Universidade Federal do ABC

Search for narrow high-mass resonances in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ decaying to a Z and a Higgs bosons in the CMS detector at LHC

César Augusto Bernardes

Advisor: Pedro Galli Mercadante

January 25, 2016

Ph.D. Thesis in Physics

Busca por ressonâncias massivas em colisões próton-próton a $\sqrt{s} = 8\,\mathrm{TeV}$ no estado final com um bóson Z e um bóson de Higgs no detector CMS do LHC

Resumo

Apresentamos uma busca no detector CMS por ressonâncias massivas decaindo em um bóson de Higgs e um bóson Z do Modelo Padrão no estado final com um par de léptons tau e um par de quarks. Analisamos $19,7/\mathrm{fb}$ de luminosidade integrada em colisões próton-próton com energia no centro de massa de $8\,\mathrm{TeV}$ no LHC. No intervalo de massa da ressonância entre $1,0 - 2,5\,\mathrm{TeV}$, os bósons Z e Higgs são produzidos com um momento grande comparado a suas massas, isso implica que os produtos finais provenientes de cada par de taus ou quarks devem ser detectados dentro de pequenas regiões espaciais. Técnicas de subestrutura de jatos são utilizadas para identificar o bóson Z que decai em quarks. Para o Higgs decaindo em taus é definido um método modificado de reconstrução de pares de taus com alto momento e colimados.

Combinando os resultados das análises de cada modo de decaimento do tau, são excluídas com 95% de confiança seções de choque de produção de ressonâncias com spin-1 no intervalo de $0,9 - 27,8\,\mathrm{fb}$, dependendo da massa da ressonância. Consideramos como modelo de referência uma teoria efetiva de Higgs composto, que prediz ressonâncias massivas com pequena razão de ramificação em férmions. Comparando os limites experimentais com expressões analíticas dos parâmetros de uma Lagrangeana de um modelo simplificado, obtemos limites sobre esses parâmetros, os quais estão diretamente relacionados com a produção e decaimento das ressonâncias.

Search for narrow high-mass resonances in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ decaying to Z and Higgs bosons in the CMS detector at LHC

Abstract

We present a search in the CMS detector for a high-mass and narrow resonance decaying into a Higgs and a Z SM bosons in the final state with a pair of tau leptons and a pair of quarks. We analyze $19.7/\mathrm{fb}$ of integrated luminosity of $\sqrt{s} = 8\,\mathrm{TeV}$ proton-proton collisions from LHC. In the resonance mass range of interest $(1.0 - 2.5\,\mathrm{TeV})$, the Z and Higgs bosons are produced with large momentum compared with their masses, which implies that the final products of the two quarks and the two taus must be detected within a small angular separation (boosted topology). Jet substructure techniques are used to identify the boosted Z boson decaying hadronically and a modified approach is considered to reconstruct very collimated pairs of tau leptons from boosted Higgs decay.

From a combination of all possible decay modes of the tau lepton, heavy spin-1 resonances production cross sections are excluded at 95% C.L. in a range between 0.9 and 27.8 fb, depending on the resonance mass. We consider as benchmark model an effective description of a strongly coupled theory based on the Minimal Composite Higgs Model, which predicts heavy resonances with suppressed branching fractions into fermions. Comparing the experimental limits with analytical expressions of the parameters of a simplified model Lagrangian we obtain observed limits on these parameters, which are directly connected with the heavy resonance production and decay.

Agradecimentos

Esta é uma seção bastante "perigosa" porque a chance de se cometer injustiças é bastante grande...Peço desculpas às pessoas que foram importantes na execução deste projeto e que não receberam os devidos agradecimentos.

Primeiramente, gostaria de agradecer a meu orientador, Pedro Galli Mercadante, que teve participação crucial no desenvolvimento deste trabalho, passando-me conhecimentos importantíssimos para minha carreira profissional.

Gostaria de agradecer a todo o grupo SPRACE pela estrutura proporcionada, tanto do ponto de vista de ambiente de trabalho, com salas de estudos/reuniões e equipamentos próprios para computação científica, como também do ponto de vista de auxílio profissional com integrantes experientes em diversas áreas como física, engenharia e ciências da computação. Neste grupo também tive a sorte de fazer grandes amizades e gostaria de agradecer explicitamente a: Sérgio Novaes, Eduardo Gregores, Sandra Padula, Pedro Mercadante, Sérgio Lietti, Ailton Shinoda, Mário Vaz, Eduardo Bach, Jadir Marra, Márcio Costa, Rogério Iope, Sidney Santos, Angelo Santos, Chang-Seong Moon, Sudha Ahuja, Sunil Dogra, Thiago Tomei, Thiago Paiva, David Romero, José Vargas e Lucas Ramalho.

Gostaria de agradecer à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa de estudos concedida, processo 2012/24464-3, e pelo auxílio através do projeto temático, processo 2008/02799-8.

À Universidade Federal do ABC (UFABC) pela bolsa de estudos concedida e pelo ambiente de estudos, com excelentes funcionários, e pelos auxílios em participação de eventos e estágio de pesquisa.

À CAPES que através do programa Ciência Sem Fronteiras me concedeu uma bolsa de estudos, processo BEX 0494/13-5, com duração de um ano no CERN.

In the period I was at CERN I had the pleasure to work with very nice people. I was lucky and I had many co-advisors there. I would like to thanks: Maurizio Pierini, Thiago Tomei, Alessio Bonato, Francesco Santanastasio, Roberto Covarelli, Clemens Lange, Andreas Hinzmann, and Christian Veelken for all the help and guidance.

I would like to thanks my workmates at CERN: Aniello Spiezia, Alexandra Carvalho, Camilla Galloni, Jennifer Ngadiuba, Mengmeng Wang, Shuai Liu, and Raffaele Gerosa, for all the help and friendship.

I would like to thanks my housemates at CERN: Daniel Berkowitz, Cristhian Lizarraga, Michel Reymond, Georg Sieber, Daniel Funke, and Matthias Michels, for the pleasant stay I had in Saint-Genis-Pouilly.

Agradeço à toda minha família pelo apoio total em minha escolha pela carreira de física. Em especial a meus pais, Carlos Roberto Bernardes e Antonia Soares de Lima Bernardes, que estiveram do meu lado durante toda a minha vida. À minha noiva, Milene da Silva, por todo o incentivo, carinho e compreensão.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Very recently, the validity of the standard model (SM) of particle physics has been confirmed by the discovery of a Higgs boson with mass near $125\,\mathrm{GeV}$ by the ATLAS and CMS experiments [1, 2]. Though the SM successfully describes a broad range of high energy phenomena, the answer to remaining questions regarding the theoretical structure of the SM, particularly the hierarchy problem, leads naturally to the introduction of physics beyond the standard model (BSM), possibly at the TeV scale [3, 4, 5, 6, 7, 8]. Many of the BSM models predict the existence of heavy resonances with masses of the order of a TeV, which may have sizable couplings to gauge and Higgs boson fields of the SM [9, 10, 11, 12]. We consider here one important family among these models, which incorporate composite Higgs bosons [11, 12]. In these models, the Higgs boson is a pseudo-Nambu-Goldstone boson of a broken global symmetry. Other composite bound states beyond the Higgs are expected to exist and could be experimentally observed.

Several searches for TeV scale resonances decaying into pairs of vector bosons or Higgs bosons have been performed by the ATLAS and CMS experiments [13, 14, 15, 16, 17, 18, 19, 20, 21]. Both collaborations have studied criteria for reconstruction and identification of high-momentum bosons decaying leptonically and hadronically. Kinematic selections have been tuned in order to have the best discrimination between background and signal events. In CMS, since 2011, there is a group in charge of the experimental searches for these signatures. We have investigated semileptonic channels (one boson decaying into leptons and the other hadronically) and all-hadronic channels (with both bosons decaying hadronically).

In this thesis, we search for a resonance with a mass in the range $0.8 - 2.5\,\mathrm{TeV}$ decaying to ZH, where the Z boson decays to $\mathrm{q\bar{q}}$ and the Higgs boson decays to $\tau^+\tau^-$. It is assumed that the natural width of the resonance is negligible in comparison to the experimental mass resolution, which is between 6% and 10% of the mass of the resonance, depending on the mass.

The theoretical model used as benchmark in this work is described in Ref. [22]. In this model a heavy $SU(2)_L$ vector triplet (HVT) containing neutral (Z') and charged (W'$^{\pm}$) spin-1 states is introduced. This scenario is well-motivated in cases where the new physics sector is either weakly coupled [23], or strongly coupled, e.g., in the minimal composite model [24]. The cross sections and branching fractions ($\mathcal{B}$) for the heavy triplet model depend on the new physics scenario under study and can be characterized by three parameters in the phenomenological Lagrangian: the strength of the couplings to fermions c_{F}, to the Higgs c_{H}, and the self-coupling g_{V}. In the case of a strongly coupled sector, the new heavy resonance has larger couplings to the W, Z, and H bosons, resulting in larger branching

fractions for the diboson final states.

In the high-mass case under study, the directions of the particles stemming from Z and H boson decays are separated by a small angle. This feature is referred to as the "boosted regime" or "boosted topology". For the case of $Z \to q\bar{q}$, this results in the presence of one single reconstructed jet after hadronization called a "Z-jet" or "fat-jet". For $H \to \tau^+\tau^-$, it is observed two narrow jets separated by a small angle and with a low charge multiplicity. This last feature is because most τ decays are in one or three charged tracks plus some neutral hadrons (one or three prongs). The novel feature of this study is the reconstruction and identification of a τ pair in the boosted regime. For this purpose, it is important to define criteria to avoid at maximum the interference between the two τ candidates.

In the following, we label τ decays in a simplified way: $\tau^\pm \to e^\pm \nu \bar{\nu}$ as "τ_e", $\tau^\pm \to \mu^\pm \nu \bar{\nu}$ as "τ_μ", and $\tau^\pm \to (n\,\pi)(m\,K)\nu$ as "τ_h", where n and m can be 0, 1, 2, or 3, and the pions and kaons can be either charged or neutral. Six channels, depending on the combinations of τ decays, are studied separately and labeled as all-leptonic ($\tau_e\tau_e$, $\tau_e\tau_\mu$, $\tau_\mu\tau_\mu$), semileptonic ($\tau_e\tau_h$, $\tau_\mu\tau_h$), and all-hadronic ($\tau_h\tau_h$). In this thesis, we will describe in details the all-hadronic channel.

The experimental strategy is to reconstruct and identify the two bosons and to combine their information into a variable that can discriminate between signal and background and on which a statistical study can be performed. This variable is the estimated mass of the Z' after applying dedicated reconstruction techniques to the boosted $q\bar{q}$ and $\tau\tau$ pairs (m_{ZH}). The m_{ZH} distribution would show an excess of events at the assumed Z' mass if a signal is present.

Chapter 2

The CMS experiment at the LHC

2.1 Accelerator physics

In this section we will discuss briefly some important characteristics of some types of particle accelerators. More specifically, we are going to talk about some basics of colliders. In general, the circular colliders use two beams of particles accelerated in opposite directions, which interact in a given point where the detectors are located. The energy of the particles that make up the beams are the same, so that the energy in the center-of-mass (CM) is:

$$E_{\mathrm{CM}} = 2E_{\mathrm{p}}, \tag{2.1}$$

where E_{p} refers to the energy of each particle in the beam.

The research with accelerators has a very vast applicability, and not only in physics, but in medicine, biology, among other areas [25, 26]. In the case of the Large Hadron Collider (LHC), the main purpose is in the physics area, more specifically in High-Energy Physics. In this case, we need a very big amount of energy for each particle in the beam, so that it is possible to create heavy particles (based on the Einstein relation $E = mc^2$), like the Z boson, and investigate deeply the subatomic structure (as expected by the de Broglie relation $\lambda = 2\pi/p$).

Two common types of colliders are the linear and circular. In the linear collider, the bunch crossing occurs just one time (between the beam and target), then it is not possible to reuse the particles that did not interact with this target. In the case of the circular collider this is not a problem, because the two beams are reused many times while circulating in the ring. However, to maintain the circular orbit we have to use charged particles, which lose energy by the synchrotron radiation,

$$\Delta E_{\mathrm{sinc}} \propto \frac{E_{\mathrm{p}}^4}{m^4 R_{\mathrm{o}}}, \tag{2.2}$$

where m is the mass of the accelerated particle and R_{o} is the radius of the orbit.

The type of the particles in the bunch is related with the purpose of the experiments. For example, the $e^- e^+$ colliders are in general dedicated to precision physics. We know with precision the energy of the electrons in the initial state, and since the electron is an elementary particle, the environment is much cleaner than in a hadronic collision. In a proton-proton collider, for example, we are interested in collisions among the partons that make up the protons. In this case we do not know very well the energy of the initial partons. Since we fix a direction to accelerate the particles, it is reasonable to infer that the momentum of the

partons in the perpendicular direction is approximately zero, but in the direction of the beam it is not easy to estimate the parton momentum because they share in a non-stationary way the momentum of the proton in this direction [1]. In searches for new physics, since in general we do not know in which energy scale it will appear, the hadron colliders are more appropriated because for a given energy of the hadron, the colliding partons will have CM energies in a considerable range below the hadronic CM energy. Since hadrons are heavier than the electrons, they lose less energy by the synchrotron radiation, allowing higher energies for the colliding particles.

In order to relate the theoretical quantities with the experimental ones (number of events), it is defined a quantity called instantaneous luminosity ($\mathcal{L}$), which is related with the number of events produced in a collision per unit of time ($\frac{dN}{dt}$). These events occurring with a probability described by a cross section σ (in general σ is defined in units of barn (b) with $1\,\text{b} = 10^{-24}\,\text{cm}^2$), are related with the instantaneous luminosity by,

$$\frac{dN}{dt} = \sigma \cdot \mathcal{L}. \tag{2.3}$$

In the case where the beams have N_B bunches, each with N particles, which cross with a frequency f, the instantaneous luminosity is defined by,

$$\mathcal{L} = \frac{f \cdot N_{\text{B}} \cdot N^2}{4\pi\sigma_{\text{x}}\sigma_{\text{y}}}, \tag{2.4}$$

where each bunch has a gaussian shape in the orthogonal directions with respect to the beam, with width σ_{x} in x and σ_{y} in y. In the case of the LHC, the expression above is multiplied by an additional factor to take into account the fact that, in general, the bunches cross with an angle with a small deviation from π [27].

The total number of events in an interval of time is obtained integrating Eq. 2.3. Supposing σ independent of the time, it is possible to write the total number of events (N_{tot}) as function of a quantity called integrated luminosity L_{int} ($L_{\text{int}} = \int dt\mathcal{L}$, with units of $[L_{\text{int}}] = (10^{-12}\text{b})^{-1} = \text{pb}^{-1}$, for example):

$$N_{\text{tot}} = \sigma \cdot L_{\text{int}}. \tag{2.5}$$

2.2 The Large Hadron Collider

The LHC complex is located in the border between France and Switzerland, and its accelerators and detectors are located in the underground between 50 e 175 m below the surface, see Fig. 2.1. The LHC [27] is a circular accelerator that operates with protons and ions (Pb ions). Since the two beams circulate in opposite directions, it is necessary two rings with superconducting dipole magnets in many points of the accelerator in order to bend the particles trajectory. In Fig. 2.2 we show a transverse section of the accelerator and a magnet dipole represented by its field lines.

The protons used in the collisons are obtained ionizing hydrogen atoms. Before they reach the LHC ring, the beam passes through a process of acceleration called injection chain,

---

[1] The fractions of the proton energy carried by the partons are described by complicated distributions (parton distribution functions), which are deduced based on observed events.

where the protons pass in a sequence of accelerators "Linear Accelerator 2 (Linac2)", "Proton Synchroton Booster (PSB)", "Proton Synchroton (PS)", and "Super Proton Synchroton (SPS)", so that when the beam reach the LHC the protons have an energy of around $450 \ \mathrm{GeV}$, see Fig. 2.3. The total energy is reached in the LHC accelerator.

The main ring where the particles collide has a circumference of approximately $26.7 \ \mathrm{km}$. There are four points where the bunches are bent to cross and the detectors are built around this region of collision. There are four main detectors:

- **A L**arge **I**on **C**ollider **E**xperiment (ALICE): Specialized in ion collisions. It is studied, for example, quark/gluon plasma systems.

- **A T**oroidal **L**HC **A**pparatu**S** (ATLAS): General purpose detector, with the main topics in Higgs physics and searches for physics beyond the SM.

- **C**ompact **M**uon **S**olenoid (CMS): General purpose detector like ATLAS. Possibility to perform independent studies and combine the results with ATLAS.

- **L**arge **H**adron **C**ollider **b**eauty (LHCb): Focus mainly in the b-physics, for example, study of CP violation in systems involving B mesons.

Besides the detectors above, there are other two: TOTEM (**TOT**al **E**lastic and diffractive cross section **M**easurement) and LHCf (**L**arge **H**adron **C**ollider **f**orward), which study regions closer to the beam pipe (forward detectors). They are important in the measurement of total proton-proton cross section and monitoring of the LHC luminosity.
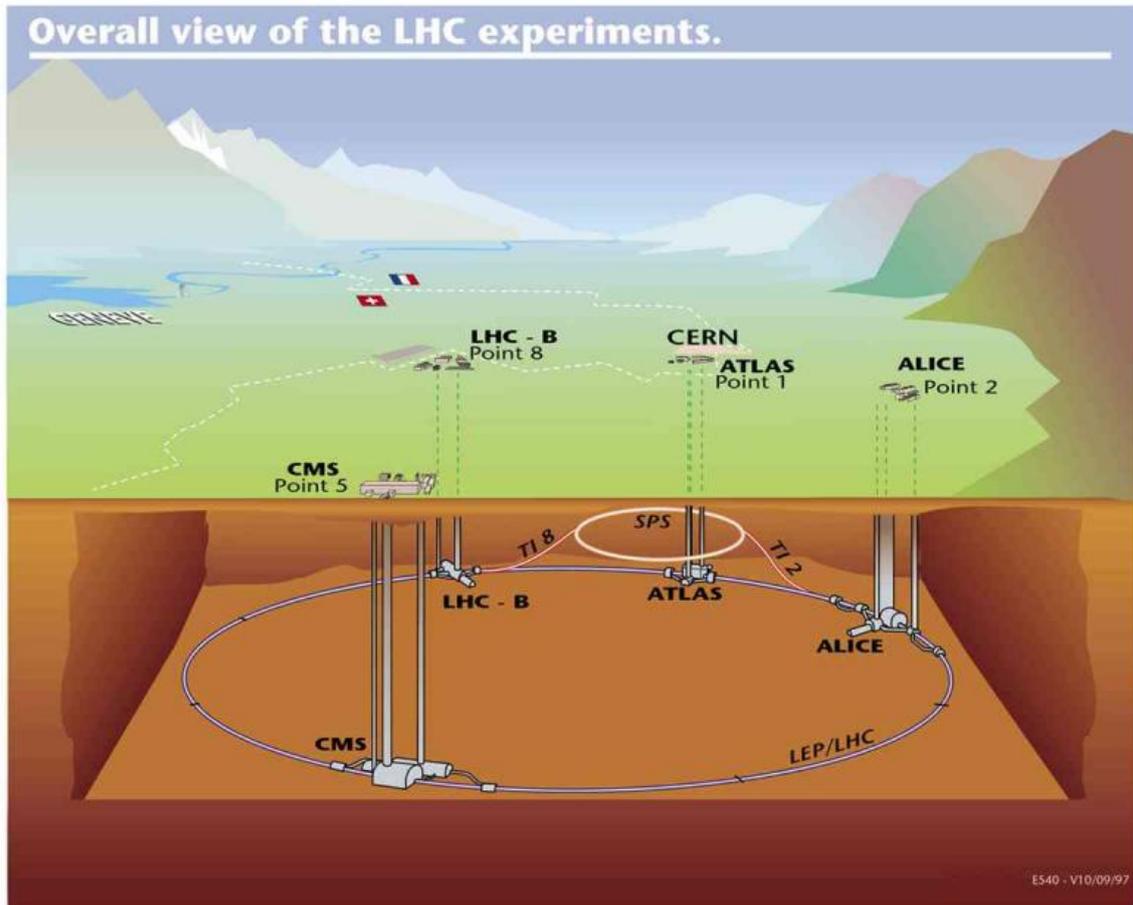
Figure 2.1: Geographical position of the LHC complex [27].

| (a) | (b) |
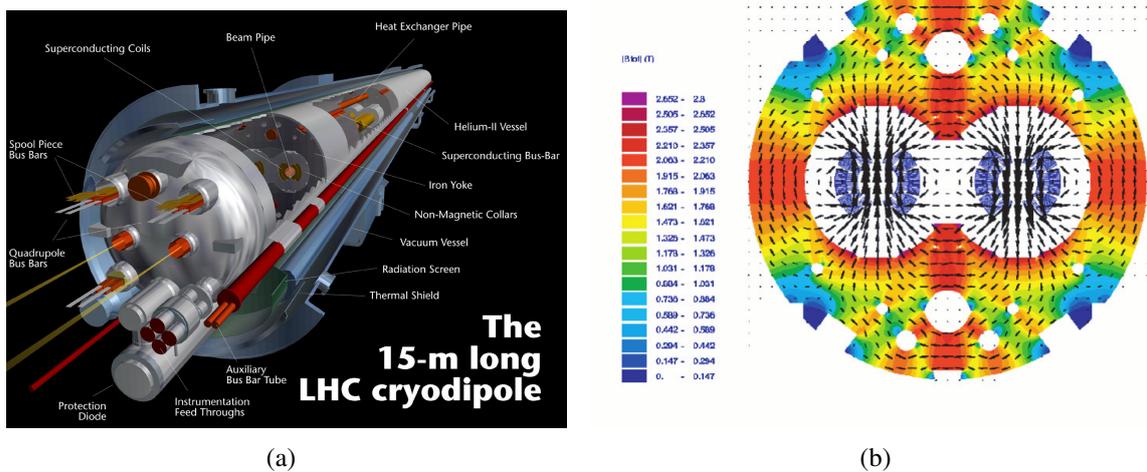
Figure 2.2: (a) transverse section of the LHC [28]. (b) representation of a magnet dipole of the LHC with the field lines [29].

In this thesis we studied proton-proton collisions collected in 2012 by the CMS detec-

Table 2.1: Beam parameters for the LHC in 2012 Run and nominal values.

Parameters of the beam	2012 Run	Nominal value
Energy of each proton	4.0 TeV	7.0 TeV
Number of protons per bunch	1.7×10^{11}	1.15×10^{11}
Maximum number of bunches per beam	1374	2808
Mean lifetime of the beam	15 h	15 h
Distance between two bunches	50 ns	25 ns
Maximum instantaneous luminosity	7.7×10^{33} cm^{-2}s^{-1}	1.0×10^{34} cm^{-2}s^{-1}

tor, which had an energy of 8.0 TeV in the CM. The LHC was designed to work with an instantaneous luminosity of 10^{34} cm^{-2}s^{-1} and a CM energy of 14 TeV, but due to some initial incidents, it was convenient to work in the first years with different parameters from the nominal ones. In Tab. 2.1 we show the values for some important parameters used in the Run of 2012 and their respective nominal values.
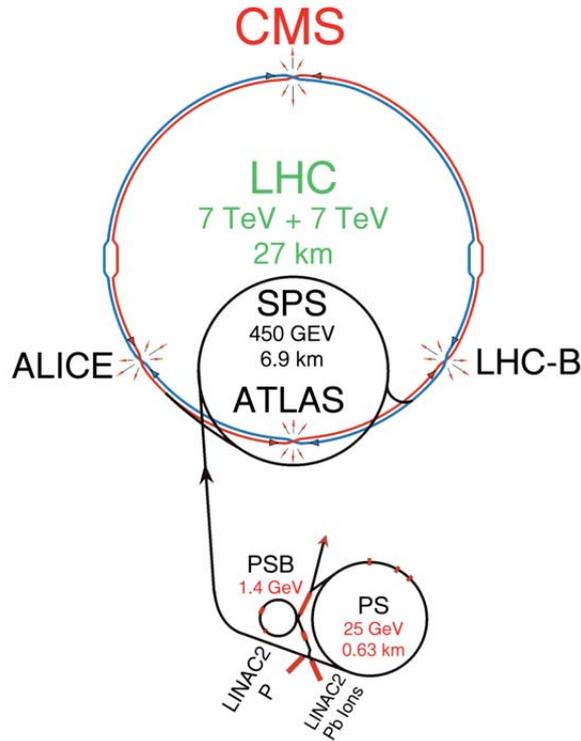
Figure 2.3: Proton injection chain in the LHC complex [28].

CMS Integrated Luminosity, pp, 2012, $\sqrt{s} = 8$ TeV

Data included from 2012-04-04 22:38 to 2012-12-16 20:50 UTC
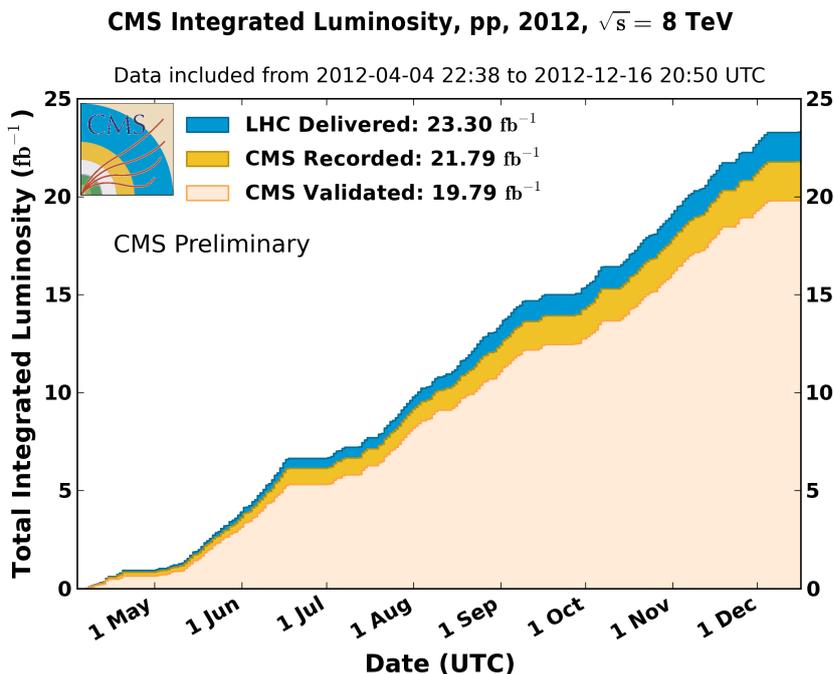
Figure 2.4: Integrated luminosity delivered by LHC, recorded and certified as good for physics by CMS during the Run of 2012 [30].

