai et al., “New parton distributions for collider physics”, *Phys. Rev. D* **82** (2010) 074024, `doi: 10.1103/PhysRevD.82.074024, arXiv: 1007.2241.`
- [21] R. D. Ball et al., “Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology”, *Nucl. Phys. B* **849** (2011) 296,
`doi: 10.1016/j.nuclphysb.2011.03.021, arXiv: 1101.1300.`
- [22] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852, `doi: 10.1016/j.cpc.2008.01.036, arXiv: 0710.3820.`
- [23] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, `doi: 10.1088/1126-6708/2006/05/026, arXiv: hep-ph/0603175.`
- [24] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128,
`doi: 10.1007/JHEP06(2011)128, arXiv: 1106.0522.`
- [25] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070,
`doi: 10.1088/1126-6708/2007/11/070, arXiv: 0709.2092.`
- [26] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in heavy flavor production”, *JHEP* **08** (2003) 007,
`doi: 10.1088/1126-6708/2003/08/007, arXiv: hep-ph/0305252.`
- [27] GEANT4 Collaboration, “GEANT4 — a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, `doi: 10.1016/S0168-9002(03)01368-8.`
- [28] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV”, *JHEP* **09** (2011) 109,
`doi: 10.1007/JHEP09(2011)109, arXiv: 1107.0330.`
- [29] CMS Collaboration, “Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9, 2.76$, and 7 TeV”, *JHEP* **04** (2013) 072,
`doi: 10.1007/JHEP04(2013)072, arXiv: 1302.2394.`
- [30] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, “Heavy Vector Triplets: Bridging Theory and Data”, *JHEP* **09** (2014) 060, `doi: 10.1007/JHEP09(2014)060, arXiv: 1402.4431.`
- [31] G. Burdman, M. Perelstein, and A. Pierce, “Large Hadron Collider tests of a little Higgs model”, *Phys. Rev. Lett.* **90** (2003) 241802,
`doi: 10.1103/PhysRevLett.90.241802, arXiv: hep-ph/0212228.` [Erratum: *Phys. Rev. Lett.* 92, 049903(2004)].

- [32] J. M. Campbell, R. K. Ellis, and D. L. Rainwater, “Next-to-leading order QCD predictions for $W + 2$ jet and $Z + 2$ jet production at the CERN LHC”, *Phys. Rev.* **D68** (2003) 094021, doi:10.1103/PhysRevD.68.094021, arXiv:hep-ph/0308195.
- [33] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.
- [34] J. M. Campbell and R. K. Ellis, “Top-quark processes at NLO in production and decay”, *J. Phys.* **G42** (2015), no. 1, 015005, doi:10.1088/0954-3899/42/1/015005, arXiv:1204.1513.
- [35] J. M. Campbell, R. K. Ellis, and F. Tramontano, “Single top production and decay at next-to-leading order”, *Phys. Rev.* **D70** (2004) 094012, doi:10.1103/PhysRevD.70.094012, arXiv:hep-ph/0408158.
- [36] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ”, *Phys. Rev.* **D86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.
- [37] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [38] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.
- [39] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [40] 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.
- [41] W. Adam, B. Mangano, T. Speer, and T. Todorov, “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Dec, 2006.
- [42] 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.
- [43] R. Frhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.
- [44] T. Speer et al., “Vertex Fitting in the CMS Tracker”, Technical Report CMS-NOTE-2006-032, CERN, Feb, 2006.
- [45] 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.
- [46] W. Waltenberger, R. Frhwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 12, N343.

- [47] CMS Collaboration, “CMS Tracking Performance Results”.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTRK>.
- [48] S. Baffioni et al., “Electron reconstruction in CMS”, *Eur. Phys. J.* **C49** (2007) 1099–1116, doi:10.1140/epjc/s10052-006-0175-5.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] CMS Collaboration, “CMS ECAL Performance Results”.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResults>.
- [53] 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.
- [54] 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.
- [55] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, 2006.
- [56] CMS Collaboration, “Additional Performance Results for EPS 2015”,.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] 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.