In the 8 TeV Run of 2012 the LHC delivered $\sim 23\text{fb}^{-1}$ of integrated luminosity. CMS recorded ~ 21.8 fb^{-1}, with ~ 19.8 fb^{-1} classified as good for physics analysis [30] (see Fig. 2.4). In this scenario we observed a mean of 20 primary vertices for a single bunch crossing (pile-up).

2.3 The Compact Muon Solenoid

The CMS [31] is a general purpose detector, which has as main goals:

- Search and study properties of the Higgs boson. Provide a better understanding of the spontaneous symmetry breaking mechanism in the electroweak sector of the SM.

- Search for new physics in the TeV scale like supersymmetry, extra dimensions, composite Higgs, etc...

In order to perform these studies, the CMS was built with a wide angular coverage and high resolution in trajectories and energy deposits measurements. The detector:

- Is compact: The material density is in general higher than the other LHC detectors.

- Has a dedicated system for muon trajectory measurements and trigger.

- Has as one of the main components a superconductor solenoid, which is responsible for an uniform magnetic field of 3.8 Tesla in its center. The uniformity of the magnetic field is increased by a steel return yoke external to the solenoid. In Fig. 2.5 it is shown the degree of uniformity of the magnetic field.
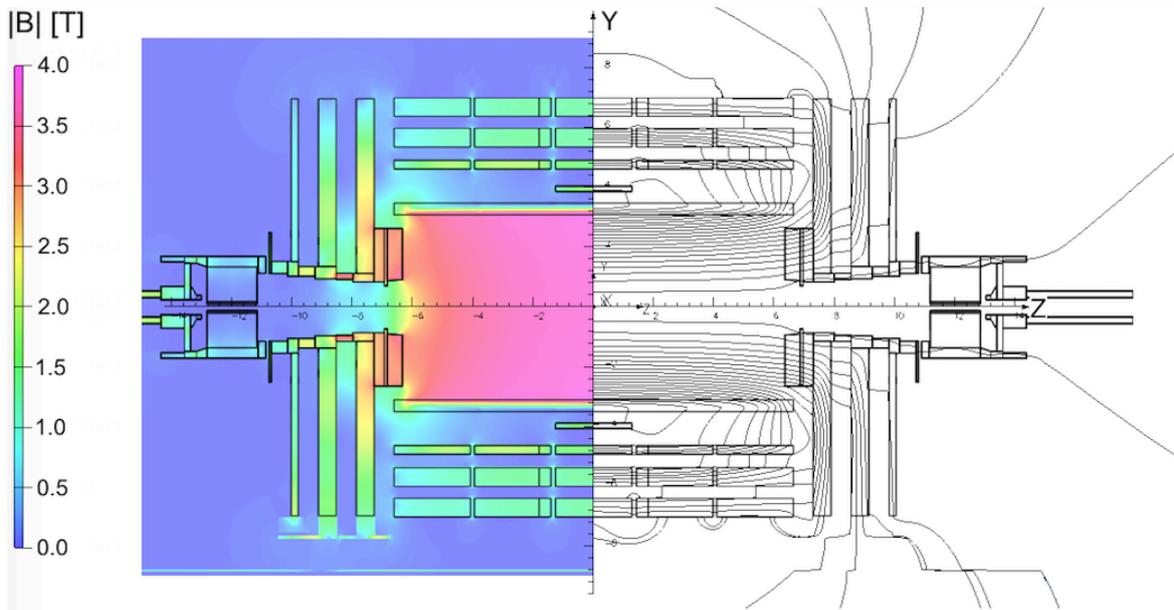
Figure 2.5: Field map of the magnetic field in the CMS detector [32].
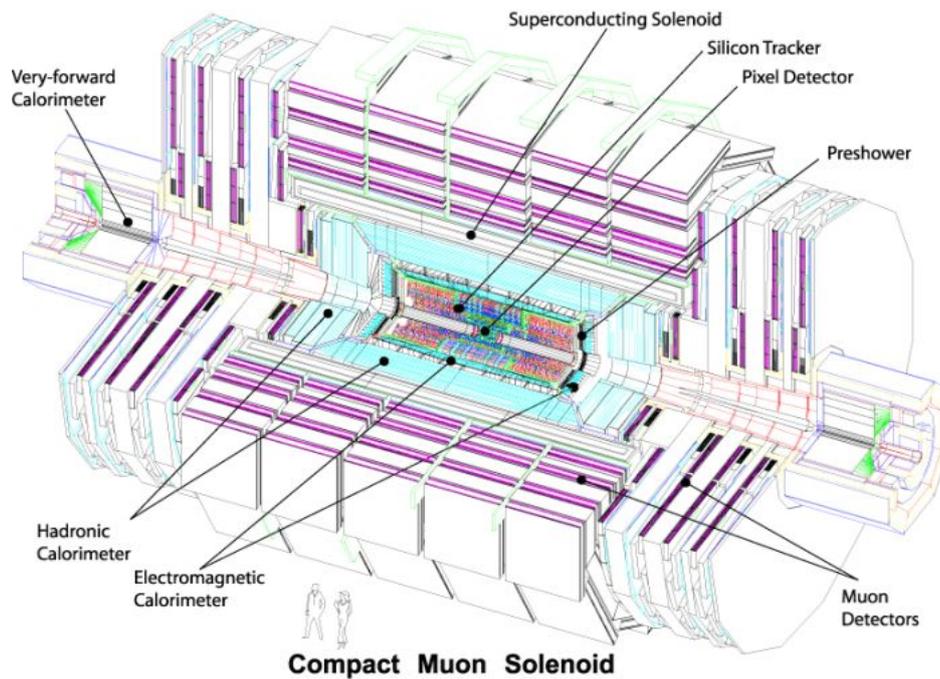
Figure 2.6: General view of the CMS detector [31].

With its cylindrical geometry, see Fig. 2.6, the CMS surrounds the beam pipe of the LHC in one of the interaction points, the P5, located in Cessy (France). The cylinder is 21.6 m long, has radius of 7.3 m, mass of 12500 t, and is divided in three regions: end-cap forward, end-cap backward, and barrel. The main detection systems are:

- Trajectory system.

- Calorimetry system.

- Superconducting solenoid with a steel return yoke.

- Muon system.
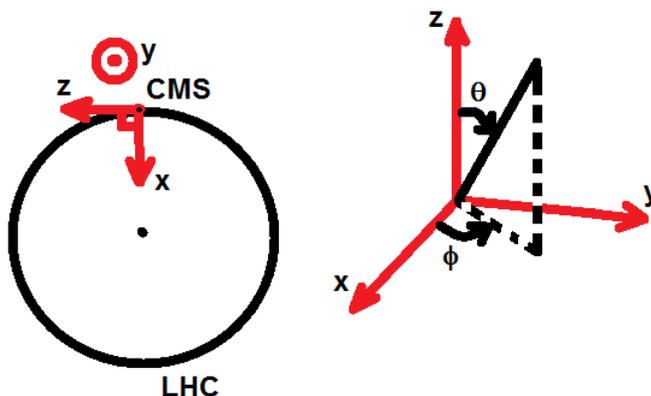
Figure 2.7: Scheme with the CMS coordinate system.

It is defined a cartesian coordinate system with origin in the center of the detector. The z axis is in the beam direction, the x axis points to the center of the LHC ring, and y points to the surface in the cavern[2]. It is common to use spherical coordinates (see Fig. 2.7) and a quantity called pseudorapidity (η) that is related with the θ coordinate in the spherical system. The pseudorapidity is obtained in the high-energy limit of the rapidity, defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_{\mathrm{L}}}{E - p_{\mathrm{L}}} \right),$$

where p_{L} is the longitudinal momentum of the particle with respect to the z axis and E is its energy. Therefore, in the high-energy limit ($E^2 = m^2 + \vec{p}^2 \rightarrow E \approx |\vec{p}|$), we have:

$$y \approx \eta = \frac{1}{2} \ln \left(\frac{|\vec{p}| + p_{\mathrm{L}}}{|\vec{p}| - p_{\mathrm{L}}} \right) = \frac{1}{2} \ln \left(\frac{|\vec{p}| + |\vec{p}| \cos \theta}{|\vec{p}| - |\vec{p}| \cos \theta} \right).$$

Using the trigonometric identities: $\cos^2(\theta/2) = (1 + \cos \theta)/2$ e $\sin^2(\theta/2) = (1 - \cos \theta)/2$, we have:

$$\eta = - \ln \left(\tan \left(\frac{\theta}{2} \right) \right), \tag{2.6}$$

i.e., from a quantity that depends on the energy and momentum of the particles we derived a quantity that only depends on the geometrical coordinate θ.

The main reason that we use the pseudorapidity is due to the fact that the rapidity and consequently the pseudorapidity are invariant under Lorentz boosts in the z direction aside

---

[2]In fact, the LHC ring is located in an inclined plane with 1.4% slope [33].

of one constant ($\eta' = \eta + c$). And based on this fact it is possible to define a quantity that is invariant under boosts in the z direction:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2},\tag{2.7}$$

which describes distances in the plane $\eta - \phi$.

These coordinates are used in the description of the CMS. In addition they are very useful in the description of the trajectories of the particles and energy deposits. In the following we will talk briefly about some subdetectors of CMS.

2.3.1   Tracking system

The tracking system (tracker) is responsible for the reconstruction of the trajectories of charged particles. Due to the magnetic field, the trajectories of the particles are bent allowing to estimate their momenta. The tracker has a very good spatial resolution, which allow a precise mapping of all the trajectories. It is possible to reconstruct not only the primary but also secondary vertices, like the ones from b hadron decays. In addition, the tracking system is used in the identification of particles isolated from extra hadronic activity from the same event.

The tracker is the sub detector closest to the beam pipe. It is important that it does not interfere much in the energy of the particles, so that the bias in the measurements of other sub detectors is reduced as much as possible. The tracker was built using two technologies: silicon pixel and strips sensors. Below we describe briefly these components.

- **Silicon pixel detector**: This is the component closer to the beam, see Fig. 2.8. Since it is very close to the interaction point, the pixel detector needs to deal with an environment with very high radiation dose[3]. The pixel detector is made up by three cylindrical layers with radii 4.4, 7.3, and 10.2 cm with respect to the beam and two disks in each extremity, see Fig. 2.9. The detector has a total of 66 M silicon sensors (cells) with area of 100 μm $\times$ 150 μm each, resulting in an active area of approximately 1.1 m^2. When a particle passes through a sensible region of the detector, it ionizes the material creating electron-hole pairs (~ 110 $pairs/\mu$m). These charge carriers are separated in opposite directions due to an electric field applied in the silicon. The induced charges are then collected and amplified by a low-noise electronic system. Position measurements in the pixel detector have a spatial resolution of order 10 μm in the plane xy and 20 μm in the z direction. With this very good resolution, it is possible to reconstruct primary and secondary vertices in an event. The pixel detector is used in the measurement of the luminosity using a pixel cluster counting technique [34].

- **Silicon strip tracker**: This component is divided in two regions: inner silicon strip tracker with internal radius of 20 cm and outer silicon strip tracker with internal radius of 55 cm. The inner silicon strip tracker is composed by two sub detectors, the tracker inner barrel (TIB) and tracker inner disks (TID), which are respectively made up by four cylindrical layers and three disks in each extremity. The outer silicon strip tracker is composed by two sub detectors, the tracker outer barrel (TOB) and tracker end caps (TEC), which are respectively made up by six cylindrical layers and nine disks (see

---

[3]For example, in a distance of 8 cm with respect to the interaction point, we have around 10^6 particles/cm^{2}s.

Fig. 2.8). Increasing the radial distance from the interaction point decreases the particle flux in the detector so that we need a better resolution for the internal components. The separation between the strips is lower in the TIB and TID resulting in a spatial resolution of approximately $23-25 \ \mu$m in comparison with a resolution of $35-53 \ \mu$m in TOB/TEC.
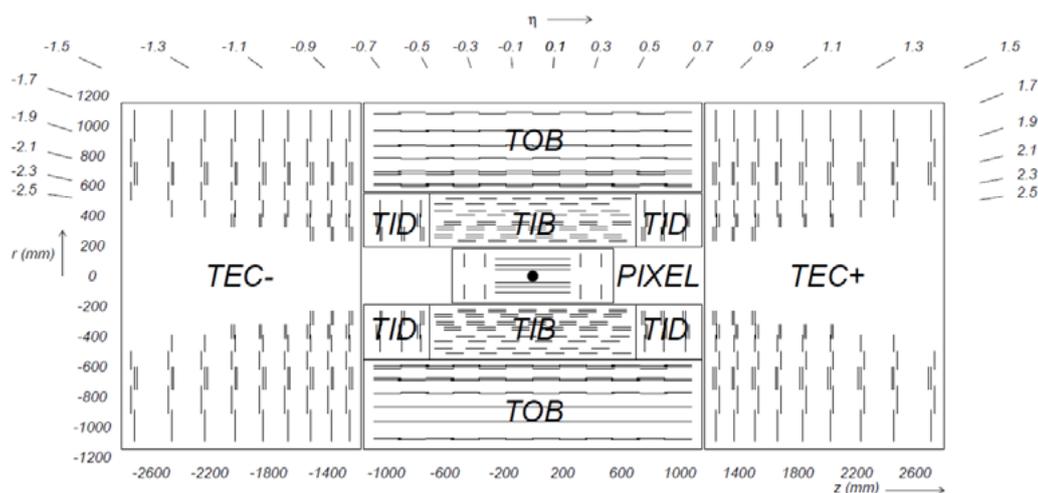
Figure 2.8: zy view of the tracking system [31].

(a)   (b)
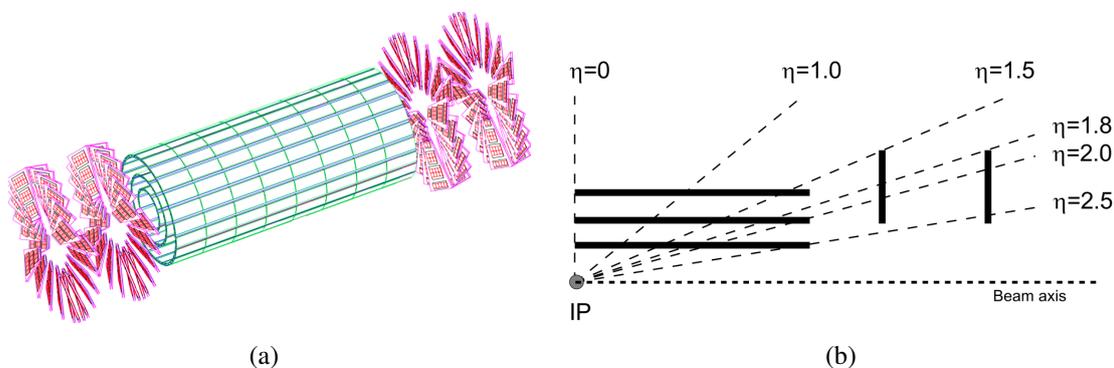
Figure 2.9: (a) sketch of the pixel detector. (b) one quadrant in the zy plane of the pixel detector [31].

In Fig. 2.10 we show the transverse momentum resolution and spatial resolution in the xy plane and z direction. All these quantities are studied as a function of muon transverse momentum equal to 1, 10, and 100 GeV and $0 < |\eta| < 2.4$ [35].

2.3.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) measures the energy of high-energy electrons and photons, which interact with the ECAL material until being absorbed. They interact with a scintillator material that emits photons in the optical spectrum, which are detected by photodetectors allowing an estimate of the energy deposited by the particles. The ECAL is mainly
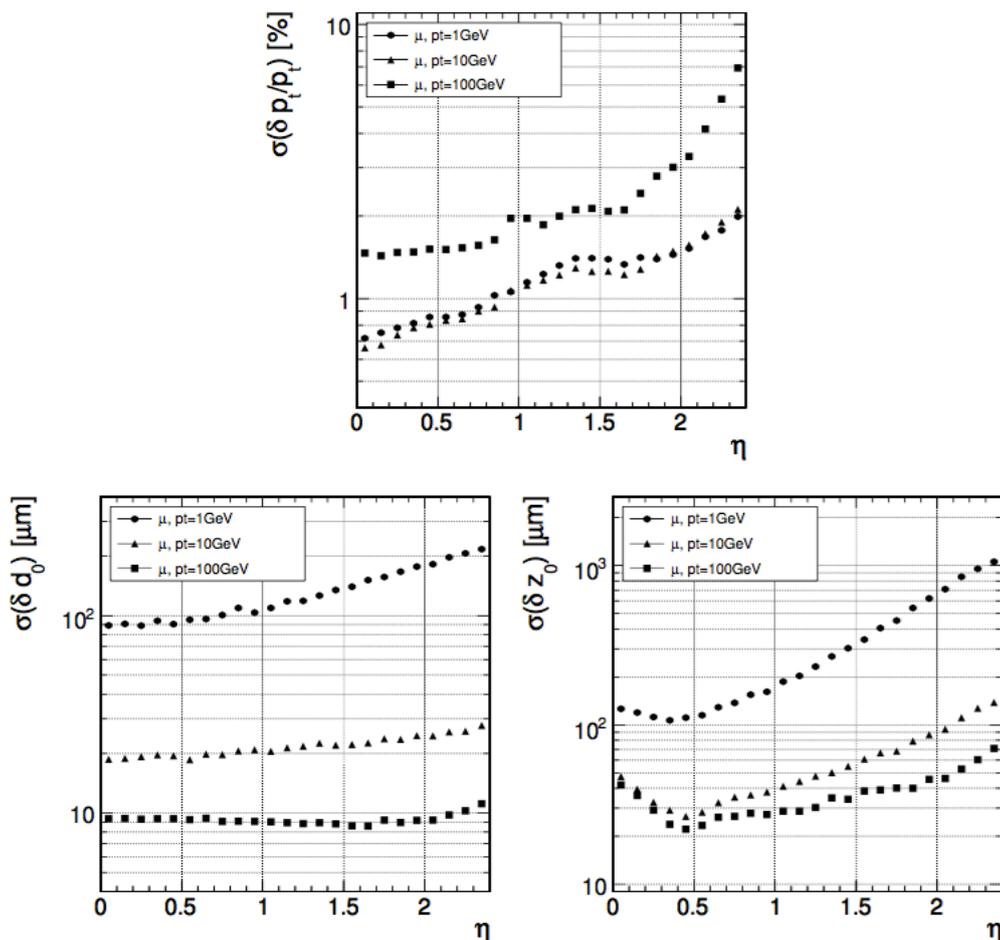
Figure 2.10: Transverse momentum and spatial resolution in single muon events with transverse momenta of 1, 10, and 100 GeV as a function of $|\eta|$. Upper: transverse momentum resolution. Lower left: spatial resolution in xy plane. Lower right: spacial resolution in z direction [31].

made up by $\mathrm{PbWO_4}$, which is a fast response scintillator material with 80% of the optical photons emitted within 25 ns, allowing a separation between the bunch crossings. The atoms that constitute this material have a huge concentration of electrons (high atomic number), allowing higher electromagnetic interaction with the particles. The ECAL is composed by three sub detectors (see Fig. 2.11):

- **ECAL Barrel (EB)**: Composed by 61200 $\mathrm{PbWO_4}$ crystals, the EB has a cylindrical shape with pseudorapidity coverage of $|\eta|< 1.479$.

- **ECAL Preshower (ES)**: Composed by two lead disks followed by silicon detectors similar to the ones from the tracking system, the ES distinguishes π^0, which decay in two photons very close in space, from a single high-energy photon. When passing through the lead disks, the photons create an electromagnetic shower (composed mainly by electron-positron pairs), which is detected by the silicon sensors. The ES is located in a η region where these signatures frequently occur, $1.653 < |\eta|< 2.6$. This sub detector has an important role in analyses with the Higgs boson in the diphoton channel.

- **ECAL Endcaps (EE)**: The EE in conjunction with the ES compose the extremities of the ECAL. The EE is composed by 3662 $PbWO_4$ crystals, covering a η region of $1.479 < |\eta| < 3.0$.

Several components spoil the measurement of the deposited energy in ECAL: calibration uncertainties, noises from electronic components, pile-up, and others. The resolution in the energy measurement in ECAL is described by the following expression [31]:

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

where E is the energy in GeV and the values of each term were obtained with a fit to data from an electron test beam for $0 < E < 250$ GeV.

2.3.3 Hadron calorimeter

In general the hadrons have mass (order of 1 GeV) much higher than the electron's mass (around 0.5 MeV). At typical energies of the LHC, the hadrons are produced with energies of the order of several GeV. These features make useful the implementation of a sub detector composed by materials with atoms with very high atomic number, where these hadrons will predominantly interact via strong interaction, complementing the measurements from the electromagnetic calorimeter. The hadron calorimeter (HCAL) is located externally to the ECAL and has the following components (see Fig. 2.12):

- **HCAL Barrel (HB)**: This sub detector is composed by brass plates that absorbe the particles. These plates are alternated with scintillator layers that generate signals associated to the energy deposits. The HB is located in a region between ECAL Barrel and the superconductor solenoid, with a coverage of $|\eta| < 1.3$.

- **HCAL Outer (HO)**: With a similar coverage as in the HB, the HO is in a region between the superconductor solenoid and the steel return yoke. The main function of the HO is to detect and absorb the very energetic hadron showering that passes through the whole HB. The HO is composed by five rings of scintillator material. In the central region, where the particles fly through a smaller distance in the HB (smaller absorption power), two layers of scintillator material are filled by an iron absorber.

- **HCAL Endcaps (HE)**: This component has a similar composition and function as HB, but it covers the $1.3 < |\eta| < 3.0$ region.

- **HCAL Forward (HF)**: This component covers a region of $2.9 < |\eta| < 5.2$, which is not covered by any other CMS sub detector. Due to its very close position with respect to the beam pipe, the HF is composed mainly by quartz fiber, resistant to high radiation, combined with a steel absorber.

With this very high η coverage, the HCAL is a key component in the estimate of the transverse missing energy, which is an important object in many analyses in the collaboration. The expected resolution in the energy measurement of HCAL is approximated by [36]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{120\%}{\sqrt{E}}\right)^2 + (5\%)^2, \tag{2.9}$$

(a)

(b)
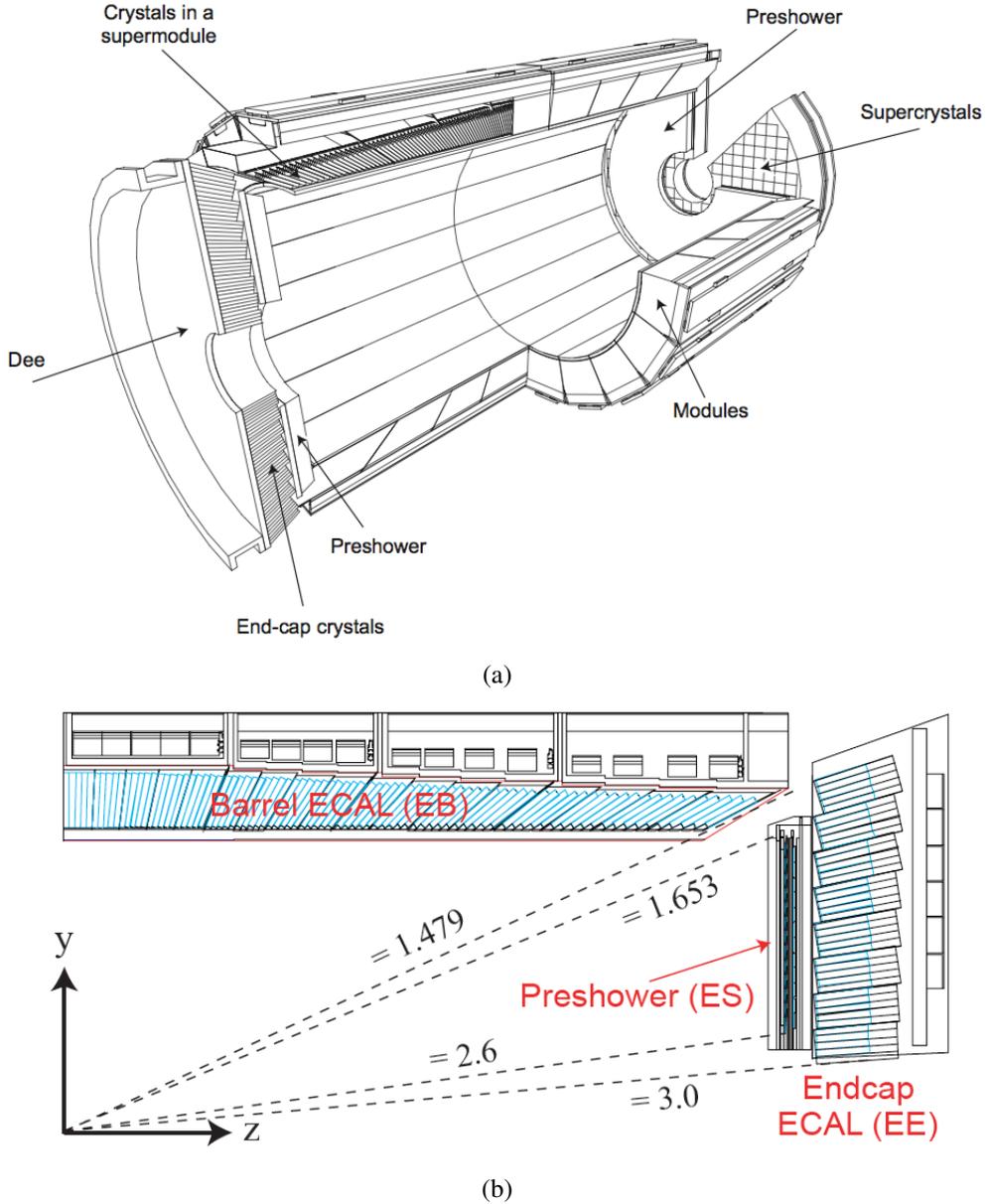
Figure 2.11: (a) sketch of the ECAL detector with their main components [31]. (b) transverse section of the ECAL showing one quadrant in the plane yz and pointing the η coverage [35].

where E is in GeV and the values of each term were obtained with a fit to data from test beam for particles' energy in the range $30 < E < 1000$ GeV.

2.3.4 Superconducting magnet

This component is responsible for a magnetic field of 3.8 T in the center of CMS. With this field magnitude it is possible to estimate the electric charge and momentum of a charged particle even at very high energy (order of TeV). The solenoid is located between the HB and HO and it has a cylindrical shape with diameter of 6 m and length of 12.5 m. Outside the solenoid, giving mechanical support to the muon system, is located the steel return yoke, which is responsible for the homogeneity of the magnetic field outside the solenoid. In
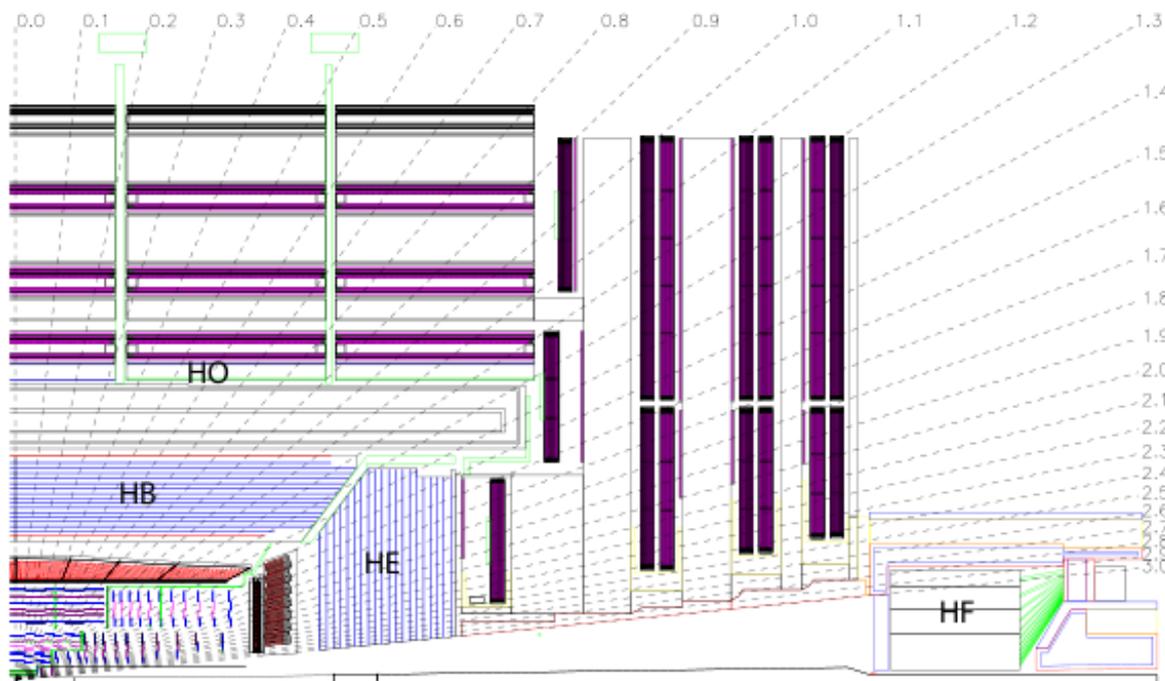
Figure 2.12: Sketch of the hadron calorimeter of CMS pointing its main components and the η coverage [31].

the barrel region this system is composed by five "rings" with a transverse section with a dodecagon shape. Each "ring" has three layers of steel and between these layers are located the muon system, see Fig. 2.13. In the end-caps of the detector, the return yoke system is composed by six disks, three in each extremity (see Fig. 2.14).

2.3.5 Muon system

The muons are object of study in many analyses in the LHC. Due to this importance, it was developed a system of detectors specialized in complementing the position and momentum measurements from the tracker system, providing a more precise muon identification. Due to its relatively high mass and the fact that it does not interact strongly, the muon passes through all the material of the tracker and calorimeters without losing a considerable amount of its energy. The muon system is located outside the calorimeters between the layers of the steel return yoke in the barrel and end-caps, see Fig. 2.15. The sub detectors were built using three technologies: Drift Tubes (DT), Resistive Plate Chambers (RPC), and Cathode Strip Chambers (CSC).

- **Drift Tubes**: With a coverage of $|\eta| < 1.2$, the DT system has a spatial resolution of approximately $100 \ \mu$m. In total there are 250 chambers, which are located in 4 concentric stations inside the steel return yoke of CMS. These stations are divided into 5 wheels with 12 sectors each. Each sector covers $\sim 30°$ in ϕ, see Fig. 2.16. The DT chambers are composed by 12 silver layers grouped in three sets of four layers. Each set of layers has 60 tubes with transverse section of width of 4.2 cm and height of 1.3 cm. The tubes are filled with a mixture of argon (Ar) and carbon dioxide (CO_2),

16

Figure 2.13: Superconducting solenoid surrounded by one of the wheels of the steel return yoke in the barrel region of CMS [31].

see Fig. 2.17. When a charged particle passes through the gas, it ionizes the atoms. In each tube there is a wire (anode) that together with a cathode create an electric field that accelerates the electron from the gas ionization. With a gain in energy due to the electric field, the electron has energy enough to ionize other atoms in the gas resulting in an avalanche effect, which is recorded by the electronic system. Since the mean velocity of the electrons in the gas is a known parameter and that is possible to obtain the mean time that the electrons pass through the gas untill reach the anode, it is possible to obtain the position of the charged particle.

- **Cathode Strip Chambers**: This component uses a similar approach as the Drift Tubes, but covering the region of $0.9 < |\eta| < 2.4$ and with a spatial resolution of $50 \ \mu m$. Since they are located in the end-caps, the CSC were designed in order to work more efficiently in a region with higher magnetic fields. The sets of CSC make up four disks for each end-cap, which are composed by 18 or 36 chambers in trapezium shape (in total 468 chambers), see Fig. 2.18(a). Each chamber is composed by seven layers with cathode strips and six layers filled with a mixture of gases (40% of Ar, 50% of CO_2, and 10% of CF_4) and with many wires (anode) as showed in Fig. 2.18(b).

- **Resistive Plate Chambers**: This sub detector is located in both barrel and end-caps. It helps the other components (DT and CSC) in the trajectory reconstruction and provides

17
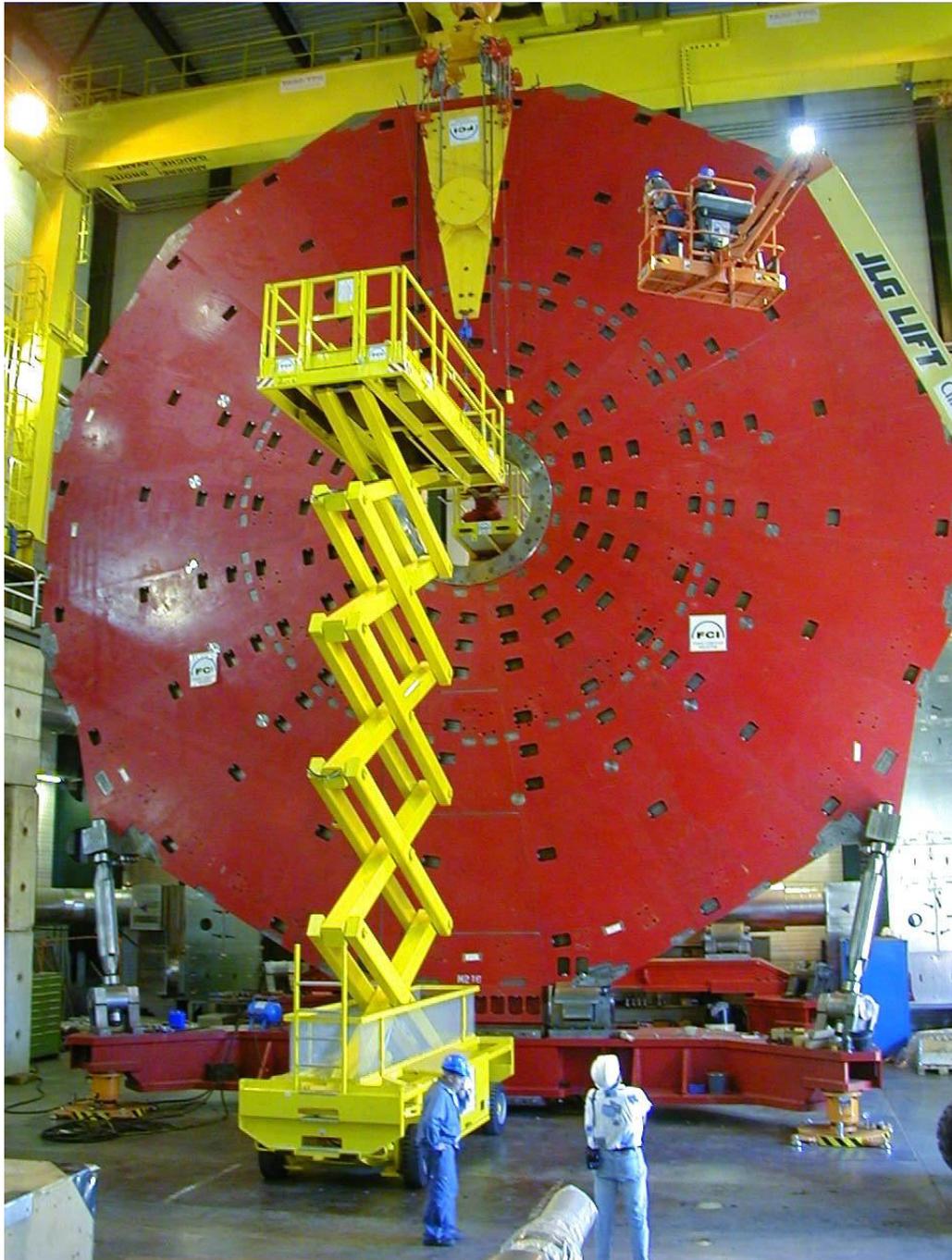
Figure 2.14: Photo of one return yoke disk [31].

a trigger system. There are six layers of RPC in the barrel and three in each end-cap. The RPC is composed by two bakelite plates, which are parallel and separated by a small distance. Between these two plates it has gas and it is fixed a high electric potential difference, see Fig. 2.19. A charged particle passes trough the gas and ionizes some atoms producing an avalanche inside the chamber. Combined with the plates there are some silver strips responsible by the read-out of the signal created by the avalanche of electrons. Due to the small distance between the bakelite plates, we have a very fast response to the passage of the charged particle, order of 5 ns, which is

enough for a trigger system given that the time between each collision at the nominal luminosity is about 25 ns.
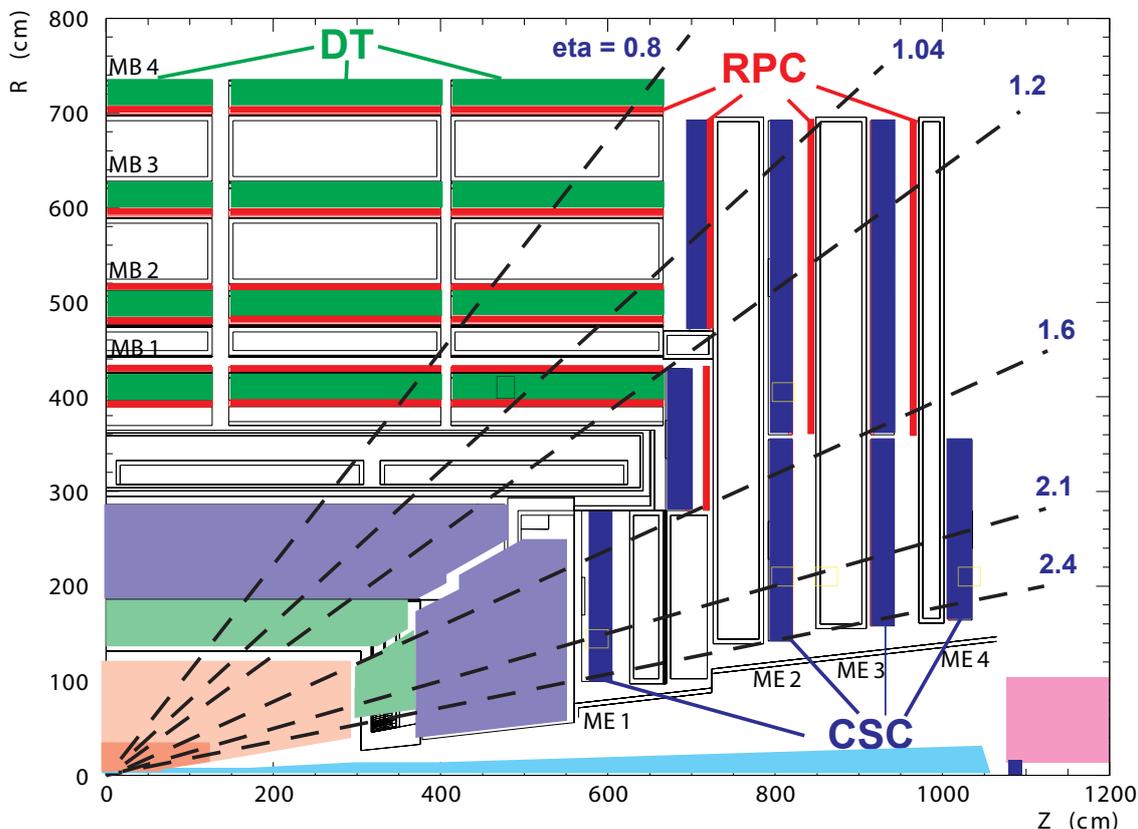
Figure 2.15: Sketch of the muon system in one quadrant in the yz plane [31].

In Fig. 2.20 we show curves describing the muon transverse momemtum (p_T) resolution, looking all the covered η region (except the crack region between $0.8 < |\eta| < 1.2$, see Fig. 2.15). We present the resolution by the muon system only, the tracking system, and the combination of the two systems. For the low p_T region, below ~ 100 GeV, the muon system provides much worse resolution than the tracking system. For higher momenta, the resolution for the tracker degrades faster than the muon system, so that they have comparable resolutions for $p_T > 2$ TeV in the central region, $0 < |\eta| < 0.8$.

2.3.6 Trigger system

The trigger system [37] was developed in order to select fastly and precisely "interesting events" (with characteristics compatible with the physics analyses program of the LHC), and to reduce considerably the amount of data to be stored. In the nominal LHC luminosity, we expect around 20 proton-proton collisions [4] for each bunch crossing, which occur in 25 ns intervals. This rate of events translates in an amount of data that, with the current technology, is not viable to be recorded. Therefore, it is necessary a trigger system with fast response

---

[4]It is important to remember that we have other parameters in heavy-ion collisions, which imply in different trigger strategies for these studies.
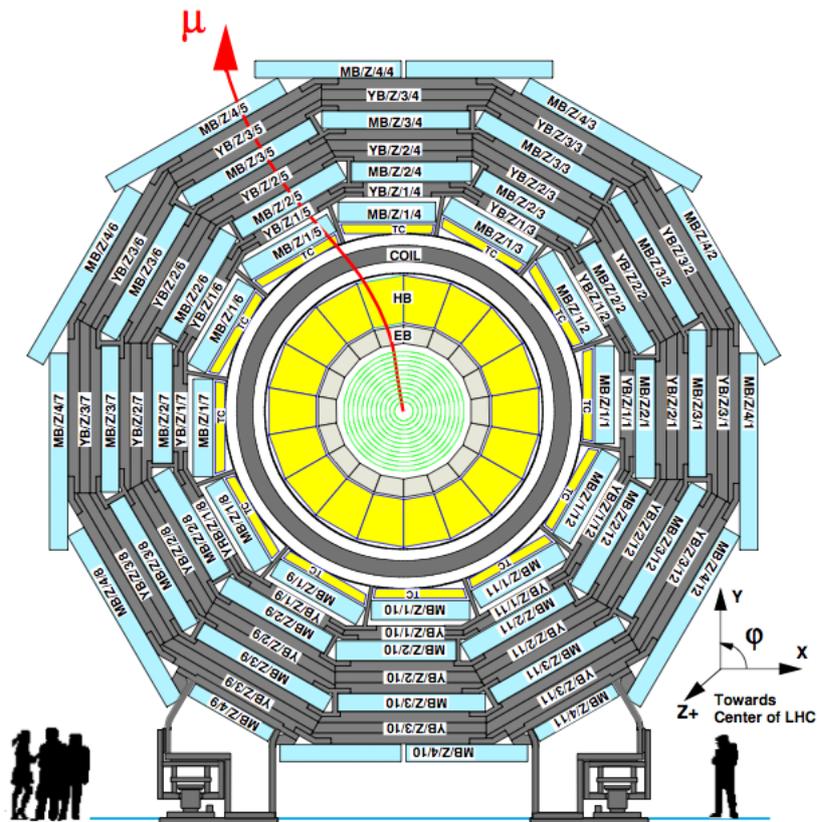
Figure 2.16: Sketch of the transverse section of one of the wheels of CMS barrel. The DT chambers are labeled by "MB" [31].
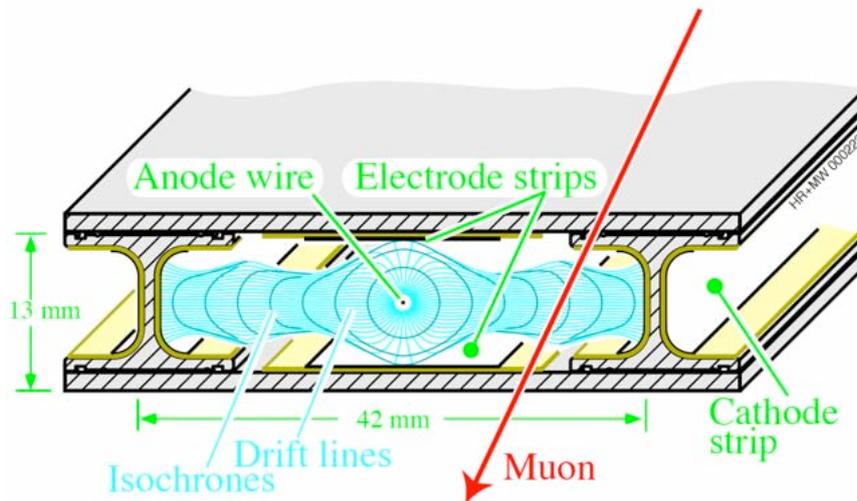
Figure 2.17: Drift Tube transverse section [31].

and efficient algorithms to select the required events. It is possible to divide the CMS trigger system in two big categories:

- **Level 1 Trigger (L1 Trigger)**: Currently most of this step is performed in the custom electronics of CMS. It is divided in some sectors that reconstruct and classify
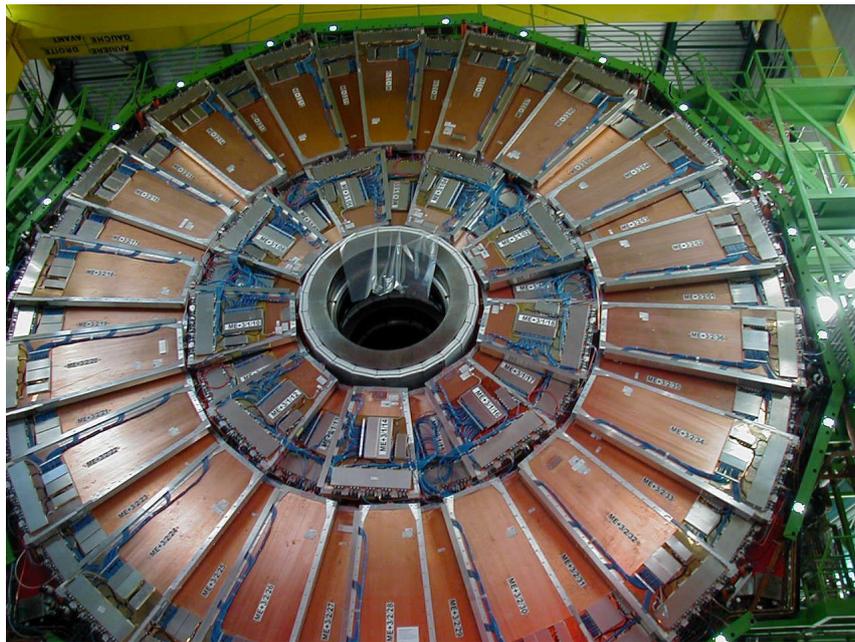
(a)
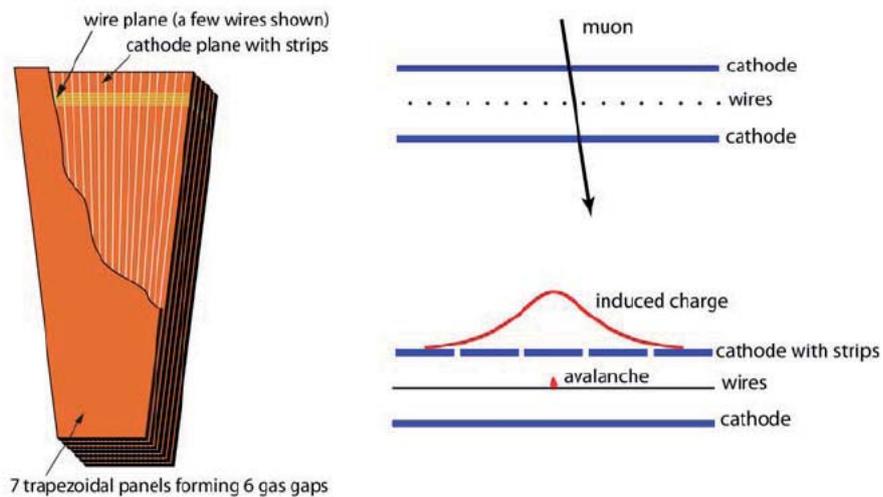
(b)

Figure 2.18: (a) disks with the Cathode Strip Chambers. An example of chamber is showed in (b), together with a scheme showing how the anode and cathode are organized inside the chambers [31].

some physical objects (like muons, electrons, jets, etc...) using informations from the calorimetry and muon systems. This analysis is passed to a final component that decides to save or not the event. This last decision need to be done in a time interval of around 3.2 μs after each collision because the data are stored in a temporary area and are overwritten after this time interval. In the nominal luminosity, the L1 Trigger reduces the rate of events from 40 MHz to 100 kHz. Each event with a mean of 20 primary vertices as described above.
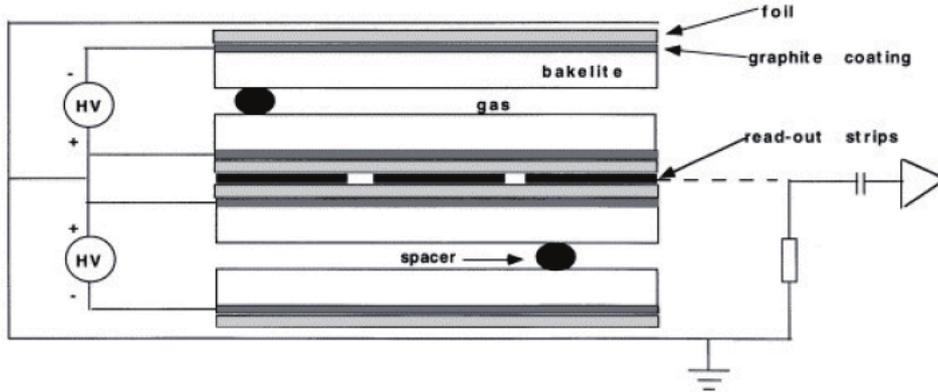
Figure 2.19: Sketch of one Resistive Plate Chamber [31].
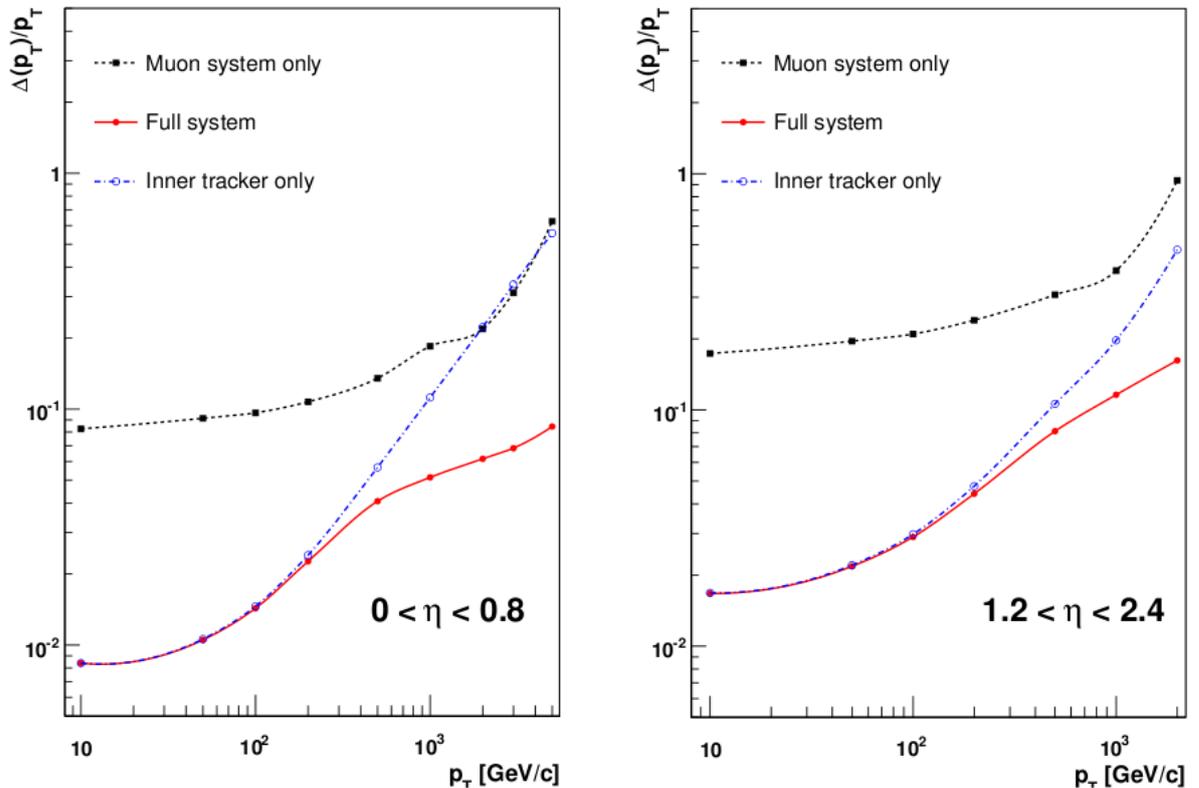
Figure 2.20: Muon transverse momentum resolution. We present this quantity for three situations: considering only the muon system, only the tracking system, and both systems. In the left plot we show the resolution for $0 < |\eta| < 0.8$ and in the right plot for $1.2 < |\eta| < 2.4$ [31].