- [63] 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](https://arxiv.org/abs/hep-ph/9907280).
- [64] 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](https://arxiv.org/abs/0802.1189).
- [65] 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](https://arxiv.org/abs/1410.4227).
- [66] 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.
- [67] 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.
- [68] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *Submitted to: JINST* (2016) [arXiv:1607.03663](https://arxiv.org/abs/1607.03663).
- [69] T. C. collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *Journal of Instrumentation* **6** (2011), no. 11, P11002.
- [70] CMS Collaboration, “Pileup Jet Identification”, Technical Report CMS-PAS-JME-13-005, CERN, 2013.
- [71] CMS Collaboration, “Pileup Removal Algorithms”, Technical Report CMS-PAS-JME-14-001, CERN, 2014.
- [72] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, 2014.
- [73] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, [arXiv:1111.6097](https://arxiv.org/abs/1111.6097).
- [74] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder”, *Phys. Lett.* **B641** (2006) 57–61, doi:10.1016/j.physletb.2006.08.037, [arXiv:hep-ph/0512210](https://arxiv.org/abs/hep-ph/0512210).
- [75] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett.* **B659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, [arXiv:0707.1378](https://arxiv.org/abs/0707.1378).
- [76] “Jet energy scale and resolution performances with 13TeV data”,.
- [77] 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](https://arxiv.org/abs/1211.4462).
- [78] W. Waltenberger, “Adaptive Vertex Reconstruction”, Technical Report CMS-NOTE-2008-033, CERN, Geneva, Jul, 2008.
- [79] 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](https://arxiv.org/abs/1102.3194).

- [80] CMS Collaboration Collaboration, “Performance of b tagging at $\sqrt{s}=8$ TeV in multijet, $t\bar{t}$ and boosted topology events”, Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, 2013.
- [81] 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.
- [82] T. C. collaboration, “Missing transverse energy performance of the CMS detector”, *Journal of Instrumentation* **6** (2011), no. 09, P09001.
- [83] 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.
- [84] 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.
- [85] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.
- [86] J. Bauer, “Prospects for the Observation of Electroweak Top-Quark Production with the CMS Experiment”. PhD thesis, Karlsruher Institut für Technologie (KIT), 2010.
- [87] 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.
- [88] 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.
- [89] 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.
- [90] 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.
- [91] 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.
- [92] 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.
- [93] 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.
- [94] CMS Collaboration, “V Tagging Observables and Correlations”, Technical Report CMS-PAS-JME-14-002, CERN, 2014.
- [95] 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.

- [96] M. Czakon, D. Heymes, and A. Mitov, “High-precision differential predictions for top-quark pairs at the LHC”, *Phys. Rev. Lett.* **116** (2016), no. 8, 082003, doi:10.1103/PhysRevLett.116.082003, arXiv:1511.00549.
- [97] CMS Collaboration, “Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos, and b quarks”, arXiv:1610.08066. Submitted to PLB.
- [98] M. J. Oreglia, “A study of the reactions $\psi' \rightarrow \gamma\gamma\psi'$ ”. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236.
- [99] CMS Collaboration, “Measurement of the single-top-quark t -channel cross section in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **12** (2012) 035, doi:10.1007/JHEP12(2012)035, arXiv:1209.4533.
- [100] P. Kant et al., “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Comput. Phys. Commun.* **191** (2015) 74–89, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.
- [101] CMS Collaboration, “Measurement of W+W- and ZZ production cross sections in pp collisions at $\text{sqrt}(s) = 8$ TeV”, *Phys. Lett.* **B721** (2013) 190–211, doi:10.1016/j.physletb.2013.03.027, arXiv:1301.4698.
- [102] T. Gehrmann et al., “ W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD”, *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.
- [103] CMS Collaboration, “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV”, *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016. [JINST8,9009(2013)].
- [104] CMS Collaboration, “CMS Luminosity Measurement for the 2015 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-15-001, CERN, 2015.
- [105] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013.
- [106] M. Cacciari et al., “The $t\bar{t}$ cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence”, *JHEP* **04** (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.
- [107] S. Catani, D. de Florian, M. Grazzini, and P. Nason, “Soft gluon resummation for Higgs boson production at hadron colliders”, *JHEP* **07** (2003) 028, doi:10.1088/1126-6708/2003/07/028, arXiv:hep-ph/0306211.
- [108] M. Botje et al., “The PDF4LHC Working Group Interim Recommendations”, (2011). arXiv:1101.0538.
- [109] S. Alekhin et al., “The PDF4LHC Working Group Interim Report”, (2011). arXiv:1101.0536.
- [110] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC”, *Eur. Phys. J. C* **63** (2009) 189, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