- **High Level Trigger (HLT)**: The events selected by the L1 trigger are transferred to a computer farm, where a more detailed event selection is done. The HLT does a similar reconstruction and identification as in the physics analysis level, which uses more specific criteria for the objects like requirements on the transverse momentum and energy of jets, muon/electron isolation, etc... In this layer of the trigger system the events that pass at least one pre-defined set of requirements (trigger path) are stored,

which reduces the rate of events from $100\,\mathrm{kHz}$ to $100\,\mathrm{Hz}$ ($100\,\mathrm{MB/s}$ to be permanently stored).

2.4 General view of an event and analysis framework

When we have a proton bunch crossing in the center of the detector, many interactions take place among these protons. The interactions that we are interested in (and to be selected by most of the trigger paths) are the ones in which the proton constituents (partons) interact producing other particles with transverse momentum of tens of GeV (hard process). In the interactions, these partons can exchange, for example, gluons, Z/W bosons, photons. These bosons result in more stable particles that are detected by CMS. In the case of production of high-p_{T} quarks/gluons (or hadronic decay of the bosons), these final state partons go through a serie of intermediate processes like radiation emission, combination into color singlets (hadrons), hadron decay. Since the mother partons have a high-momentum, in general the stable final state particles are detected in localized regions in the detector. These complex objects are called jets.
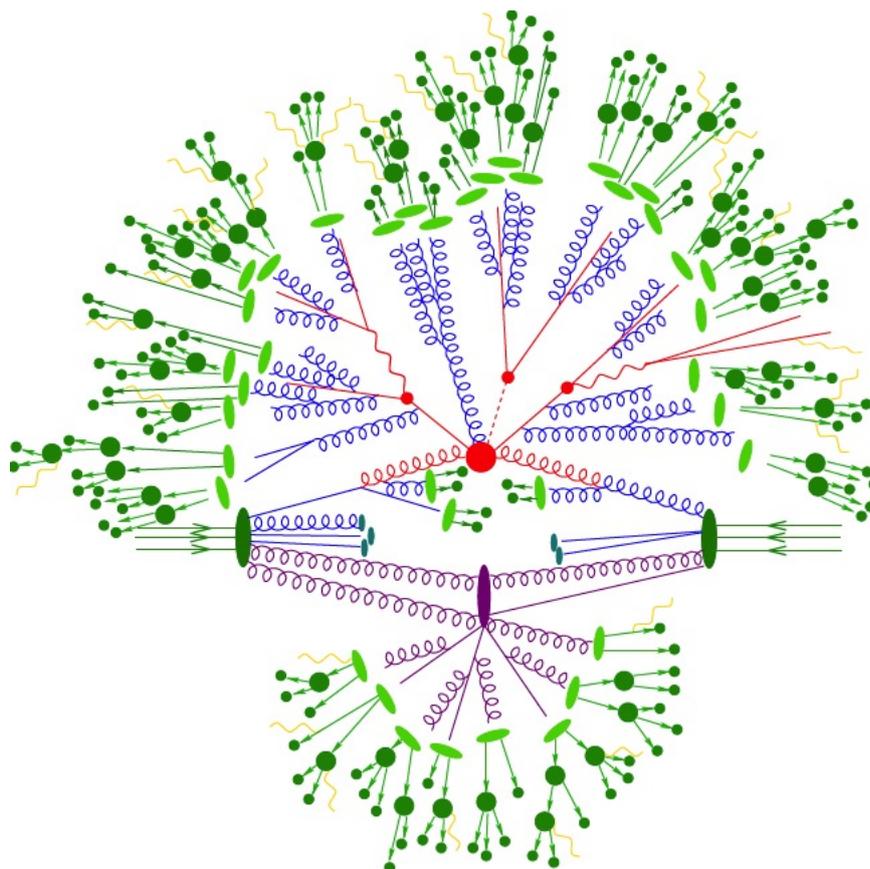
Figure 2.21: Pictorial representation of a Monte Carlo simulation of a hadron collision [38]. The hard process is described by the red color. Secondary interactions are in purple. In blue is represented the radiations from the parton showering. In green are described the hadrons after the hadronization step and dark green the decay of unstable hadrons. Finally, in yellow we have QED Bremsstrahlung.

Before the hard interaction, the partons radiate other particles like gluons and photons, called initial-state radiation (ISR). Just after the hard process, it is possible that the final state particles radiate, which is known as final-state radiation (FSR). It is observed the production of particles coming from secondary interactions among the other partons constituents of the protons that originated the hard interaction (underlying-event). In Fig. 2.21 we show how Sherpa [38], a Monte Carlo generator, models these processes, including the non-perturbative hadronization and unstable hadron decay steps.

Due to the high-luminosity of LHC, it is possible to have more than one proton-proton hard interaction in one single bunch crossing, pile-up. As described in Sec. 2.2 we had a mean of 20 hard collisions per bunch crossing in the run of 2012. All these processes can interfere in the particle reconstruction for each event. For example, radiation and hadronization tend to smear out the energy of jets from the hard process ("splash-out" effect), while the underlying-event and pile-up tend to add energy to the hard interaction ("splash-in" effect) [39].

The stable particles interact with the sensible material in the detector generating signals, called hits. These signals are measurements of the position of the particles when passing through the sensitive volume of the CMS. An algorithm of pattern recognition uses these hits for the reconstruction of the trajectories, which are combined with energy deposits in the calorimeters providing more details about the physics objects. More complex objects like jets and taus are reconstructed using these primary objects. This "offline" reconstruction is done using an analysis framework called CMSSW (**CMS S**oft**W**are) [40], which has a modular structure, where each module written in C++ is controlled by configuration files written in Python. The CMSSW is also used to perform the analysis selection itself, it is possible to access several information about the events (particles' momentum, charge, spatial position, etc...) using dedicated analysis modules. These last information are directly related with theoretical particle physics frameworks. In the next chapters some theoretical models and their phenomenology are briefly described.

Chapter 3

The standard model and some possible extensions

3.1 The standard model of particle physics

The standard model (SM) [41, 42, 43, 44, 45, 46] is nowadays the main paradigm to describe elementary particles and their interactions. This model has been tested successfully by many experiments [1, 2, 47, 48, 49, 50, 51, 52, 53, 54]. It uses quantum field gauge theories that describe the fundamental interactions among the particles fixing local symmetries in the corresponding Lagrangians. In particular, the current formulation of the SM is built defining the most general renormalizable Lagrangian with a set of particles described below and invariant under symmetries associated to the Poincaré and gauge $\mathrm{SU(3)} \times \mathrm{SU(2)} \times \mathrm{U(1)}$ groups[1].

Many experimental results gave guidance to the development of the SM[2]. For instance, the observation of parity violation by the weak interaction defined the structure of the weak current in the SM.

The masses of the Z^0 (from now on Z) and $\mathrm{W}^\pm$ bosons are generated by the Higgs mechanism. In this procedure, new degrees of freedom are introduced in the theory, represented by massless scalar particles (Goldstone bosons), which are "transferred" in a consistent way to become new degrees of freedom for these massive vector bosons. A consequence of this mechanism is the appearance of an additional massive scalar particle, the Higgs boson, which has a crucial role in the perturbative unitarization of the theory. The masses of the fermions are generated by Yukawa couplings between the fermions and the Higgs boson.

The fermions are the matter fields defined by the fundamental representations of the gauge groups. They are divided into quarks and leptons, which are classified in three generations or families. The elementary bosons are associated to the interactions in the model. In Figs. 3.1 and 3.2 we show, respectively, the elementary particles and the interactions among them in the framework of the SM. In the next subsections we describe briefly some properties of each elementary particle that is useful for the understanding of the analysis procedure.

---

[1]The gravitational interaction is not included in the SM. In fact, for the energies tested until now (at order of TeV), the introduction of quantum effects from gravity is not necessary since it is expected to be much weaker than the other known interactions. It is worth to mention that nowadays we do not have a standard theoretical framework to describe quantum-gravity in the energy regime (i.e. Planck scale, $M_\mathrm{P} \sim 10^{19}$ GeV) that the gravitational interaction is expected to be comparable with the other known fundamental interactions.

[2]See Ref. [55] for a historical summary.
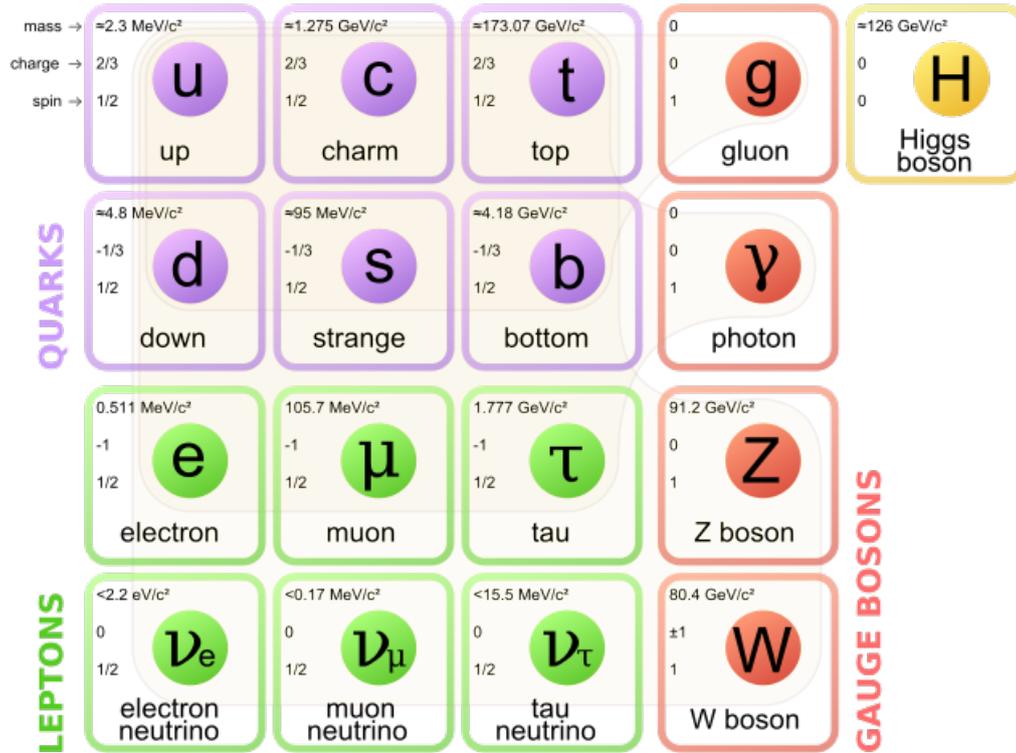
Figure 3.1: Scheme describing the elementary particles in the standard model. We present values for the mass, charge, and spin of each component.

3.1.1 Leptons

Leptons are fermions that interact both electromagnetically and weakly, with the exception of the neutrinos that only interact weakly.

- **Electron** (e): The lightest charged lepton in the SM is the electron with a mass of around 0.5 MeV. Since it is charged and stable, the electron production in the collisions results in a track in the silicon detector and a considerable amount of energy deposit in the electromagnetic calorimeter of CMS.

- **Muon** (μ): With a mass of around 200 times bigger than the electron mass (≈ 106 MeV), the muon is an unstable lepton with a mean lifetime of $t \approx 2~\mu$s, which corresponds to $ct \approx 600$ m. With this mass and not interacting strongly, the μ does not interact much with the material of the calorimeters, instead its reconstruction and identification are mainly done by the tracker and muon systems.

- **Tau** (τ): This is the heavier lepton observed to date, with a mass of approximately 1.8 GeV and a mean lifetime much smaller than the muon, $t \approx 3 \times 10^{-13}$ s ($ct \approx 90~\mu$m). The τ lepton decays before interacting with the tracker, therefore only its decay products are detected in CMS, i.e., pions (π^0, $\pi^{\pm}$), kaons (K^0, $K^{\pm}$), electrons, muons, etc... In Tab. 3.1 we present the dominant τ decay processes with the respective branching fractions, and intermediate vector meson resonances production.

- **Neutrino** (ν): For each $e/\mu/\tau$ described above exist a corresponding neutrino, which interact only through the weak force. In the simplest formulation of the standard model
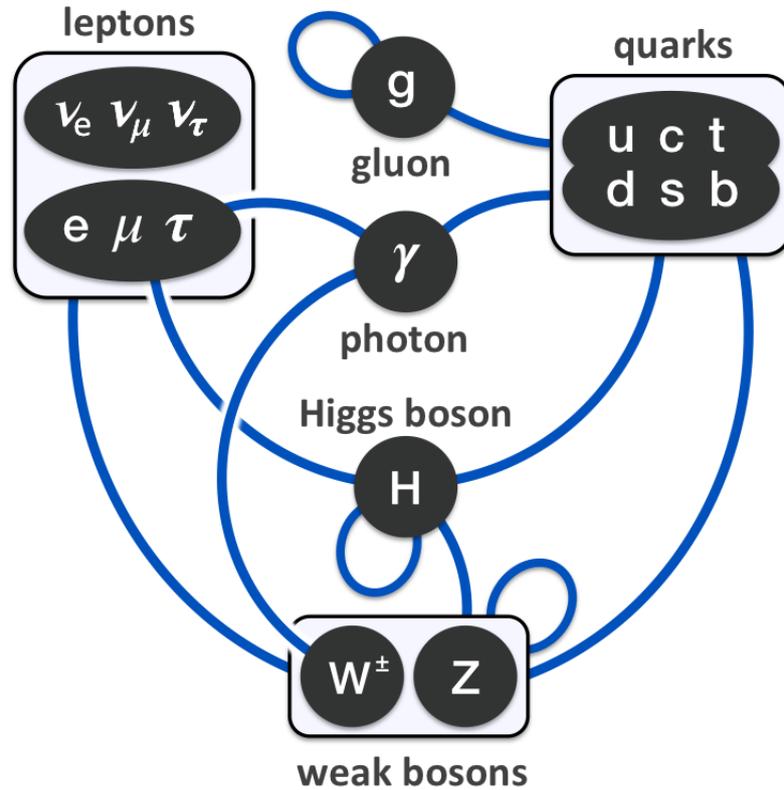
Figure 3.2: Scheme describing the interactions among the elementary particles in the standard model. For example, the lines connecting the leptons with gauge bosons mean that the leptons can interact exchanging these bosons. The closed lines refer to self couplings.

they are massless. Nowadays, from neutrino experiments we have some facts (neutrino flavor oscillation [56]) indicating that neutrinos have mass. The neutrinos do not interact with any component of the CMS detector. They are included in the transverse missing energy, which is defined as the negative of the vector sum in the transverse plane (with respect to the beam line) of all visible particles.

3.1.2 Electroweak vector bosons

- **Photon** (γ): The photons are responsible for the electromagnetic interaction. They are massless and electrically neutral, interacting mainly with the material of the electromagnetic calorimeter creating pairs of electron-positron.

- **Charged bosons** ($W^{\pm}$): Both have the same mass ($m_{W^{\pm}} \approx 81.4$ GeV) and enter in the diagrams of many processes in the SM. In Tab. 3.2 we list the decay modes and the respective branching ratios.

- **Boson Z**: Electrically neutral boson with a mass of approximately 91 GeV. The Z width for decay into leptons and hadrons is, respectively, $\Gamma_{ll} \approx 84$ MeV and $\Gamma_{hh} \approx 1.7$ GeV. The Z can interact off-shell (Z^*), so that the invariant mass of the

Table 3.1: Dominant decay processes for the τ^- lepton with the respective branching fractions [57]. For the decay modes intermediated by vector meson resonances, we present their type and mass. The symbol h^- represents both charged pions and kaons.

Decay mode	Resonance	Resonance mass (MeV)	Branching fraction (%)
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$			≈ 17.4
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$			≈ 17.8
$\tau^- \to h^- \nu_\tau$			≈ 11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	ρ^-	≈ 770	≈ 26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	a_1^-	≈ 1200	≈ 9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	a_1^-	≈ 1200	≈ 9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$			≈ 4.8

Table 3.2: Branching fractions for the W^- decay modes [57].

Decay modes	Branching fraction (%)
$e^- \bar{\nu}_e$	≈ 10.8
$\mu^- \bar{\nu}_\mu$	≈ 10.6
$\tau^- \bar{\nu}_\tau$	≈ 11.3
hadrons	≈ 67.6

decay products can be much smaller than the mass of the Z. In this case the process is indistinguishable from virtual photon (γ^*) exchange. In hadron colliders, when a produced lepton-antilepton is mediated by Z/Z*/γ^*, we call it a Drell-Yan process. In Tab. 3.3 we list the decay modes of the Z boson with the respective branching fractions. The leptonic decay is in general very clean in CMS (exception for the τ decay mode) and is used, for example, in efficiencies studies and calibrations.

Table 3.3: Branching fractions for the Z decay modes [57]. "Invisible" means decay into a pair of neutrinos.

Decay modes	Branching fraction (%)
$e^- e^+$	≈ 3.4
$\mu^- \mu^+$	≈ 3.4
$\tau^- \tau^+$	≈ 3.4
invisible	≈ 20.0
hadrons	≈ 69.9

3.1.3 Quarks and gluons

There are six quarks in the SM, which interact electromagnetically, weakly, and strongly. The strong interaction is mediated by vector bosons called gluons, and due to the confinement nature of these interactions, the quarks are not observed isolated in the detector. Instead we detect the hadrons, which are quark bound states. The hadrons are classified in baryons (made up by three quarks) and mesons (made up by one quark and one antiquark) [3].

---

[3]Recently it was observed by the LHCb experiment [58] compound states of $c\bar{c}uud$, called charmonium-pentaquark states.

- **Quarks up** (u)**, down** (d)**, and strange** (s): They are the lightest quarks with masses of a few MeV for u and d and approximately 100 MeV for s. These quarks make up many hadrons, some of them with very low mean lifetime and others with much bigger lifetime, for example, the protons and neutrons are made up by quarks up and down.

- **Quarks bottom** (b) **and charm** (c): The hadrons made up by these quarks are unstable with some decay products including non-isolated leptons. The hadrons made up by quarks c have mean lifetime between around 0.5 and 1.0 ps ($ct \approx 150 - 300 \ \mu$m). The hadrons made up by quarks b have mean lifetime of the order of 1.5 ps, i.e., $ct \approx 450 \ \mu$m. In general, we reconstruct these hadrons with a measurable displacement (or impact parameter) from the initial parton interaction in the proton-proton collision.

- **Quark top** (t): With a mass of around 172 GeV, the top quark is the heaviest elementary particle to date. Most of the time in proton-proton collisions they are produced in pairs. They decay before hadronize, so that we do not observe hadrons made up of top quarks, which decay with almost 100% of branching fraction in a W boson and a b quark. The production of pairs of top quarks ($t\bar{t}$) are included in the background of many analyses of physics beyond the standard model. In Tab. 3.4 we list the main decay products of ($t\bar{t}$) with the respective branching fractions.

Table 3.4: Braching fraction for the main decay modes of pairs of top quarks [57].

Decay modes	Branching fraction (%)
Semileptonic	
$e\nu_e b\bar{b}q\bar{q}'$	≈ 14.8
$\mu\nu_\mu b\bar{b}q\bar{q}'$	≈ 14.8
$\tau\nu_\tau b\bar{b}q\bar{q}'$	≈ 14.8
Leptonic	
$e^+\nu_e \mu^- \bar{\nu}_\mu b\bar{b}$	≈ 2.5
$e^+\nu_e \tau^- \bar{\nu}_\tau b\bar{b}$	≈ 2.5
$\mu^+\nu_\mu \tau^- \bar{\nu}_\tau b\bar{b}$	≈ 2.5
$e^+\nu_e e^- \bar{\nu}_e b\bar{b}$	≈ 1.2
$\mu^+\nu_\mu \mu^- \bar{\nu}_\mu b\bar{b}$	≈ 1.2
$\tau^+\nu_\tau \tau^- \bar{\nu}_\tau b\bar{b}$	≈ 1.2
Hadronic	
$b\bar{b}q\bar{q}'q''\bar{q}'''$	≈ 44.4

3.1.4 Higgs boson

In 2012 a candidate for the last unobserved elementary particle in the SM was discovered by ATLAS and CMS experiments [1, 2]. The Higgs boson (H) from the SM is a scalar elementary particle with its mass defined as a free parameter. The Higgs couplings with massive gauge bosons (proportional to the square of the boson mass) are a corollary of the Higgs mechanism and the couplings with the fermions (proportional to the fermion mass) are introduced as Yukawa terms. The branching fractions of the Higgs bosons to decay in SM particles depend on the mass of the Higgs (see Fig. 3.3). ATLAS and CMS measured

the Higgs mass with a very good precision, and the combined[4] result is $m_{\rm H} = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})$ GeV [59], which is in agreement with the SM, taking into account previous theoretical and experimental constraints. There are some preliminary measurements about the properties of the Higgs bosons that agree with the SM expectations [60, 61, 62, 63], however there are many options beyond the SM to be investigated, like the possibility of the Higgs to be a composite state.
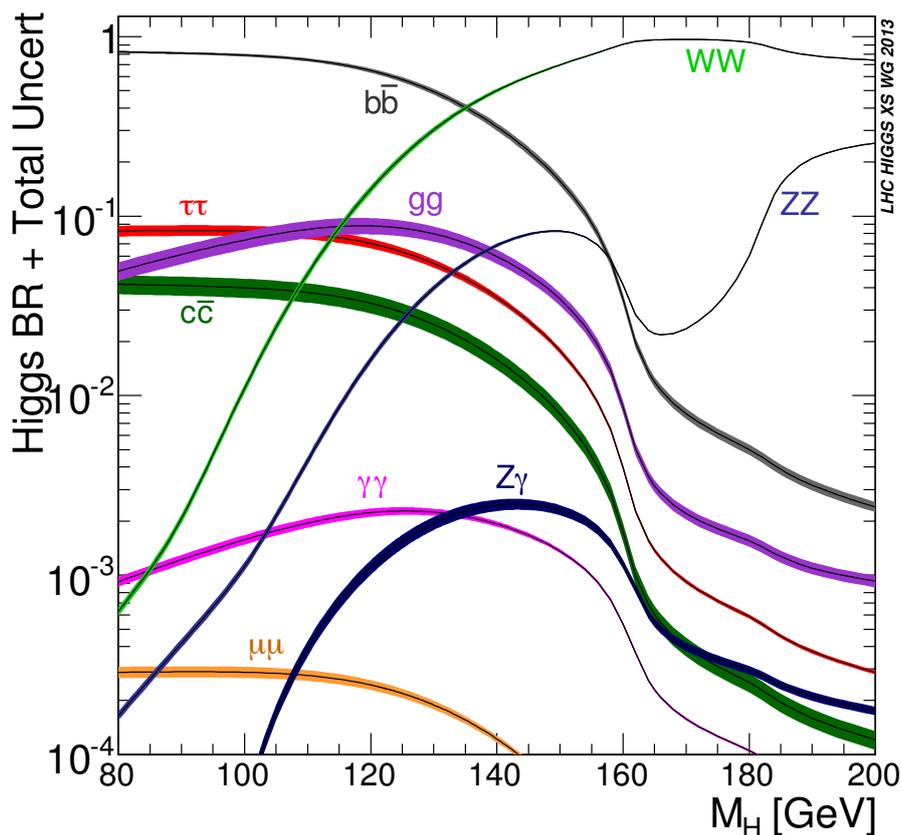
Figure 3.3: Higgs branching fractions in different decay channels as a function of its mass [64].

3.2 Going beyond the SM

As we have briefly described above, the SM is a successful framework that explains and predicts many phenomena in particle physics. Despite it is robust description of the strong and electroweak interactions and particle spectra, there are some theoretical and experimental facts that make us think about the validity of the SM at an arbitrary energy scale [65, 66, 67, 68]. Below we list three possible examples of limitations of the SM:

1. It does not account for the Dark Matter content estimated in the universe;

---

[4]The collaborations consider $\sqrt{s} = 7$ TeV and 8 TeV proton-proton collisions from both experiments, and they analyse the channels: H $\to \gamma\gamma$ and H $\to$ ZZ $\to ll\,ll$.

2. Neutrino masses are indirectly inferred by the observation of neutrino flavor oscillations. At least in its minimal formulation, the SM does not include mass for the neutrinos;

3. We do not have a consistent description of quantum-gravity at energies $\gtrsim M_{\mathrm{P}}$.

In fact, from the points above we do not have any strong hint that new physics should appear at the energies tested by LHC. From points "2" and "3" we have some indications [65] that new physics can appear at scales $\sim 10^{15}$ GeV and $\sim 10^{19}$ GeV, respectively, which are very far away from our experimental reach nowadays. This is in general the case of many other questions raised about the characteristics of the SM: if there is new physics we do not know exactly at which energy scale it will appear. Even harder is to know the nature of the theory that will describe a possible new phenomenon.

Looking deeply into the structure of the SM, we see a feature that we are not accostumated to deal with in quantum field theory: there is a hierarchy between the Higgs mass (or the Electroweak Symmetry Breaking (EWSB) scale v), and the untraviolet (UV) cutoff of the SM, Λ_{SM}. Since in the SM the Higgs is an elementary scalar, its mass receives one-loop corrections that are proportional to $\Lambda_{\mathrm{SM}} \sim M_{\mathrm{P}}$, requiring a very fine tunning in order to have the observed value of the Higgs mass. This is called the "Hierarchy Problem" (HP) [66, 65][5]. In fact, the HP is not an inconsistency of the SM. In principle, there is no problem to think that Nature has these kind of hierarchies. The problem can be formulated in other words: suppose the SM is valid up to a very high-energy scale Λ_{SM}, where will appear new particles associated with the "New Physics" (NP). How could this NP manage to generate at low energy a scale v so much different from the typical masses of the particles that is of order Λ_{SM}?

A way out of this HP is to suppose that NP will instead appear at the TeV scale. It is possible to identify three general approaches [65]:

1. Supersymmetry [69, 70]: Where bosons and fermions are connected with their respective fermions and bosons superpartners. In this theory the quadratic corrections to the Higgs mass are canceled by the contributions from the superpartners.

2. Composite Higgs [71]: Where the Higgs is a pseudo-Nambu-Goldstone boson with its mass protected by an approximate Goldstone symmetry;

3. Extra dimensions [72, 73]: In these classes of models it is possible to reduce the hierarchy allowing the gravitational interaction to propagate through the extra spatial dimension(s). The cutoff scale (in 4D) is reduced to the electroweak scale.

All these theories need to describe the physics at the energies tested so far. In this case many constraints are obtained and sometimes additional requirements are needed. Some of these approaches appear with important additional properties like in the case of supersymmetry that propose a candidate for Dark Matter.

These NP classes of models in general predict new particles that couple with SM particles resulting in signatures that can be probed at LHC. The signatures can present different features. For example, it can happen through a cascade decay resulting in distortions in SM distributions or decay of a TeV scale resonance resulting in a peak at high values of invariant

---

[5]There is another "Hierarchy Problem" about the cosmological constant (Λ_{cosmo}) [68], which will not be discussed here.

mass of final state candidates. In particular, in this thesis we will be interested in classes of models that predict the production of heavy resonances, which couples preferably with dibosons. This feature is very common, for example, in composite Higgs [6, 7, 74] and extra dimension [75, 76, 77] models.

Chapter 4

Benchmark model and simplified model approach

In order to define a strategy for the analysis, it is essential to have a benchmark model to be simulated and compared with our expectations from SM and the observed events from LHC. We intend to search for a hypotetical high-mass ($M_V \gtrsim 1$ TeV) resonance decaying into a pair of SM bosons (Higgs or massive vector bosons $W^\pm$ and Z). This kind of signature can appear, for example, in an effective description of a strongly coupled theory presented in Ref. [74]. This model is based on the coset $SO(5)/SO(4)$, i.e. at some energy scale F a global $SO(5)$ symmetry is spontaneously broken into a symmetry associated with the $SO(4)$ group, which is isomorphic to $SU(2) \times SU(2)$, $SO(4) \sim SU(2) \times SU(2)$. The model is inspired in the non-linear sigma model [68] developed to describe light mesons in QCD, like pions and kaons. In the new approach, the Higgs is analogous to these light mesons, with its mass protected by the explicit breaking of the Goldstone symmetry $SO(5)$, like the chiral symmetry in QCD that is explicitly broken by the quark masses[1]. As a consequence, it is expected to appear other composite resonances heavier than the Higgs, analogous to the vector meson ρ in QCD.

We used the implementation of the model ("Model B") from [22] (HVT paper). The main purpose of this paper is to present a model-independent analysis for possible new narrow resonances based on a simplified phenomenological Lagrangian, which can be directly identified with some explicit models of heavy vector triplets (HVT). In fact, the relation between the phenomenological Lagrangian and the experimental output (for example, limits on $\sigma \times \mathcal{B}$) has some caveats, because the reconstructed resonance mass distribution receives corrections from the parton distribution functions (PDFs) and quantum interferences from non-negligible irreducible backgrounds. It is hard to control these interferences and consequently to find an explicit connection with the parameters of the phenomenological Lagrangian, mainly in the tails of the distributions that receive higher interferences. In the region very close to the peak it is a good approximation to describe the invariant mass distribution with a Breit-Wigner (BW)[2]. Consequently, it is possible to relate in an analytical way the experimental results

---

[1]In the Goldstone theorem [78] it is expected exactly massless bosons (called Nambu-Goldstone bosons) if there is a global symmetry in the Lagrangian. When we say "explicitly broken", it means that the symmetry is broken by a small amount in the Lagrangian, resulting in low-mass bosons (called pseudo-Nambu-Goldstone bosons).

[2]Actually the experimental output is not a simple BW but a different distribution that includes the experimental resolution. In the peak region it is fine to use a convolution between a BW and a Gaussian to describe detector resolution.

and the parameters defined in the phenomenological Lagrangian. The tails of the invariant mass distribution is in general sensible to additional parameters that are especific of each explicit model, not included in the simplified description.

As we will describe in the analysis section, since the number of events from SM background is very small for our specific channel, we perform a cut-and-count analysis, with a selection optmized for a narrow resonance[3]. In this case we are almost free from effects in the tail of the resonance mass distribution. Then, in a good approximation, we can interpret the limits in terms of the couplings of the simplified model Lagrangian. We need to be caution to not consider such values of the parameters that, for example, result in signals with a non-narrow-width.

The following simplified phenomenological Lagrangian is used to describe the dynamics of the heavy vector triplet:

$$
\begin{aligned}
\mathcal{L}_{\mathrm{V}} = \;&- \frac{1}{4} D_{[\mu} V_{\nu]}^a D^{[\mu} V^{\nu] a} + \frac{m_V^2}{2} V_\mu^a V^{\mu a} \\
&+ i g_{\mathrm{V}} c_{\mathrm{H}} V_\mu^a H^\dagger \tau^a \overset{\leftrightarrow}{D}{}^\mu H + \frac{g^2}{g_{\mathrm{V}}} c_{\mathrm{F}} V_\mu^a J_{\mathrm{F}}^{\mu a} \\
&+ \frac{g_{\mathrm{V}}}{2} c_{\mathrm{VVV}} \epsilon_{abc} V_\mu^a V_\nu^b D^{[\mu} V^{\nu] c} + g_{\mathrm{V}}^2 c_{\mathrm{VVHH}} V_\mu^a V^{\mu a} H^\dagger H \\
&- \frac{g}{2} c_{\mathrm{VVW}} \epsilon_{abc} W^{\mu\nu a} V_\mu^b V_\nu^c,
\end{aligned}
\tag{4.1}
$$

where V_μ^a ($a = 1, 2, 3$) are real vector fields in the adjoint representation of $SU(2)_{\mathrm{L}}$ and with vanishing hypercharge. Written in the charge eigenstates below, they describe massive spin-1 charged and neutral particles:

$$
V_\mu^\pm = \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}}, \quad V_\mu^0 = V_\mu^3.
$$

The first two terms in the Lagrangian describe the kinetic and mass terms of the heavy vector bosons with:

$$
D_{[\mu} V_{\nu]}^a = D_\mu V_\nu^a - D_\nu V_\mu^a, \quad D_\mu V_\nu^a = \partial_\mu V_\nu^a + g \epsilon^{abc} W_\mu^b V_\nu^c,
$$

where g is the SM $SU(2)_{\mathrm{L}}$ gauge coupling and W_μ^a are SM fields.

The second line shows two important terms for the phenomenology of the models. They describe, respectively, the couplings of the heavy vector fields with the Higgs (c_{H}) and left-handed fermion (c_{F}) currents defined as:

$$
i H^\dagger \tau^a \overset{\leftrightarrow}{D} H = i H^\dagger \tau^a D^\mu H - i D^\mu H^\dagger \tau^a H
$$

and

$$
J_{\mathrm{F}}^{\mu a} = \sum_f \bar{f}_{\mathrm{L}} \gamma^\mu \tau^a f_{\mathrm{L}},
$$

where $\tau^a = \sigma^a/2$ (σ are the Pauli matrices) and D_μ defined as in SM:

$$
D_\mu = \partial_\mu + i g \tau^a W_\mu^a + i g' Y B_\mu.
$$

---

[3]Width smaller than the experimental resolution that is, for example, around 7% of the resonance mass for $M_{\mathrm{V}} = 1.0$ TeV and $M_{\mathrm{V}} = 1.5$ TeV.

The coupling g_V is introduced as an overall parameter in order to parametrize the interaction terms in a way that is more convenient to study weakly- and strongly-coupled theories, since similar combinations of parameters appear in the explicit models.

The three last terms in general result in small contributions for the LHC phenomenology, since they mainly involve self interaction among the heavy vector bosons. There are terms mixing V_μ^a with SM gauge bosons, but in most of the explicit models this mixing is negligible compared with the terms in the second line. We could think of contributions from cascade decays, but in order to satisfy the tree level SM prediction of $\rho = 1$, it was verified that the heavy vector triplet need to be practically degenerate in mass ($M_V := M_\pm \approx M_0$), suppressing this kind of signature. The tensor $W_{\mu\nu}^a$ is the same as in SM:

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon^{abc}W_\mu^b W_\nu^c.$$

After diagonalyzing the mass matrices that mix the heavy vector triplets with the SM gauge bosons and fixing a hierarchical relation among the masses in order to have TeV scale resonances and SM $m_{W,Z} \sim 100$ GeV, i.e., $(m_{W,Z}/M_V \lesssim 10^{-1})$, it is possible to compute approximate expressions for the width into fermions and bosons:

$$\Gamma_{V_\pm \to f\bar{f}'} \simeq 2\Gamma_{V_0 \to f\bar{f}} \simeq N_c[f]\left(\frac{g^2 c_F}{g_V}\right)^2 \frac{M_V}{48\pi}, \tag{4.2}$$

where $N_c[f]$ is the number of colors (3 for decay into quarks and 1 for decay into leptons).

$$\begin{aligned}
\Gamma_{V_0 \to W_L^+ W_L^-} &\simeq \Gamma_{V_\pm \to W_L^\pm Z_L} \simeq \frac{g_V^2 c_H^2 M_V}{192\pi}\frac{(1+c_H c_{VVV}\zeta^2)^2}{(1-c_H^2\zeta^2)^2} = \frac{g_V^2 c_H^2 M_V}{192\pi}[1+\mathcal{O}(\zeta^2)], \\
\Gamma_{V_0 \to Z_L h} &\simeq \Gamma_{V_\pm \to W_L^\pm h} \simeq \frac{g_V^2 c_H^2 M_V}{192\pi}\frac{(1-4c_{VVHH}\zeta^2)^2}{1-c_H^2\zeta^2} = \frac{g_V^2 c_H^2 M_V}{192\pi}[1+\mathcal{O}(\zeta^2)],
\end{aligned} \tag{4.3}$$

where $\zeta = \frac{g_V\hat{v}}{2M_V}$ ($\hat{v}$ is the Higgs field vacumm expectation value (VEV) $\langle HH^\dagger\rangle = \hat{v}^2/2$ in the models considered) and h is the "Higgs boson"[4]. The widths for difermions were computed using the Unitary Gauge and due to the small mixing angle between the heavy vectors and SM gauge bosons ($\theta_{N,C} \lesssim 0.1$), the decays are mainly in left-handed chiralities. In the diboson case, the widths for the longitudinal components were computed using the Equivalence Theorem [79], i.e., computing these quantities with the Goldstone bosons associated with the SM massive vector bosons. The transverse component can be computed in the Unitary Gauge, but it is negligible compared with the longitudinal component. The other dibosons decay modes like ZZ, hh or $\gamma\gamma$ are very suppressed or prohibited.

In principle, ζ can be of order of unit, but only for very strong coupled models with very high values of g_V. It is a good approximation to consider $\zeta \ll 1$ for the parameter values used as benchmarks. In this limit it is interesting to note that the most important part of the phenomenology of the simplified model, i.e. production and decay of heavy vector resonances, is described with a good precision by the couplings: $g^2 c_F/g_V$ and $g_V c_H$.

---

[4]Note that this is not really the Higgs boson of the SM, since it comes from the Higgs field (H) in the simplified Lagrangian, but we expect it to have very similar properties (at least it needs to be compatible with the properties measured until now). When working with the signal we will consider this h as the SM Higgs boson.
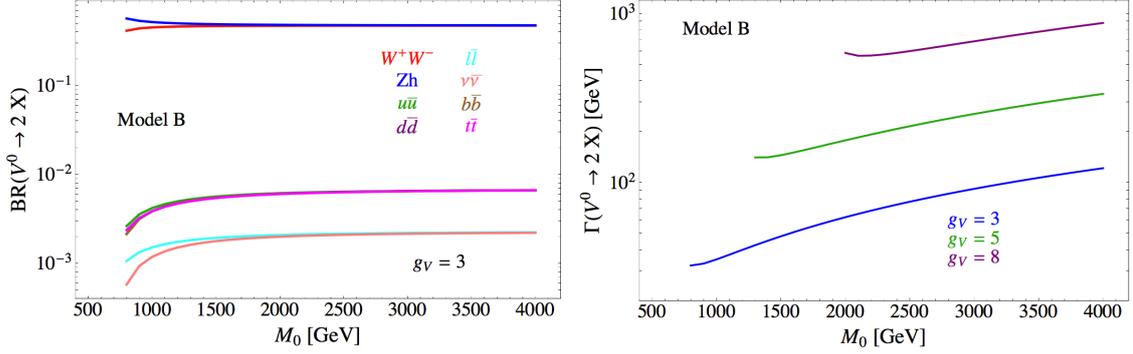
Figure 4.1: For $c_H \sim c_F \sim 1$. Left: Branching fractions of two body decay for neutral heavy vector resonance as a function of its mass and fixing $g_V = 3$. Right: Total width for the neutral heavy vector resonance as a function of its mass [22].
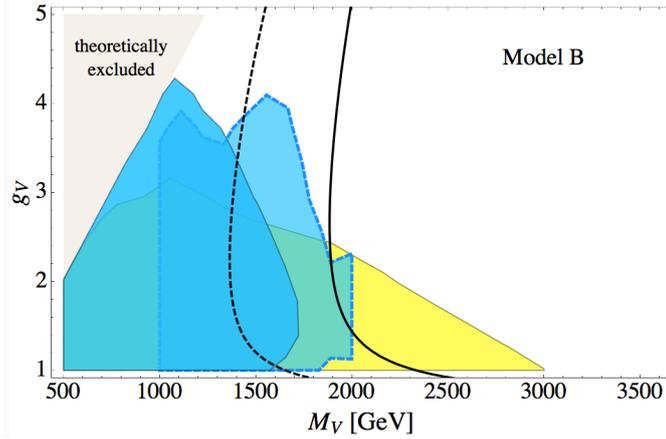
Figure 4.2: Excluded regions from direct searches and EWPT. The colored regions are excluded at 95% of C.L. by the direct searches: $l\nu$ [80] (yellow), $WZ \rightarrow 3l\nu$ [81] (dark blue), and $WZ \rightarrow jj$ with W/Z tagged jets [82] (light blue). The regions in the left of the black curves are excluded by EWPT: strict 95% of C.L limts on $\hat{S}$ of Ref. [83] (continuous line) and multiplying the bound in $\hat{S}$ by a factor two (dotted line). The theoretically excluded region corresponds to parameters values where it is not possible to reproduce the SM parameters α_{EW}, G_F, and M_Z [22].

The matching of the phenomenological Lagrangian with a specific model describing the lighest vector resonances from a composite Higgs theory described in [74] results in the characteristic values of the parameters,

$$c_H \sim c_F \sim 1.$$

To simulate signal events, we fixed $g_V = 3$ and generated resonances with $M_V = 0.8, 0.9, 1.0, 1.2, 1.5, 2.0,$ and 2.5 TeV. This combination of parameter values results in heavy resonances decaying predominantly into dibosons. From expressions 4.2 and 4.3, it is possible to observe this feature. In Fig. 4.1 we show the branching fractions for the two body decay
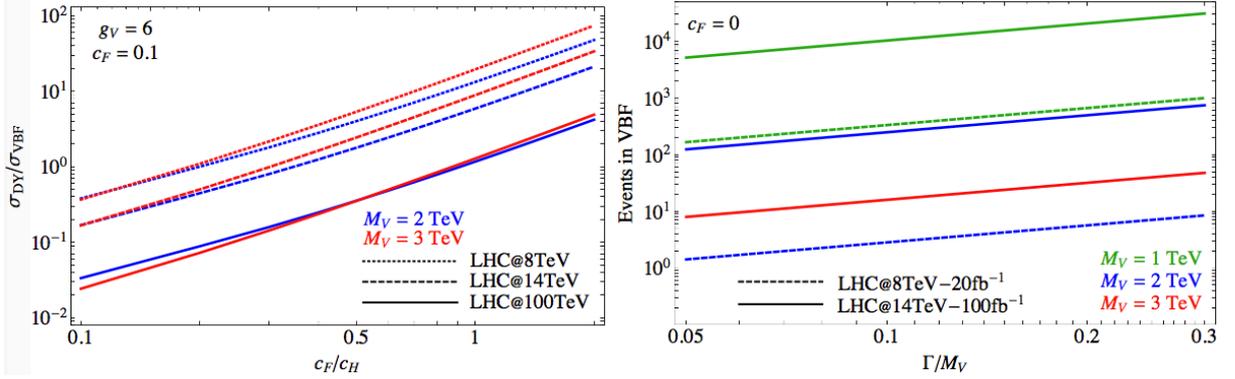
Figure 4.3: Left: ratio between DY and VBF production cross-section for different resonance masses and collider energies. Right: Considering $c_F = 0$. Number of signal events produced through VBF as a function of the ratio between the total width and resonance mass [22].

of the neutral heavy vector boson (Z′) as a function of the mass considering the parameters fixed above. The performance of the standard leptonic searches like Z′ → ll or W′ → $l\nu$ is very reduced in these scenarios. In Fig. 4.2 we present limits in the plane (M_V, g_V) with $c_H \sim c_F \sim 1$. Colored regions represent excluded values for (M_V, g_V) from the CMS direct searches in the following final states: $l\nu$ [80] (yellow), WZ → $3l\nu$ [81] (dark blue), and $WZ → jj$ with W/Z tagged jets [82] (light blue). The limits are not statistically combined but just superimposed in the plot. The regions in the left of the black curves are excluded by electroweak precision tests (EWPT). The continuous line represents strict 95% C.L. on $\hat{S}$ of Ref. [83], while the dashed line represents more conservative limits enlarging the bounds on $\hat{S}$ by a factor of two. Comparing the limits from direct searches, we observe better limits from leptonic analysis in the low g_V region ($g_V \lesssim 3$), these in general represent weakly coupled vector resonances that can be produced in models with an extended SM gauge group [84]. For $g_V \gtrsim 3$, which configures strong coupled models (non-perturbative theories) with higher branching fractions in dibosons, the limits are much better in diboson searches. Comparing with the constraints from EWPT, for the low g_V region the direct searches has a better performance, but the contrary happens for $g_V \gtrsim 3$.

The parameters that we chose for our benchmark model are approximately located in the border of the exclusion regions in the parameter space. For the production mechanism we will consider a Drell-Yan process, which is the dominant production considering $c_F \sim 1$, $c_H \sim 1$, and $g_V \sim 3$.

In the case of the strongly coupled scenario another possibility could be a vector boson fusion (VBF) production. From left plot of Fig. 4.3 we note that VBF can be comparable (or even dominant) to Drell-Yan for regimes with very high g_V ($g_V \sim 6 - 7$), mainly for higher energies of proton-proton collisions, for example $\sqrt{s} = 100$ TeV. Another possibility in strongly coupled theories could be the investigation of the VBF production in models where $c_F \approx 0$. From right plot of Fig. 4.3 we note that this scenario could be studied for 8 TeV proton-proton collisions, but the mass reach would be very small (around ~ 1 TeV) in order to have a considerable number of signal events and sufficient narrow resonances ($\Gamma/M_V \lesssim 0.1$).

Since V or H have many decay modes, the diboson channels have lots of possible final state configurations. There are many analyses in CMS and ATLAS trying to cover as much
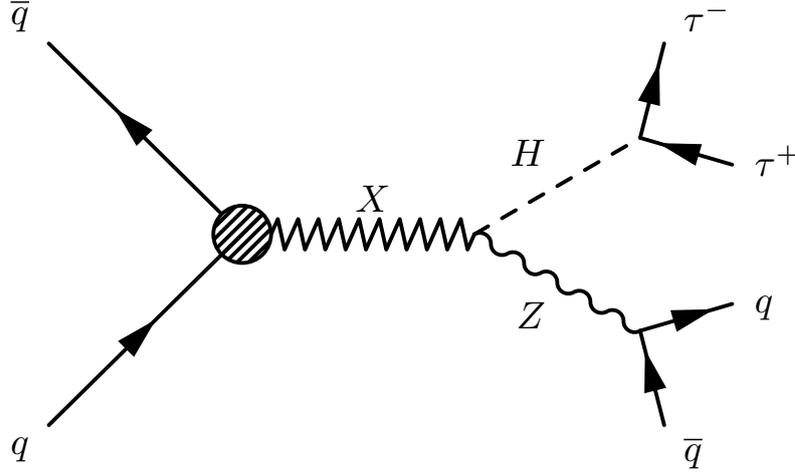
Figure 4.4: Feynman diagram for the signal reaction.

as possible these possibilities. The first channels investigated in proton-proton collisions at $\sqrt{s} = 8$ TeV were VV (V = W, Z) in semileptonic and all-hadronic final states. After, as a continuation of these analyses both collaborations started to cover VH and HH channels.

In the ZH channel, the following options for the final state are possible (higher branching fractions):

1. All-hadronic channel: $H \to b\bar{b}$, $Z \to q\bar{q}$;

2. Semileptonic channel: $H \to b\bar{b}$, $Z \to l^-l^+$;

3. Hadronic-invisible channel: $H \to b\bar{b}$, $Z \to \nu\bar{\nu}$;

4. Semileptonic $\tau\tau$ channel: $H \to \tau^-\tau^+$, $Z \to q\bar{q}$;

5. All-leptonic $\tau\tau$ channel: $H \to \tau^-\tau^+$, $Z \to l^-l^+$;

6. Leptonic-invisible $\tau\tau$ channel: $H \to \tau^-\tau^+$, $Z \to \nu\bar{\nu}$.

In CMS, the first three channels have other groups working on them. We analyze here the channel in item "4" (see corresponding Feynman diagram in Fig. 4.4). The "all-leptonic $\tau\tau$" channel has a very low branching fraction, but the signal is very clean, with zero background contamination at high-mass. Despite the lower branching fraction with respect to the "semileptonic $\tau\tau$", the "leptonic-invisible $\tau\tau$" channel seems really challenging because we have a "mixture" of the neutrinos from the Z and from the τ decays.

In the "semileptonic $\tau\tau$" channel, the main expected backgrounds come from fakes. These fakes can be jets that pass the reconstruction and identification criteria used to define the τ object in the analysis. It is possible as well to misidentify electrons and muons as τ decaying hadronically. In the $Z \to q\bar{q}$ side it is possible to have fakes from quark/gluon-jets. Events with a genuine boosted Z/W bosons decaying hadronically or $Z \to \tau\tau$ events that enter in the Higgs mass window are also possible contributors for the background. With these considerations, the main processes that contribute to the background are: QCD multijets, $t\bar{t}$ and Single $-$ top production, $Z/W +$ jets and SM dibosons (VV). In principle, one could consider the irreducible SM background: $pp \to Z^* \to Z$ (q$\bar{q}$) H ($\tau\tau$). But the cross-section

of this process at $\sqrt{s} = 8$ TeV is very small compared with the other backgrounds and we expect a negligible contribution in the signal region (after all selection cuts). For example, comparing with the ZZ background that has a production cross-section of around 8 pb, the ZH production has a cross-section supressed by a factor of 10^{-2} [85]. We will show in the analysis section that the number of events expected for the ZZ background in the signal region is compatible with zero. The total selection efficiency is higher for the ZH process because of the $H \rightarrow \tau\tau$ reconstruction, but we will show that even using a looser selection, without the Higgs identification, the ZZ background is already negligible.

The proportion among the background components depends on the decay modes of the τ leptons that come from the Higgs. In this thesis I will detail the channel where both τ leptons decay hadronically. In collaboration with other students from University of Perugia and Zurich we covered the leptonic decay modes as well (see Ref. [86]).

Before describing the specific analysis channel $pp \rightarrow Z' \rightarrow ZH \rightarrow q\bar{q}\,\tau\tau$, we will present a more general description about the searches for high-mass resonances in the diboson channels in proton-proton collisions with $\sqrt{s} = 8$ TeV at LHC.

Chapter 5

Overview of heavy resonances searches in the VV/VH/HH channels by CMS and ATLAS

There are many models that predict heavy resonances decaying preferably into SM bosons with no strong limits from standard fermionic final state searches. In this chapter, we will give a general idea about the difficulties and search strategies on this topic used by ATLAS and CMS experiments. We are going to describe the analyses using $\sqrt{s} = 8$ TeV proton-proton collisions of 2012 LHC Run, which present the more stringent limits for resonances with masses $\gtrsim 1$ TeV.

The production of TeV resonances decaying into SM boson pairs, implies that these bosons will be created with a high-momentum. In this case, the decay products of the bosons are within a small angular interval, with distance in $\eta\phi$ plane given by:

$$\Delta R_{z_1 z_2} \equiv \sqrt{(\phi_{z_1} - \phi_{z_2})^2 + (\eta_{z_1} - \eta_{z_2})^2} \approx \frac{2M^Y}{p_T^Y}, \tag{5.1}$$

where M^Y and p_T^Y are respectively the mass and transverse momentum of the SM boson. For example, for a resonance with mass of 2 TeV decaying into two Z bosons, the tipical distance in $\eta\phi$ plane between the Z decay products is $\Delta R \sim 0.2$. This feature of very collimated decay products introduces additional complications in the reconstruction and identification of the bosons.