- [111] R. D. Ball et al., “Impact of heavy quark masses on parton distributions and LHC phenomenology”, *Nucl. Phys. B* **849** (2011) 296,
`doi:10.1016/j.nuclphysb.2011.03.021`, [arXiv:1101.1300](https://arxiv.org/abs/1101.1300).
- [112] A. L. Read, “Presentation of search results: The CL_S technique”, *J. Phys. G* **28** (2002) 2693, `doi:10.1088/0954-3899/28/10/313`.
- [113] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, `doi:10.1016/S0168-9002(99)00498-2`, [arXiv:hep-ex/9902006](https://arxiv.org/abs/hep-ex/9902006).
- [114] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554,
`doi:10.1140/epjc/s10052-011-1554-0`, `10.1140/epjc/s10052-013-2501-z`,
[arXiv:1007.1727](https://arxiv.org/abs/1007.1727). [Erratum: *Eur. Phys. J. C* **73** (2013) 2501].
- [115] A. Wald, “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large”, *Transactions of the American Mathematical Society* **54** (1943), no. 3, 426–482.
- [116] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, *Ann. Math. Statist.* **9** (03, 1938) 60–62,
`doi:10.1214/aoms/1177732360`.
- [117] E. Gross and O. Vitells, “Trial factors or the look elsewhere effect in high energy physics”, *Eur. Phys. J. C* **70** (2010) 525–530,
`doi:10.1140/epjc/s10052-010-1470-8`, [arXiv:1005.1891](https://arxiv.org/abs/1005.1891).
- [118] F. Garwood, “Fiducial Limits for the Poisson Distribution”, *Biometrika* **28** (1936) 437, `doi:10.1093/biomet/28.3-4.437`.
- [119] CMS Collaboration, “Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **02** (2016) 145, `doi:10.1007/JHEP02(2016)145`, [arXiv:1506.01443](https://arxiv.org/abs/1506.01443).
- [120] CMS Collaboration, “Search for new resonances decaying via WZ to leptons in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett. B* **740** (2015) 83,
`doi:10.1016/j.physletb.2014.11.026`, [arXiv:1407.3476](https://arxiv.org/abs/1407.3476).
- [121] CMS Collaboration, “Search for massive WH resonances decaying into the $\ell\nu b\bar{b}$ final state at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **76** (2016) 237,
`doi:10.1140/epjc/s10052-016-4067-z`, [arXiv:1601.06431](https://arxiv.org/abs/1601.06431).
- [122] CMS Collaboration, “Search for Narrow High-Mass Resonances in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV Decaying to a Z and a Higgs Boson”, *Phys. Lett. B* **748** (2015) 255, `doi:10.1016/j.physletb.2015.07.011`, [arXiv:1502.04994](https://arxiv.org/abs/1502.04994).
- [123] ATLAS Collaboration, “Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *JHEP* **12** (2015) 055, `doi:10.1007/JHEP12(2015)055`, [arXiv:1506.00962](https://arxiv.org/abs/1506.00962).