In the case of a hadronic decay, the final state in the detector will be characterized by a wide jet with high-energy. If we use only the common jet kinematical quantities (like mass, p_T, η, ϕ) to distinguish these "fat-jets" from the ones originated primarily from quark/gluon, we will conclude that the fake-rate is still very high and the bosons will be badly identified. In addition, since the area ocupied by the jet is higher, the contamination from secondary processes like underlying-event, pile-up, final state radiation, will be higher [87], introducing considerable interferences in the jet kinematics. In order to improve the boson identification (tagging), it was developed several procedures/algorithms that we can roughly divide into two categories [1]:

1. Groomers [88, 89, 90, 91]: It tries to minimize as much as possible the effect from soft (low p_T) and high-angle jet constituents in the jet kinematics. As a result of the jet

---

[1]In fact, there are algorithms that consider both techniques at once.

grooming, the interferences from pile-up, underlying-event and initial state radiation
will have a considerable reduction. Since these soft and high-angle radiations tend to
considerably modify (in general increase) the mass of the jet, the groomed jet mass
tends to be closer to the mass-scale of the primary parents, i.e., it shifts toward zero for
quark/gluon jets and toward the pole mass of the boson for boson-jets. In this scenario
the mass of the groomed jet becomes a really powerfull variable to distinguish signal
jets (from boson decay) from quark/gluon jets (from QCD background). In general we
optimize a window for the jet groomed mass around the pole mass of the boson.

2. Taggers (or shape variables) [89, 92, 93, 94, 95, 96, 97, 98]: As we said before, in the
   boosted regime we have wide jets representing the bosons. If we look "inside" these
   jets, it is possible to identify a different substructure as compared with quark/gluon
   jets. For example, in the case of the SM bosons (Z, W, and H) with a two-prong
   decay, the wide jets are expected to have two localized energy clusters that can be
   defined as two "subjets". The taggers use these features to discriminate the boson-jets
   from quark/gluon-jets.

In order to improve the sensitivity of the 8 TeV analyses with respect to 7 TeV, ATLAS
and CMS have used some combinations of groomers and taggers to identify boosted bosons
decaying hadronically. Both collaborations have studied the performance of many groomers
and taggers [99, 100, 101, 102]. For example, in Fig. 5.1, the efficiency of a mass cut,
$60 < m_{\mathrm{jet}} < 100$ GeV, for the groomed jets and a tagger selection represented by a cut in
a variable defined as τ_2/τ_1 are shown as a function of the jet p_{T} and the number of vertices
[2]. It was used a simulation of WW signal events (both W decaying hadronically) in the
CMS detector. There is a turn-on behavior at low p_{T} representing the p_{T} region where both
subjets start to be reconstructed within a single jet. There is a decrease in efficiency for
high-p_{T}, representing a degradation in the detector resolution for jet substructure. The τ_2/τ_1
variable suffers from these effects, increasing more the p_{T} dependence of the efficiency. The
efficiency in the mass cut drops by $\sim 6\%$ from 5 to 30 reconstructed vertices. Also, the
τ_2/τ_1 variable shows a pile-up dependency, dropping by $\sim 12\%$ from 5 to 30 reconstructed
vertices.

For the leptonic decay we have a cleaner situation, but some care is necessary. For
example, it is common to define an isolation criteria for muons, electrons, and taus, which are
very important to discriminate signal events from QCD background. The isolation basically
quantify the activity in the tracker and calorimeters detectors around the respective lepton
(p_{T}-sums), excluding the lepton p_{T} itself. Since in this boosted topology the directions of
the leptons have a small angular separation, it is possible that one lepton spoil the isolation
criterium of the other and vice-versa. One common approach is to identify the leptons and
then remove their contributions for the isolation p_{T}-sums.

CMS and ATLAS used a common general approach to perform the searches for heavy
resonances: when the signal is expected to have a boson decaying hadronically, it is identified
using an optimized combination of a mass window for groomed jets and a shape variable
criterion - in the case of the Higgs decaying into $b\bar{b}$ in addition it was used b-tagging. In
the leptonic side, the visible candidates (electrons and muons) are identified applying the
corrections described above. From hadronic and/or leptonic identification, it is possible to

---

[2]The τ_2/τ_1 variable quantifies the tendency of the jet to be composed by two subjets, typical of a hadronic
V decay.
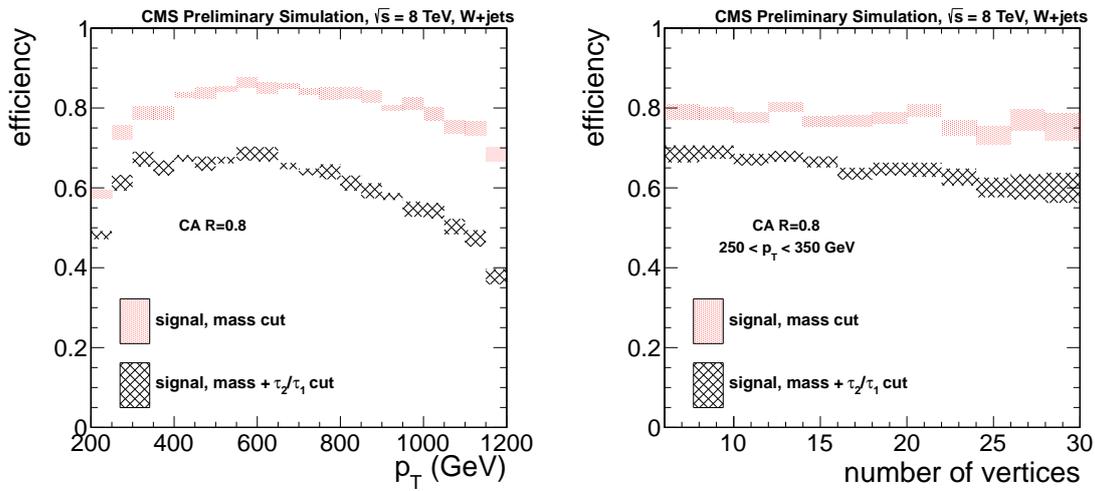
Figure 5.1: Efficiency of the m_{jet} cut and the τ_2/τ_1 cut combined with the m_{jet} cut [101]. Left: Efficiency as a function of the jet p_{T}. Right: Efficiency as a function of the number of reconstructed vertices, which quantifies the amount of pile-up.

estimate the bosons 4-momenta and consequently reconstruct the heavy resonance. After defining a dedicated kinematical selection to isolate as much as possible the signal events from SM background, a search for bumps is performed in the smooth diboson invariant mass distribution.

In Tab. 5.1 we list some analyses by ATLAS and CMS searching for heavy resonances decaying into dibosons applying jet substructure techniques. We have selected the analyses using $\sqrt{s} = 8$ TeV proton-proton collisions that became public before 25/06/2015. Besides the channels, we show the models used as benchmark, the resonance mass range investigated and the 95% C.L. upper limit on the resonance production cross-section times the branching fraction into dibosons. We have just picked the upper limits corresponding to the mass values in the extremity of the search mass range [3].

There are some other analyses with boosted topology in progress by CMS and ATLAS using $\sqrt{s} = 8$ TeV proton-proton collisions, mainly investigating VH channels. At the same time both collaborations are working hard in the preparation of the analyses for the RunII of LHC that will use $\sqrt{s} = 13/14$ TeV proton-proton collisions. This increase in the center-of-mass energy provides higher mass reach in the searches, due to the increase in the parton luminosities. In addition, it is being studied new techniques for jet substructure [108] and pile-up/underlying-event mitigation [109] that tend to improve the performance of the analyses as a whole.

It is important to note that some analyses using 2012 data-samples, in both experiments, observed small excesses of events for masses around 1.8-2.0 TeV [17, 14, 13, 18]. The highest excess was observed in the dijet ATLAS analysis [17] with a global significance of around 2.5σ in the channel $X \to WZ \to JJ$. This is an additional motivation to pay special attention on these boosted signatures at LHC for the RunII.

In the next chapter we will describe the analysis listed in the last row of Tab. 5.1, detailing the all-hadronic decay mode of the pairs of τ leptons.

---

[3]In the VV all-hadronic searches ATLAS and CMS had splitted the limits optimizing the groomed jet mass window for Z and W candidates. In this case we just showed the more sensitive result.

Table 5.1: CMS and ATLAS searches for heavy-resonances decaying into dibosons. It is shown the searches with pp at $\sqrt{s} = 8$ TeV using jet substructure techniques and that became public before 25/06/2015. Note that some searches include non-boosted topology as well.

Channels	Experiments	Benchmark models	Search mass range (TeV)	$\sigma(\text{pp} \to \text{X} \to \text{diboson})$ limit (pb)
$ZV \to ll\, q\bar{q}$	CMS [14]	Bulk graviton [75, 76, 77]	0.6 to 2.5	$\sim 1 \times 10^{-1}$ to 1×10^{-2}
	ATLAS [15]	Bulk graviton	0.3 to 2.0	$\sim 1 \times 10^{0}$ to 1×10^{-2}
		EGM [103]	0.4 to 2.0	$\sim 1 \times 10^{0}$ to 1×10^{-2}
$WV \to l\nu\, q\bar{q}$	CMS [14]	Bulk graviton	0.8 to 2.5	$\sim 7 \times 10^{-2}$ to 3×10^{-3}
	ATLAS [16]	Bulk graviton	0.3 to 2.5	$\sim 1 \times 10^{1}$ to 7×10^{-3}
		EGM	0.3 to 2.5	$\sim 2 \times 10^{1}$ to 1×10^{-2}
	CMS [13]	q* [104, 105]	1.0 to 3.9	$\sim 7 \times 10^{-1}$ to 3×10^{-3}
		Bulk graviton	1.0 to 2.9	$\sim 1 \times 10^{-1}$ to 5×10^{-3}
		RSG [106, 107]	1.0 to 2.9	$\sim 2 \times 10^{-1}$ to 1×10^{-2}
		EGM	1.0 to 2.9	$\sim 1 \times 10^{-1}$ to 5×10^{-3}
$VV \to q\bar{q}\, q\bar{q}$, $q'V \to q'q\bar{q}$	ATLAS [17]	Bulk graviton	1.3 to 3.0	$\sim 3 \times 10^{-2}$ to 8×10^{-3}
		EGM	1.3 to 3.0	$\sim 2 \times 10^{-2}$ to 1×10^{-2}
$VH \to$ all-hadronic includes $H \to WW^{*} \to$ all-hadronic	CMS [20]	HVT [22]	1.0 to 2.6	$\sim 4 \times 10^{-2}$ to 4×10^{-3}
$HH \to$ bb bb	ATLAS [21]	Bulk graviton	0.5 to 2.0	$\sim 4 \times 10^{-2}$ to 1×10^{-2}
$WH \to l\nu$ bb	CMS [18]	HVT	0.8 to 2.1	$\sim 1 \times 10^{-1}$ to 2×10^{-2}
$ZH \to q\bar{q}\,\tau\tau$	CMS [86]	HVT	0.8 to 2.5	$\sim 1 \times 10^{0}$ to 3×10^{-2}

Chapter 6

Heavy resonances search in the ZH channel at CMS detector

A high-mass spin-1 resonance decaying into a Higgs and a Z bosons is searched. The events are caracterized by two hemispheres: one where the Higgs boson is reconstructed through its decay in a pair of τ leptons and the other where the Z boson is reconstructed from its hadronic decay.

In H $\to \tau\tau$, for TeV scale resonances, the directions of the two τ leptons have a small angular separation. In Fig. 6.1 we show the distributions of the ΔR between the two τ leptons at generator level [1] using four mass values for the signal resonance: 1.0, 1.5, 2.0, and 2.5 TeV. Increasing the resonance mass, ΔR moves to lower values. This overlap introduces inefficiencies in the standard τ reconstruction. To avoid this, a different technique for τ lepton reconstruction have been studied.

In Z $\to$ qq, as in the case of the τ leptons, the two quarks are produced with a small angular separation, being identified as a single jet (Z-jet) in the detector. In order to discriminate these objects from quark/gluon-jets we use substructure techniques.

After reconstructing and identifying the Z and Higgs bosons, we combine these informations to reconstruct the heavy resonance. The mass of this object helps in the caracterization of the signal region. We search for excess of events with respect to the SM for masses above 0.8 TeV. In this search we follow a strategy called "blinded analysis", where we define and validate background prediction methods in a kinematical region that does not contain a significant number of signal events (side-band region). From the statistical point of view the "blinding" is important because it avoids bias from our knowledge about the observed events in signal region. Once we have the background estimate, signal, and observed events, we perform a statistical test in order to see if we can discard at 95% confidence level a possible class of models resulting on this signature.

I will describe in details the final state where both τ leptons decay hadronically (all-hadronic channel), since this was the channel that I have worked in the CMS collaboration. I will present the final results considering all the categories (all-leptonic, semileptonic, and all-hadronic). These results were published in PLB [86].

---

[1] It is used the kinematics of the Higgs decay into tau leptons provided by Monte Carlo generator without considering detector and reconstruction effects.
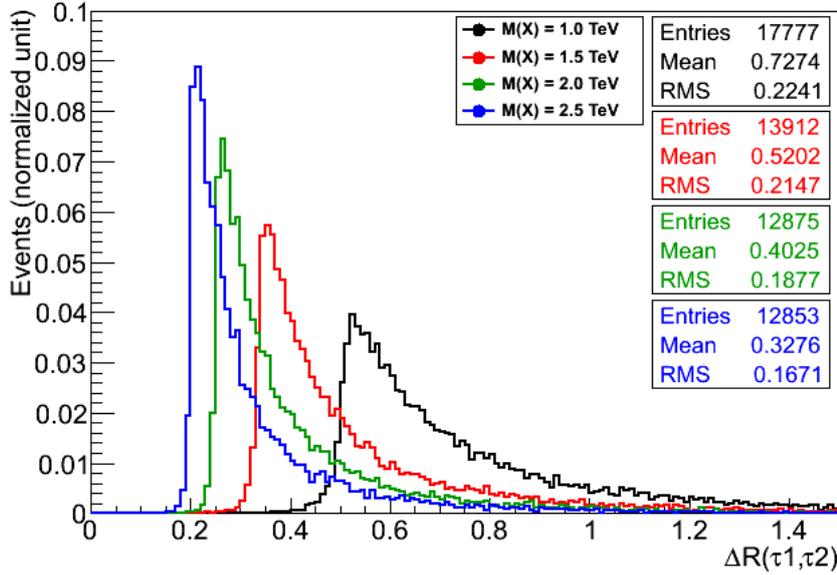
Figure 6.1: Spacial distance between the two τ leptons coming from the decay of the Higgs boson for four values of the resonance mass.

6.1 MC simulation and observed events

Here we describe the MC simulation for signal and background and the observed events used in the analysis, indicating the run periods, triggers, and other baseline requirements used to prepare these samples.

6.1.1 Signal samples

As a benchmark model for the signal events we consider Z' production via quark-antiquark interaction, and the decay $Z' \to ZH \to q\bar{q}\tau^+\tau^-$. For the parton level description of this reaction it was used MadGraph 5 1.5.11 [110], producing Les Houches Event (LHE) files [111] with basic kinematic quantities. The generation was based in the couplings described in [22] for strongly coupled models: $g_V = 3$ and $c_F = -c_H = 1$ (see Chapter 4) [2]. On Tab. 6.1 we list the signal samples generated, characterized by the mass of the Z'. Taking as input these LHE files, it was used PYTHIA 6.426 [112] for the showering and hadronization. Tau leptons decay were done using TAUOLA library [113]. The passage of the stable particles through the detector material and detector geometry were simulated using GEANT4 [114]. As done in real data events, detector information was digitalized. Simulated minimum bias samples have been mixed to signal events in order to emulate effects of pile-up. A set of trigger criteria was emulated and stable particles were reconstructed.

In this analysis chapter, when referring to MC simulation of signal events, we will be referring to the samples described here.

---

[2]I would like to thank Alexandra Carvalho, Angelo Santos, and Jennifer Ngadiuba for the production of the LHE files and cross-sections computation.

Table 6.1: Signal samples: Z' production cross-section (σ) and branching fraction (B) to decay in ZH as function of the Z' mass.

$M_{Z'}$ [GeV]	$\sigma(pp \to Z')$ [fb]	$B(Z' \to ZH)$
800	317.121	0.5672
900	238.800	0.5340
1000	165.454	0.5176
1200	76.261	0.5007
1500	24.370	0.4893
2000	4.013	0.4815
2500	0.708	0.4783

6.1.2   Background samples

All background MC simulation have the same setup as used for the signal events. In the Z-jet plus two collimated high-p_{T} τ leptons final state, the main backgrounds depend on the decay mode of the two τ leptons. Events considered in the all-hadronic channel are: QCD multijets with large H_{T} simulated with PYTHIA 6.426 (H_{T} is the scalar sum of the p_{T} of all jets with $p_{\mathrm{T}} > 30$ GeV), $t\bar{t}$ production simulated with POWHEG 1.0 r1380 [115, 116, 117, 118], SM diboson (PYTHIA 6.426), and Z/W+jets (MadGraph 5 1.3.30). For QCD and W+jets we use cross-sections at leading order. For the other components all the cross-sections are at next-to-leading order. On Tab. 6.2 we list the backgrounds with the number of events and cross-sections. Most of these components (except ZZ and WZ) present reducible signatures. We have tight criteria to avoid high fake-rate in the analysis, but even in this case we have fakes of τ leptons from quark/gluon-jets, isolated electrons and muons. We have fakes of Z-jets from quark/gluon-jets as well.

Table 6.2: Background samples: number of generated events and cross-sections.

Sample	Number of events	σ [pb]
QCD multijets: $250 < H_{\mathrm{T}} < 500$ GeV	27062078	2.76×10^5
QCD multijets: $500 < H_{\mathrm{T}} < 1000$ GeV	30599292	8426
QCD multijets: 1000 GeV $< H_{\mathrm{T}}$	13843863	204
Z($\to ll$)+jets: $M_{ll} > 50$ GeV and $70 < p_{\mathrm{T}}^{\mathrm{Z}} < 100$ GeV	11764538	53.0
Z($\to ll$)+jets: $M_{ll} > 50$ GeV and 100 GeV $< p_{\mathrm{T}}^{\mathrm{Z}}$	12511326	32.9
Z($\to ll$)+jets: $10 < M_{ll} < 50$ GeV and $70 < p_{\mathrm{T}}^{\mathrm{Z}} < 100$ GeV	5389313	11.05
Z($\to ll$)+jets: $10 < M_{ll} < 50$ GeV and 100 GeV $< p_{\mathrm{T}}^{\mathrm{Z}}$	4146124	4.22
W($\to l\nu$)+jets: 400 GeV $< H_{\mathrm{T}}$	4971847	25.22
$t\bar{t}$ production	21675970	225.20
WW	10000431	57.11
WZ	10000283	33.21
ZZ	9799908	8.06

6.1.3   Data samples, High Level Trigger and pile-up mitigation

For the six channels in this analysis we consider the same combination of hadronic High Level Triggers (HLT) to select the observed events. Besides the easier way to combine channels using always the same samples, we found that the efficiency in channels with muons/electrons in the final state is larger when triggering on jet activity because of the

high-energy thresholds of dilepton triggers. Events are selected online by triggers that require the presence of at least one of the following: either a hadronic jet reconstructed by the anti-k_{T} algorithm [119] with a distance parameter of 0.5, transverse momentum p_{T} larger than 320 GeV, and $|\eta| < 5.0$; or a total hadronic transverse energy, H_{T}, defined as the scalar sum of the transverse energy of all the jets of the event, larger than 650 GeV. The trigger names with the respective criteria are listed in Tab. 6.3. Using events selected by less restrictive, pre-scaled triggers, it has been verified that the efficiency of these criteria after applying the offline selection is above 99%. In Fig. 6.2 we show the efficiency for three different trigger requirements. The efficiency is calculated as a function of the jet p_{T} and a significant improvement is shown when using the "OR" between the three triggers, so that the plateau of the curve, defined at the point when the efficiency becomes higher than $\approx 99\%$, is shifted from ≈ 460 GeV to ≈ 400 GeV.

Taking into account only the certified luminosity sections, we investigated 19.7 fb^{-1} of proton-proton collisions at 8 TeV from 2012 LHC run. On Tab. 6.4 we list these samples.

Table 6.3: Trigger names and the respective criteria. H_{T} is defined as the scalar sum of the transverse energy of all the jets of the event.

Trigger names	Criteria
HLT_PFJet320 (unprescaled)	At least one anti-k_{T} jet with R = 0.5 and: $p_{\mathrm{T}} > 320$ GeV $|\eta| < 5.0$
HLT_PFJet140 (pre-scaled)	Same as in HLT_PFJet320, but jet $p_{\mathrm{T}} > 140$ GeV
HLT_PFHT650/HLT_PFNoPUHT650 (unprescaled)	$H_{\mathrm{T}} > 650$ GeV

Table 6.4: Observed events: Number of events and integrated luminosity processed. These samples were reprocessed, improving the event reconstruction, at the beginning of 2013.

Sample	Number of events	L_{int} [pb^{-1}]
Run2012A-22Jan2013	9400333	876.225
Run2012B-22Jan2013	15513284	4412
Run2012C-22Jan2013	25774250	7048
Run2012D-22Jan2013	27572330	7367

As we have discussed in the MC simulation section, we emulate the pile-up interactions in MC. In Fig. 6.3 we show the number of pile-up interations corresponding to the conditions used in all MC simulations and the corresponding distribution for observed events in the entire 2012 run. We do a reweighting in MC distributions so that the pile-up distribution in MC is compatible with the one in observed events.

Jets are the main objects in this analysis that are affected by the pile-up interactions. The energy of the jets can be biased by components coming from pile-up. It is possible to reconstruct jets totally composed by pile-up components. In order to correct this effect, before jet clustering we remove the charged constituents that do not come from the primary vertex - the primary vertex is chosen as the one with highest sum of p_{T}^2 of the tracks. To remove the bias on the jets kinematics due to the additional neutral-hadron component we use the jet-area method implemented in Fastjet software [120, 121]. Finally, for the leptons we apply only corrections due to the pile-up on the isolation criterion.
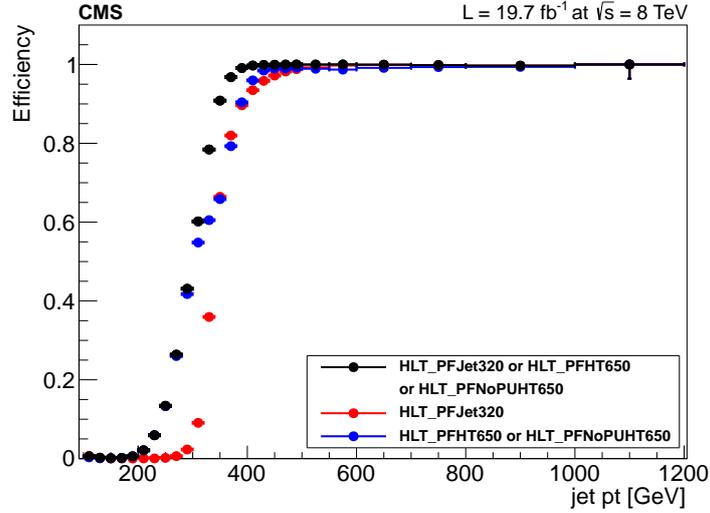
Figure 6.2: Trigger efficiency in observed events as a function of the jet p_{T}. The labels correspond to the trigger names defined in Tab. 6.3.
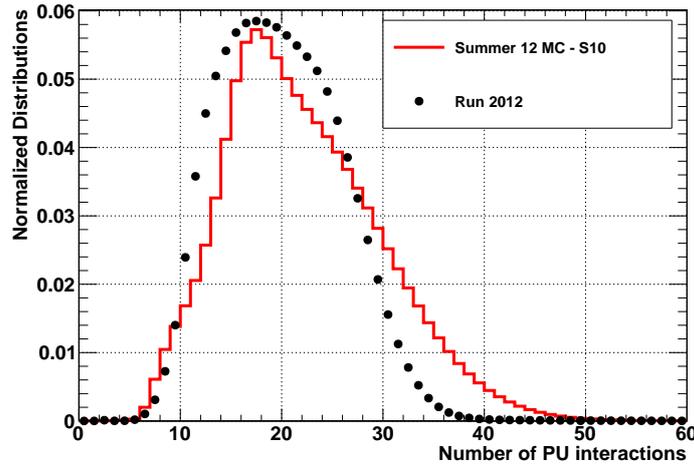
Figure 6.3: Number of pile-up interactions in MC simulations and observed events.

6.2 Physics object reconstruction and identification

Here we define the reconstruction and identification in the analysis level. We are going to show the offline reconstruction that is one step before we define a selection to improve the signal sensitivity. After this section we will have defined the τ leptons, Z-jets, missing transverse energy, and the Higgs. We will generically call them physical objects. The offline object reconstruction basically optimizes a combination of criteria based on the information from CMS subdetectors. In this analysis, the reconstruction of the missing transverse energy, hadronic tau decays, and jets starts from basic objects reconstructed and identified by the particle-flow (PF) algorithm in CMS [122, 123]. For the Higgs and Z bosons reconstruction we consider respectively, a pair of τ leptons and a wide jet.

6.2.1   Particle-flow

The particle-flow algorithm reconstructs and identifies stable particles (i.e., photons, electrons, muons, charged and neutral hadrons) using consistently the information from all the CMS subdetectors to estimate the energy, direction, and type of these particles.

The algorithm has the following three steps:

1. It reconstructs the tracks from tracker system using an iterative-tracking strategy [124, 125]. Energy clusters from ECAL and HCAL are defined by a dedicated calorimeter clustering, and muon-tracks are defined combining the three muon systems (both methods described in Ref. [122]). These objects are called "elements". One stable particle is expected to be reconstructed by a combination of many particle-flow elements.

2. The elements above are combined into objects called "blocks" by Link algorithm [122], which connects them based on $\eta \times \phi$ distance (closer elements have higher probability to be connected) and/or χ^2 of global fits that combine tracker and muon systems fits.

3. For each block, to avoid ambiguity, the PF algorithm applies criteria to define the particle candidates in a sequence starting from muons, electrons, then charged hadrons, photons, and neutral hadrons. Per block it is possible to define more than one particle. For example, the "particle-flow muon" is defined combining the tracker and muon systems information, however there might be additional elements in the block, like ECAL/HCAL clusters, that will be used to reconstruct and identify photons and hadrons.

From the last step a list of stable particles is produced, which can be used to reconstruct compound objects, like taus, jets, and missing transverse energy.

6.2.2 τ leptons

The standard τ hadronic reconstruction and identification in CMS is performed by the hadron-plus-strips (HPS) algorithm [126] in two steps: one that reconstructs the decay modes and other that use criteria (discriminators) in order to mitigate fake-rates from quark/gluon-jets, electrons, and muons.

The first step of the HPS technique uses as input the particle-flow constituents of a jet clustered with the anti-k_T algorithm with distance parameter equal to $\mathrm{R} = 0.5$. The HPS algorithm uses the PF constituents to define π^0 and charged hadron ($h^\pm$) candidates compatible with a τ decaying in $h^\pm h^\pm h^\mp$, $h^\pm \pi^0 \pi^0$, $h^\pm \pi^0$, and $h^\pm$. The four-momentum of the τ decaying hadronically (τ_h) is constructed by the four-momentum sum of all the candidates used in the reconstruction.

In the second step, an isolation criterion is used to reduce jet $\rightarrow \tau_\mathrm{h}$ fake-rate, which is produced with a very high cross-section in QCD processes. Unlike quark/gluon-jets, τ decay is described by an electro-weak process and does not have high hadronic activity in a region around the τ_h candidate. The isolation basically quantify the neutral/charged particle activity coming from the primary vertex whithin a cone of $\Delta\mathrm{R} = 0.5$ around the τ_h candidate and it does not take into account the constituents used in the τ_h reconstruction. It is also possible that electrons and muons pass the requirements in the reconstruction and isolation. Additional discriminators that compare energy deposits in the ECAL and HCAL

to discriminate τ_h from e , and consider hits in the muon chambers to discriminate τ_h from μ are used to reduce these fake-rates.

In this analysis we use a slight modification of the procedure above. Since the two τ leptons from boosted Higgs decay are very close in space, with $\Delta R \sim < 1.0$ (depending on the resonance mass - see Fig. 6.1), the standard procedure using PF anti-k_T jets with $R = 0.5$ can be spoiled because one single jet can contain the decay constituents of two τ leptons. To avoid this problem, we consider subjets found within wide jets as input to the HPS algorithm. These wide jets are clustered using the Cambridge-Aachen algorithm [127], with a distance parameter $R = 0.8$ (CA8). To find the subjets and reconstruct the τ leptons, the following algorithm (based on [89]) is used:

1. Check if the jet has two subjets passing the following criteria (we order them with $m_1 > m_2$, where m_i is the mass of the subjet i):

   - $p_T^{1,2}$ (p_T of subjet 1,2) > 10 GeV.
   - $\Delta R(\text{subjet } 1, \text{subjet } 2) < 0.8$.
   - $\max(m_1, m_2)/m_{jet} < 0.667$.

2. If this pair of subjets pass these criteria, they will be passed to the HPS algorithm.

3. If they do not pass:

   - If $p_T^1 > 10$ GeV: redo the step 1 using subjet 1 as the initial jet.
   - Else, if $p_T^2 > 10$ GeV: redo the step 1 using subjet 2 as the initial jet.

4. The maximum number of iterations is 4.

To estimate the efficiency to find the subjets corresponding to the τ leptons in simulated signal events, for each subjet candidate it is checked if it matches ($\Delta R < 0.15$) with the generator-level visible component (excluding the neutrino) of the τ leptons decaying hadronically. In Fig. 6.4 we show this efficiency for τ leptons with negative and positive charge as a function of the p_T of the generator-level visible τ. For p_T above 40 GeV the efficiency is practically constant of approximately 92%. In the first p_T bin ($20 - 40$ GeV) the efficiency is lower (around 80%). In this region it is common to have cases where some subjets are associated with pile-up or underlying-event constituents, which have in general lower p_T compared with the τ_h decay products from the signal.

After matching the subjets with the generator-level visible τ, the efficiency of the discriminators is estimated using these subjets (see results in Fig. 6.5). For p_T above 40 GeV we observe an efficiency around 88% and for p_T between 20 and 40 GeV we observe an efficiency around 75%, showing the interference from pile-up and underlying-event in the identification.

In the non-boosted regime (two isolated τ leptons), it was observed some problems of energy reconstruction, i.e. in general the energy of the reconstructed τ_h is substimated due to inefficiencies in the energy and tracking measurements and in the offline definition of the τ_h constituents. In order to verify how well the energy of the τ_h decay products is reconstructed in the boosted regime, we plot in Fig. 6.6 the ratio between the histograms of the p_T of the reconstructed τ_h passing the discriminators described above and the p_T of the generator-level visible τ. We observe an excess for the low p_T region, independent of

the τ charge. In Fig. 6.7 we plot the p_T of the leading and subleading (in p_T) reconstructed τ leptons and compare with the respective generator-level visible component. We observe that this excess mainly comes from the subleading τ leptons. In addition, we notice that excess of reconstructed vs. generator-level τ leptons observed in certain p_T bins for the leading τ distribution corresponds to deficit in the subleading one and vice-versa. This can be interpreted as a misassignment of constituents between the two subjets. In case of an imbalance in the p_T of the two τ candidates, it is probable that constituents of the subleading τ are erroneously assigned to the leading one because the decay products of the former have a larger opening angle between them.

Most of the τ_h candidates in signal events have p_T above $50\ \mathrm{GeV}$ so that all these inefficiencies at low p_T do not affect much the total signal efficiency. In the systematic uncertainty section we will present results associated with the pile-up effect, which are very small compared with the leading uncertainties in the analysis and uncertainties associated with the modified method using subjets for boosted τ leptons that represent an important component in this analysis.
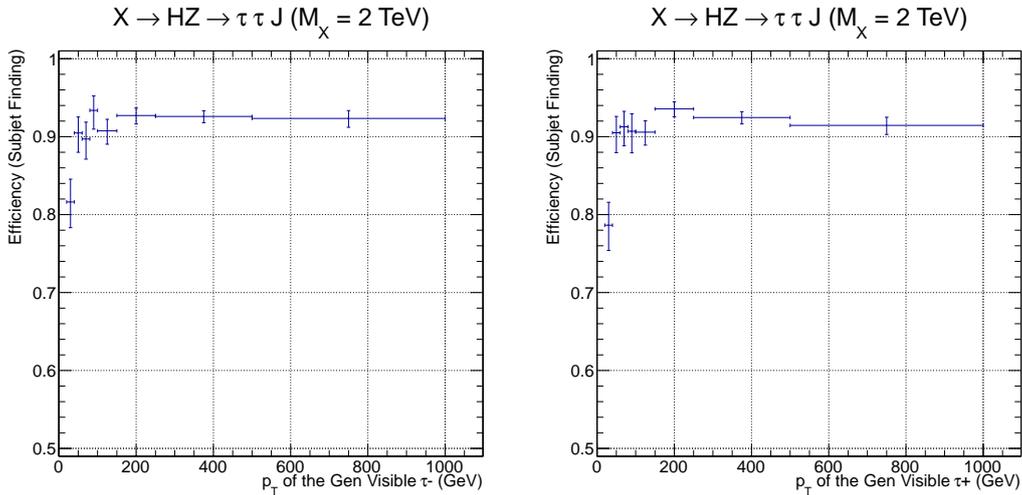
Figure 6.4: Left: efficiency to match a subjet candidate with a generator-level visible τ with negative charge as a function of the p_T. Right: efficiency to match a subjet candidate with a generator-level visible τ with positive charge as a function of the p_T.

6.2.3 Jets

The signature under investigation contains a high-p_T Z boson decaying into a quark pair, which hadronize close to each other, being detected as a single jet in the detector. Through particular studies of jet "grooming" techniques and variables looking into the jet substructure [88, 90, 92], which were used in many analyses in the collaboration (see for example [14, 13, 100]), this single jet might be identified as coming from a Z boson (Z-jet).
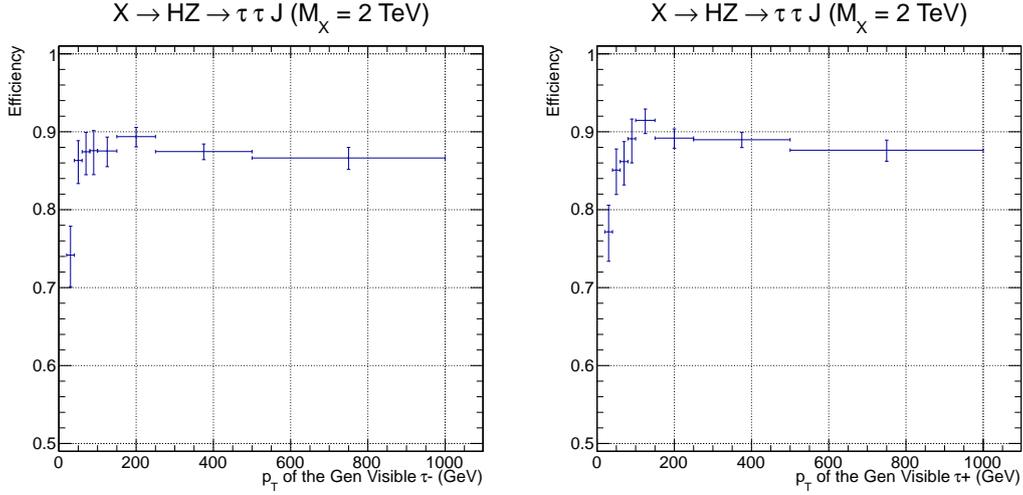
Figure 6.5: Left: decay mode reconstruction plus isolation criteria efficiency as a function of the p_T of generator-level visible τ with negative charge. Right: decay mode reconstruction plus isolation criteria efficiency as a function of the p_T of generator-level visible τ with positive charge.
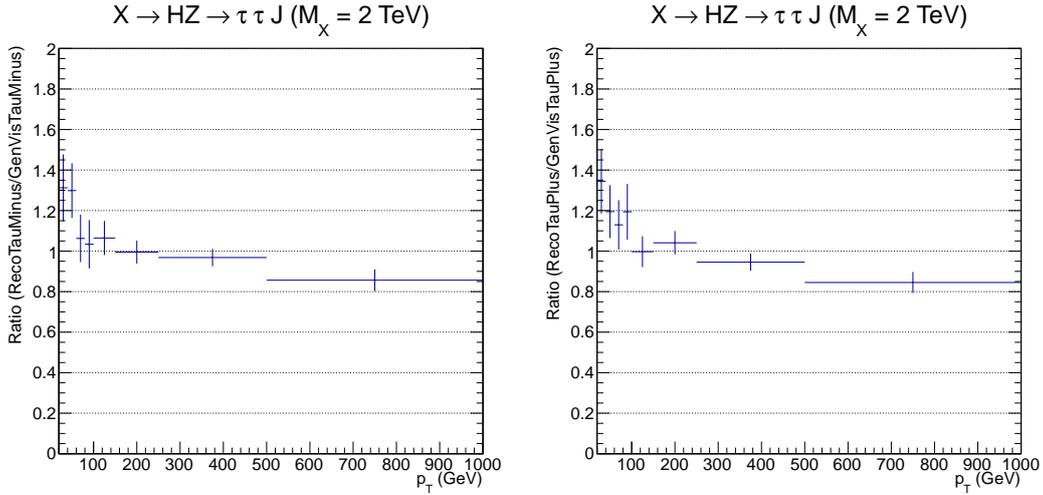
Figure 6.6: Left: ratio plot between the p_T of the τ_h candidate after decay mode reconstruction and isolation criteria and the generator-level visible τ with negative charge. Right: ratio plot between the p_T of the τ_h candidate after decay mode reconstruction and isolation criteria and the generator-level visible τ with positive charge.

In this analysis the jet clustering is performed with the Cambridge-Aachen algorithm with distance parameter $R = 0.8$ using PF constituents. These jets are then re-clustered with a k_T algorithm [128, 129] using the pruning procedure [130, 90], i.e., adding a set of requirements during the clustering in order to remove constituents that are soft (low p_T) or emitted at large angles. The pruning reduces the impact of the underlying-event and pile-up, i.e., particles that are not related to the hard process. As a result, the mass of the pruned-jet is a good discriminant to select Z bosons and reject quark/gluon-jets.
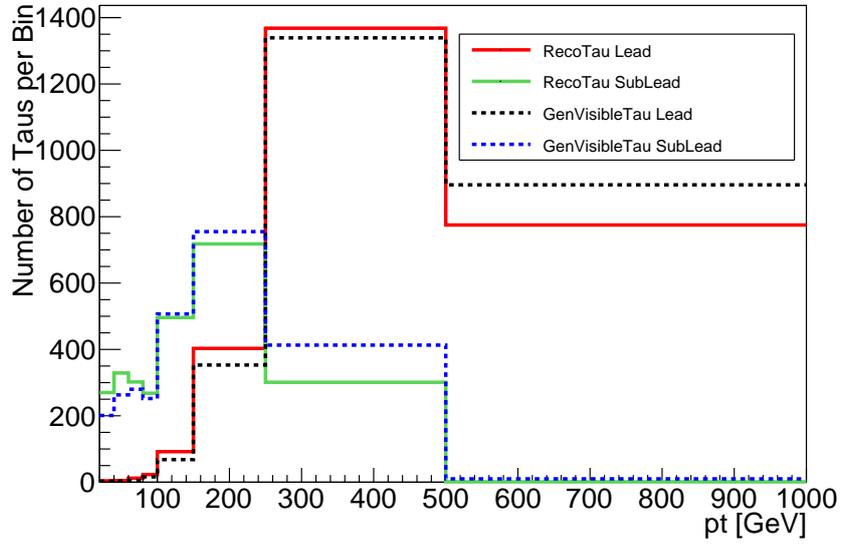
Figure 6.7: p_T of the leading and subleading reconstructed τ candidates and their matched generator-level visible τ leptons.
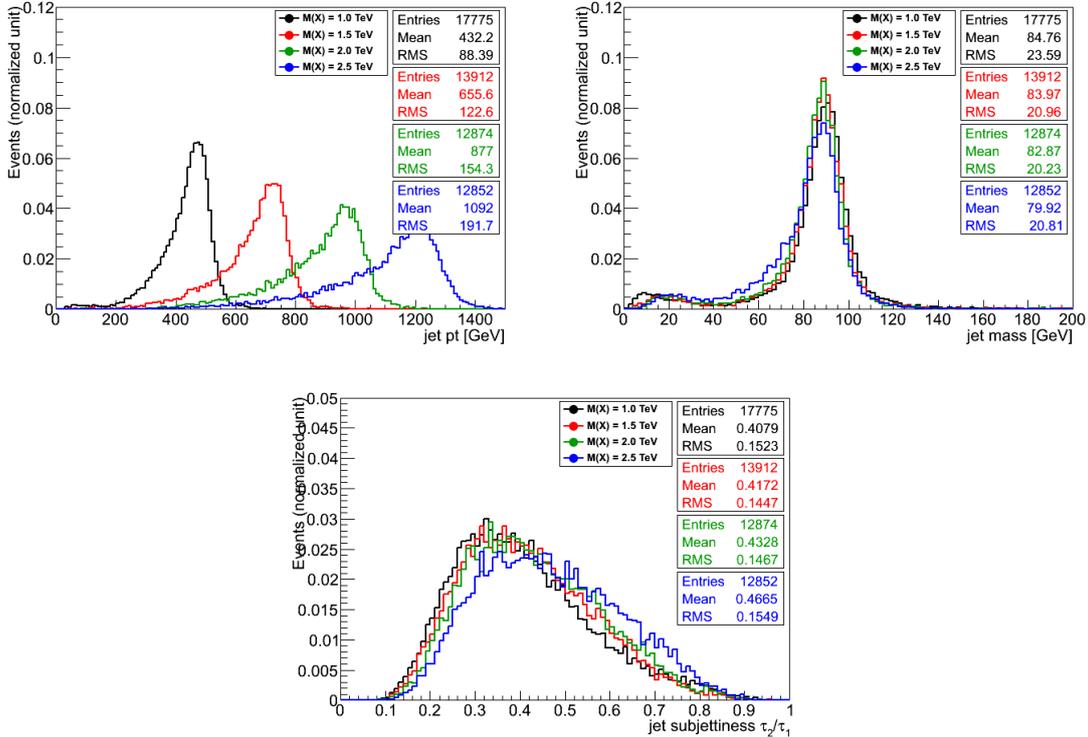
Figure 6.8: Characteristic variables for the jets using four signal mass points ($M_X = 1.0,\ 1.5,\ 2.0,\ 2.5$ TeV). Top left: p_T of the jet. Top right: mass of the jet after pruning procedure. Bottom: ratio of τ_2 over τ_1 (τ_{21}).

Another variable used to discriminate between Z and quark/gluon-jets is the N-subjettiness, which helps discriminate between a jet that is comprised of two subjets and a jet that does
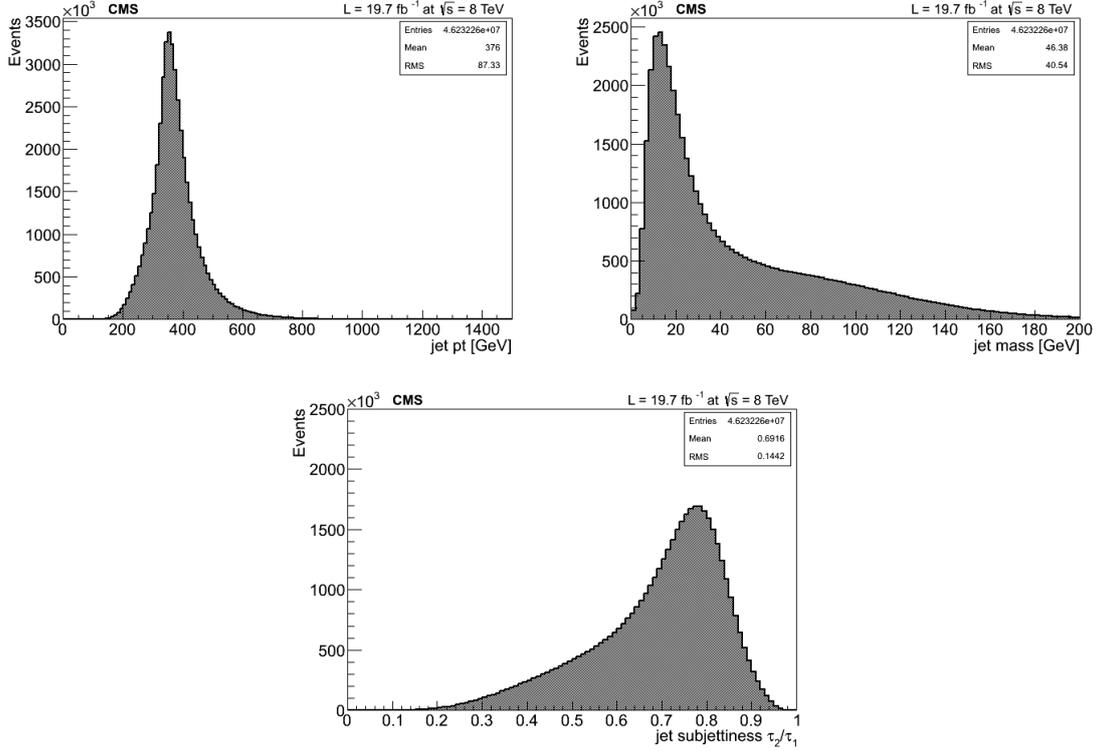
Figure 6.9: Characteristic variables for the jets using observed events with trigger selection applied. Top left: p_T of the jet. Top right: mass of the jet after pruning procedure. Bottom: ratio of τ_2 over τ_1 (τ_{21}).

not present a very well defined substructure. This variable is defined as:

$$\tau_N = \frac{1}{d_0} \cdot \sum_k [p_{T_k} \cdot min(\Delta R_{1,k}, \Delta R_{2,k}, ... \Delta R_{N,k})]$$

where k is the number of constituents and N is the number of subjets under consideration. The final variable used to discriminate Z-jets (that are expected to have two subjets) and quark/gluon-jets (that are expected to have no subjets) is $\tau_{21} = \tau_2/\tau_1$.

In Fig. 6.8 three of the variables used to select the jet coming from the Z boson are plotted for four mass points of the signal. The same variables are shown for data in Fig. 6.9, with only a trigger selection to ensure that the distributions are populated primarily by QCD multijet background events. As shown in the top right plots, the pruned jet mass (m_{jet}^P) is peaked at the mass of the Z boson for the signal and at a lower value (around 20 GeV) for the quark/gluon-jets in data events. The subjettiness variable, $\tau_{21} = \tau_2/\tau_1$, peaks at ~ 0.35 for the signal (see Fig. 6.8 bottom) and at ~ 0.8 for the background (see Fig. 6.9 bottom). The final selection on the jets is listed below:

- $p_T > 400$ GeV (to match the trigger threshold).

- $|\eta| < 2.4$ (to match tracker acceptance).

- $70 < m_{jet}^P < 110$ GeV (from previous analysis with boosted $Z \to q\bar{q}$ [14]).

- $\tau_{21} \equiv \tau_2/\tau_1 < 0.75$ (described below).

- Jet identification (loose working point):

  - muon energy fraction of the jet < 0.99.

  - photon energy fraction of the jet < 0.99.

  - charged EM energy fraction of the jet < 0.99.

  - neutral hadron energy fraction of the jet < 0.99.

  - charged hadron energy fraction of the jet > 0.00.

  - number of constituent particles larger than 1.

- $R_{CN} > 2$ for jets with $1.0 < |\eta| < 1.5$, where R_{CN} is the ratio between the multiplicity of charged PF candidates over neutral PF candidates.

The selection on the τ_{21} variable is decided looking at the expected upper limit in the cross-section of a possible signal for two values: 0.5 and 0.75 [3]. The selection with $\tau_{21} <$ 0.75 showed a more stringent limit, probing lower values of possible signal cross-sections. The requirement on R_{CN} is applied in order to remove jets where the tracking software behaves in an anomalous way, and many fake tracks are wrongly associated with the jet [14].

In addition to this selection, we apply standard CMS corrections for the jet energy calibration and transverse momentum resolution [131].

6.2.4   Missing transverse energy ($|\vec{\not{E}}_T|$)

The missing transverse emergy is constructed using particle-flow (PFMET), computed as the negative vector sum of all particles' transverse momentum. On top of the raw PFMET two kinds of corrections are considered: one that propagates the jet energy corrections (JEC) to the $|\vec{\not{E}}_T|$ computation, and other that eliminates the observed modulation of the $|\vec{\not{E}}_T|$ in the xy plane[4].

6.2.5   H $\to \tau\tau$ reconstruction

To estimate the Higgs mass two methods are tested:

- Visible mass (M_{vis}): The invariant mass of the detectable products (it does not consider the neutrinos).

- SVFit mass (M_{svfit}): When the τ lepton decays there are neutrinos that scape detection, then the direct Higgs mass reconstruction using standard objects is not possible. One possibility is to estimate the neutrinos contribution constructing and maximizing a likelihood based on the kinematics of the tau decay [132]. Using as constraint the quantities that we can measure, it is possible to do an estimate of the four-momentum of the Higgs. The SVFit mass is given by the maximization of a likelihood with respect the free parameters that correspond to the neutrinos three-momenta. The inputs of SVFit are:

---

[3]These two values were used to construct two categories in the previous analysis with boosted $Z \to q\bar{q}$ [14].

[4]These corrections are standard in CMS https://twiki.cern.ch/twiki/bin/viewauth/CMS/EXOMissingET2012 and https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis.

   – the two 3-momenta of the two visible objects (τ_h).

   – x-y components of the PFMET.

   – the covariance matrix of the PFMET.

In Fig. 6.10 the visible and SVFit masses are shown for signal events. Using the SVFit approach the signal distribution has much better resolution. For both cases there are some events at masses below $5\,\mathrm{GeV}$, mainly for the lower resonance masses $1.0 - 1.5\,\mathrm{TeV}$. This comes from events where $\Delta R(\tau, \tau)$ is larger than 0.8. In these cases the reconstruction procedure selects only one true τ, while the other is a fake object. The combination of the fake with the true τ results in a very small mass for the Higgs candidate.

In Fig. 6.11 we show the Higgs p_T as determined by the SVFit.
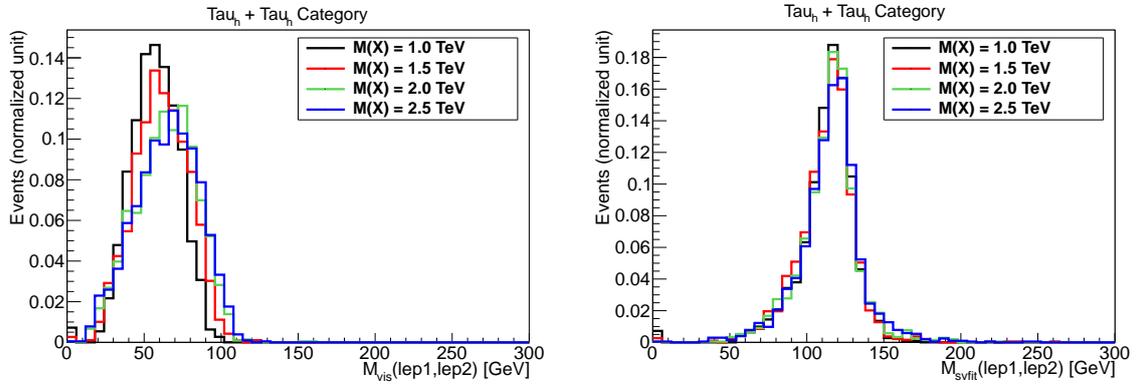
Figure 6.10: Reconstructed Higgs mass. Left: visible mass. Right: SVFit mass.
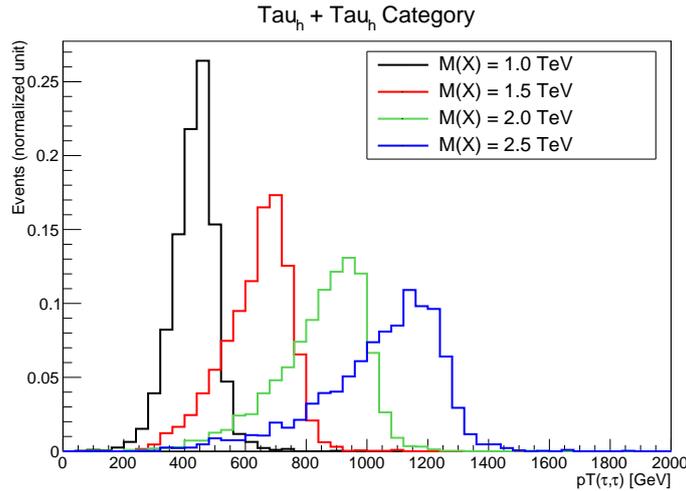
Figure 6.11: Reconstructed Higgs p_T from SVFit method for signals with $\mathrm{M_X} = 1.0, 1.5, 2.0,$ and $2.5\,\mathrm{TeV}$.

6.2.6 Resonance mass reconstruction

The resonance mass is estimated from the sum of the four-momentum of the Z-jet and the Higgs. Since the SVFit gives a better resolution for the Higgs mass compared to the visible mass, the Higgs four-momentum is taken from SVFit. In Fig. 6.12 we show the resonance reconstructed mass.
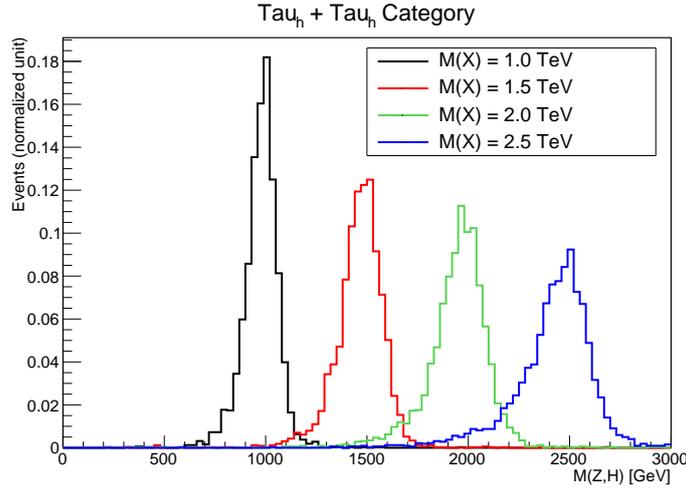
Figure 6.12: Reconstructed resonance mass for signals with M_X = 1.0, 1.5, 2.0, and 2.5 TeV. The Higgs component is estimated with the SVFit method.

6.3 Event selection

An event is selected only if it passes several filters: *primaryVertexFilter*, *noscrapingFilter*, *hcalLaserEventFilter*, *HBHENoiseFilter*, *trackingFailureFilter*, *CSCTightHaloFilter*, *eeBadScFilter*, *EcalDeadCellTriggerPrimitiveFilter*. Those filters are detailed in Ref. [14] and basically improve the quality of the event reconstruction (tracks, vertices, and energy clusters) avoiding, for example, detector interferences due to electronic noises.

In order to extract the signal, we optimize a kinematical selection. A signal event is divided into two "hemispheres" in the detector: one represented by a jet from the high-p_T $Z \to q\bar{q}$ decay and the other with two τ_h decay products plus $|\vec{\rlap{E}/}_T|$ from high-p_T $H \to \tau^- \tau^+$ decay.

6.3.1 Baseline selection

We select at least two τ candidates passing the reconstruction and identification requirements and with the kinematic cuts:

- $p_T > 20$ GeV.

- $|\eta| < 2.3$.

- The leading and subleading (in p_T) τ_h are required to have $\Delta R(\tau_h, \tau_h) < 1.0$.

Jets passing the jet identification criteria and missing transverse energy have the following kinematical selection:

- $p_T^{\text{jet}} > 400$ GeV and $|\eta^{\text{jet}}| < 2.4$.

- PFMET > 40 GeV.

The requirement of at least one jet with $p_T > 400$ GeV is due to the trigger threshold; for that range the trigger efficiency in observed events is practically constant and above 99%.

Combining the four momentum of the Z-jet and the Higgs we reconstruct the resonance candidate. Since we are studying boosted topology we require:

- $m_{\text{ZH}} > 800$ GeV.

In order to have a clear separation between the two "hemispheres" and remove events with spurious combinations of jets, τ_{h} and $|\vec{\not{E}}_T|$ we require:

- $|\Delta\phi(\text{jet}, \vec{\not{E}}_T)| > 2.0$.

- $|\Delta\phi(\tau_{\text{h}}, \vec{\not{E}}_T)| < 1.5$.

To avoid overlap with semileptonic and all-leptonic channels, events with at least one isolated muon or electron are vetoed.

In Fig. 6.13 we show the number of events after each criterion described below in signal events for mass points of $1.0, 1.5, 2.0,$ and 2.5 TeV. The description of each entry is as follows:

- **Pre-filter:** The first bin corresponds to the number of events that passed the pre-filters described at the beginning of this section. Those pre-filters have a very high efficiency around $99 - 98\%$.

- **Leptonic veto:** The second bin shows the number of events that pass the electron and muon veto. At generator level one would expect approximately 42% of the events in the first bins to survive the leptonic veto (i.e. the branching fraction of the two τ leptons to decay hadronically). But given the inefficiencies in the identification and kinematic cuts, around 50% of the events pass this veto.

- $|\vec{\not{E}}_T|$ **requirement:** The third bin is the PFMET > 40 GeV requirement, which has an efficiency always above 90%. The efficiency increases for higher resonance masses due to the higher energy of the Higgs boson from the resonance decay.

- **Jet requirement:** The fourth bin is the requirement of the presence of at least one jet with $p_T > 30$ GeV and accepted by the jet identification criteria; as expected it has a very high efficiency, around 99%.

- $\vec{\not{E}}_T$ **angular threshold:** The fifth bin is labeled "dPhiMETcut" and it represents the efficiency of the PFMET $\Delta\phi$ cut: since in signal $\Delta\phi(\vec{\not{E}}_T, \tau_{\text{h}}) < 1.5$ in practically 100% of the cases, using true quantities, an inefficiency in this requirement is indicative of a failure in the subjet finding. The efficiency due to this matching varies with the mass points in the range 77–91%. For 1 TeV the efficiency is lowest. The explanation to

this fact can be understood from the plot in Fig. 6.1 showing that for this mass point
there is a considerable number of events with $\Delta R(\tau,\tau) > 0.8$; since subjets are con-
structed from the components of CA8 jets, this decreases the subjet finding efficiency.
For higher masses the trend is a decrease of the inefficiency due to this effect. For the
2.5 TeV there is again some loss of efficiency compared with the other masses; this
is mainly due to the cases where there is one subjet containing both τ leptons and the
other subjet is mainly due to pile-up and/or underlying-event processes.

- **Tau p_{T} requirement:** The sixth bin corresponds to the number of events after the im-
  position of p_{T} thresholds to select two τ_{h} candidates; values are presented for thresh-
  olds of 20 and 40 GeV. As expected, those selection increase in efficiency for higher
  resonances masses; efficiencies are in the range of $58-67\%$ (for 40 GeV) and $73-79\%$
  (for 20 GeV).

- τ_{h} **requirements:** The remaining bins are relative to the τ_{h} discriminators (τ decay
  mode reconstruction, isolation, electron/muon rejection) and the selection in $\Delta R(\tau_{\mathrm{h}}, \tau_{\mathrm{h}}) < 1$. The total efficiency of these requirements is in the range of $63 - 73\%$ and
  depends on the signal mass point and p_{T} threshold.
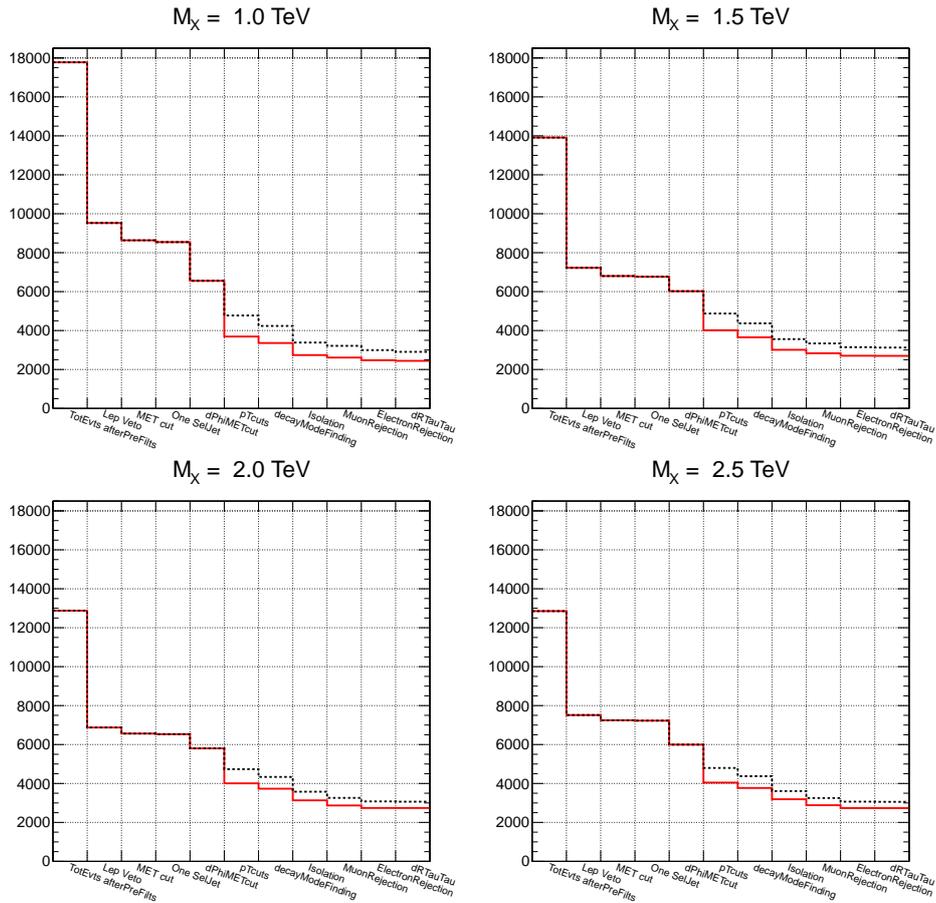
Figure 6.13: Cut flow for the baseline selection using four different signal mass points and
two p_{T} cuts to select the τ pairs ($p_{\mathrm{T}} > 20$ GeV, black dotted and $p_{\mathrm{T}} > 40$ GeV, red solid).

For both QCD and signal events the hadronic τ discriminators efficiency are investigated on top of the baseline selection and in different energy regimes: $(800 < M_X < 1200 \text{ GeV})$ and $(1600 < M_X < 2400 \text{ GeV})$, where M_X is defined as the mass of the reconstructed high-mass resonance candidate. In Fig. 6.14 we show these efficiencies for signal events and in Tab. 6.5 for QCD. We observe efficiencies between 73 and 76% for the signal and of the order of 10^{-3} for QCD in both mass regions.
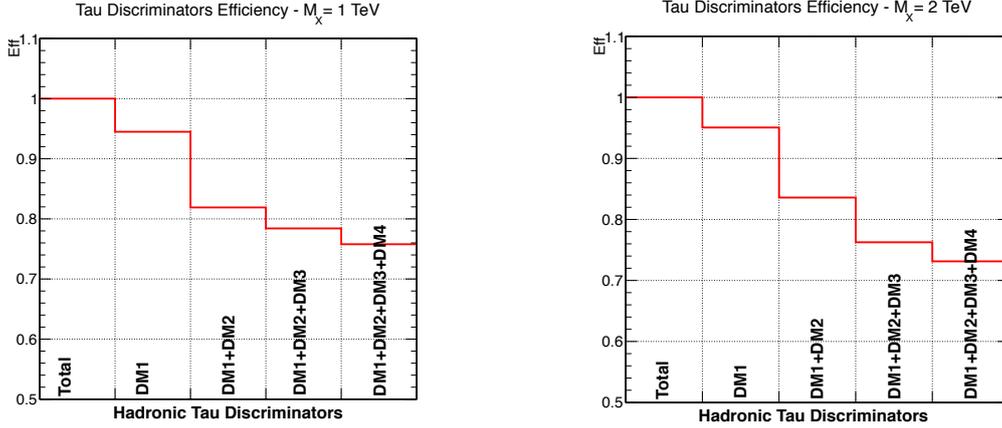
Figure 6.14: Cumulative efficiencies for τ_h reconstruction and identification criteria in signal events after aplying the baseline selection. DM1 refers to the decay mode reconstruction criteria, DM2 is the isolation requirement, and DM3/DM4 are the muon/electron rejection discriminator. Left: efficiencies for $800 < M_X < 1200$ GeV. Right: efficiencies for $1600 < M_X < 2400$ GeV.

Table 6.5: Efficiencies for τ_h reconstruction and identification criteria in QCD events in two different energy scales. Efficiencies are computed on top of the baseline selection.

Mass region (GeV)	QCD efficiency in %
$800 < M_X < 1200$	0.08
$1600 < M_X < 2400$	0.56

6.3.2 Cut optimization

On top of baseline selection additional cuts on the p_T of the τ leptons, SVFit $\tau\tau$ mass ($m_{\tau,\tau}$), and missing transverse energy are investigated to optimize signal over background.

In Fig. 6.15 it is shown the p_T of the leading τ_h in signal events (left) and QCD (right). The signal distributions peaks always for p_T above 100 GeV and in QCD at approximately 30 GeV. A supplementary requirement on the leading τ_h p_T at $p_T > 50$ GeV is defined.

Applying the baseline selection plus the boosted Z selection and the cut on the leading τ_h p_T above 50 GeV, we investigate the best selection on $m_{\tau,\tau}$ looking the Punzi factor of merit ($\mathcal{P}$), defined as: $\mathcal{P} = \epsilon_{sig}/(1 + \sqrt{B})$ [133], where ϵ_{sig} is the signal efficiency and B is the background yield after applying the selection. In Fig. 6.16 $m_{\tau,\tau}$ distributions for background and signal simulation are presented. For the optimization, a scan in the $m_{\tau,\tau}$ variable is performed, using a lower cut range between 60 and 115 GeV and an upper cut
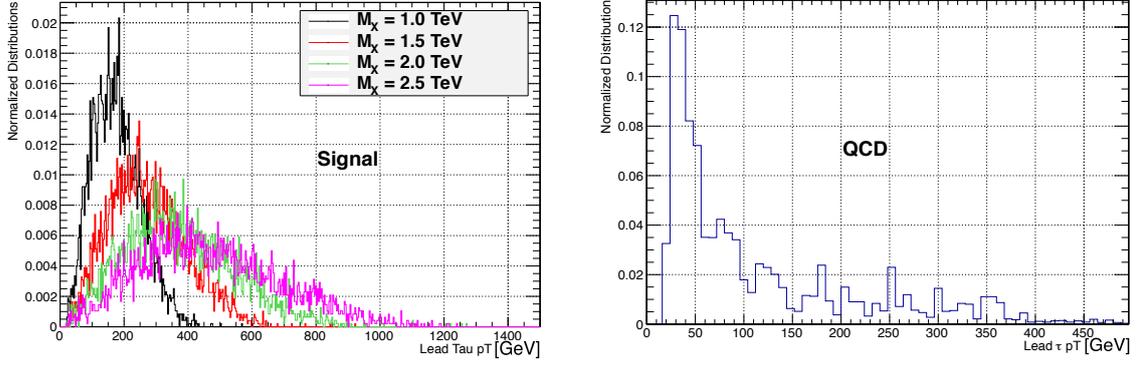
Figure 6.15: Left: leading τ_h p_T for signal events after baseline selection. Right: leading τ_h p_T in QCD events after baseline selection.

range of $150-180\,\mathrm{GeV}$, both varied in steps of $5\,\mathrm{GeV}$, and testing all possible combinations. In Tab. 6.6 we present all the selections investigated with the respective x-axis label used in the plots of the Punzi's significance in Fig. 6.17. In Fig. 6.18 the efficiencies in signal and background events are shown for the same selections defined in Tab. 6.6. For all the resonance mass points we observe a maximum in the Punzi significance in a similar region of $m_{\tau,\tau}$, which corresponds to the range $(m_{\tau,\tau}^\mathrm{lower} = 105)\,\&\&\,(150 < m_{\tau,\tau}^\mathrm{upper} < 180)\,\mathrm{GeV}$ (see Fig. 6.17). In order to get the best signal efficiency we decided to use the cut of $105 < m_{\tau,\tau} < 180\,\mathrm{GeV}$.

Table 6.6: SVFit $\tau\tau$ mass windows used in the selection optimization. The column with the labels is showing the x-axis for the plots in Figs. 6.17 and 6.18.

SVFit $\tau\tau$ mass requirements (GeV)	Labels for the plots
$m_{\tau,\tau}^\mathrm{lower} = 60$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$1-7$
$m_{\tau,\tau}^\mathrm{lower} = 70$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$8-14$
$m_{\tau,\tau}^\mathrm{lower} = 75$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$15-21$
$m_{\tau,\tau}^\mathrm{lower} = 80$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$22-28$
$m_{\tau,\tau}^\mathrm{lower} = 85$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$29-35$
$m_{\tau,\tau}^\mathrm{lower} = 90$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$36-42$
$m_{\tau,\tau}^\mathrm{lower} = 95$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$43-49$
$m_{\tau,\tau}^\mathrm{lower} = 100$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$50-56$
$m_{\tau,\tau}^\mathrm{lower} = 105$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$57-63$
$m_{\tau,\tau}^\mathrm{lower} = 110$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$64-70$
$m_{\tau,\tau}^\mathrm{lower} = 115$ and $m_{\tau,\tau}^\mathrm{upper} = 150,\ 155\ldots180$	$71-77$

Adding the $m_{\tau,\tau}$ selection defined above, the background is drastically reduced. In particular, there is no events from MC simulation of QCD. If we investigate the PFMET distributions comparing the MC simulation of signal with background, we would conclude that it is not needed to apply a tighter selection on PFMET. However, from a data-driven background estimate it seems that there are QCD events in the signal region.

As described before, after $m_{\tau,\tau}$ selection, we do not have QCD events in MC simulation. It is possible to study the efficiency of PFMET selections in QCD events by defining a QCD enriched sample with same kinematic selection as in signal region, but with at least one τ
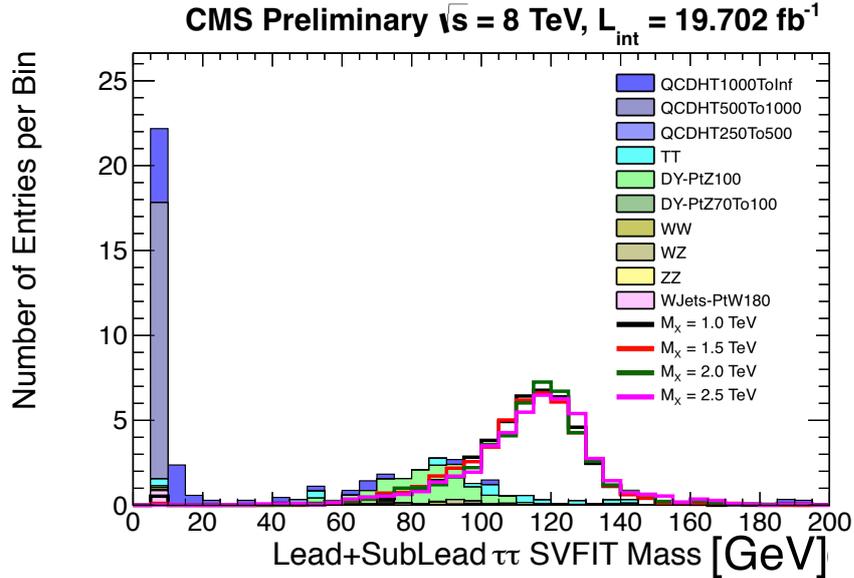
Figure 6.16: SVFit $\tau\tau$ mass distributions in background and signal events using the baseline selection plus the Z-jet selection and the p_T of the leading τ_h above 50 GeV. The signal distributions are scaled to have the same area as in the background.

candidate failing the isolation criterion. In Fig. 6.19 we show PFMET distributions in MC simulation and observed events using the criteria above. Observed events with the selection above and after subtraction of the other background components from MC simulation ($t\bar{t}$, Z/W+jets, SM dibosons) are then compared with signal events, as done for the SVFit $\tau\tau$ mass. In Figs. 6.20 and 6.21 we show the Punzi significance and signal efficiency versus background rejection, respectively. Analyzing the curves of Figs. 6.20 and 6.21, it looks resonable to use different PFMET selections for each resonance mass; on the other hand, since this control sample represents only the QCD component of the background, it is important to investigate which is the optimal selection using all background components.

We compare signal events with the background prediction from a data-driven method [5] considering different PFMET thresholds. The optimal cut is defined by the one with the best (or lower) expected upper limit in a possible signal cross-section (see Tab. 6.7). It is observed that the optimal selection in PFMET varies with the resonance mass. On the other hand, except for the 2.5 TeV mass point, the relative difference of the optimal expected limit compared with that obtained with a constant threshold of 80 GeV is always below 25%. In addition, tightening the PFMET threshold might lead to a zero background expectation, resulting in higher systematic uncertainties from the background estimation method. In view of the above, a compromise between good signal efficiency for all mass points and high power of QCD rejection, a final requirement PFMET > 80 GeV was chosen.

In Tab. 6.8 we list the full selection defined above. These criteria define the "signal-region" and it will be used for the final comparison between expected background and ob-

---

[5]This data-driven method will be described later in Sec. 6.5. It estimates all the background components using data samples.
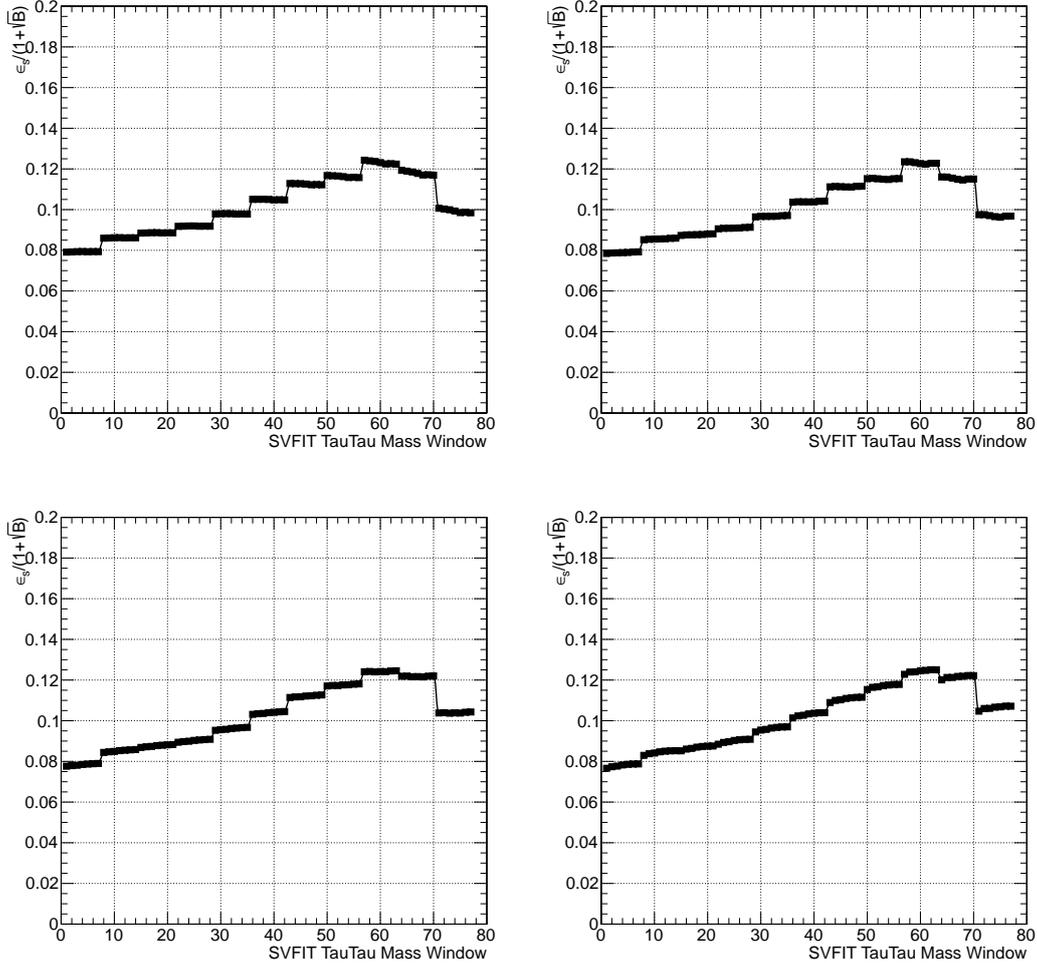
Figure 6.17: Punzi's significance as function of the SVFit $\tau\tau$ mass selection (see Tab. 6.6 to refer to the x-axis labels in the plot). We consider signals with 1.0 TeV (top left), 1.5 TeV (top right), 2.0 TeV (bottom left), and 2.5 TeV (bottom right).

Table 6.7: Number of background events estimate from ABCD method (see Subsec. 6.5) for $19.7/\text{fb}$ and using the full selection (column BG). We present as well the number of events for the signal (signal generated luminosity) and the expected upper limits in the cross-section in pb, comparing the background estimate and the signal output (no systematic uncertainties included). The red numbers correspond to the best expected upper limits.

PFMET	Number of events					Exp. limit: $pp \to \text{X} \to \text{H}(\tau\tau)\text{Z}(qq)$ [pb]			
	BG	1TeV	1.5TeV	2TeV	2.5TeV	1TeV	1.5TeV	2TeV	2.5TeV
> 40	8.0	1547	2563	2794	2625	0.0039	0.0023	0.0022	0.0023
> 80	6.1	1371	2373	2669	2519	**0.0037**	0.0023	0.0021	0.0021
> 170	2.1	789	1762	2225	2241	0.0048	**0.0021**	**0.0017**	0.0017
> 200	1.9	597	1555	2070	2133	0.0063	0.0025	0.0019	0.0018
> 300	0.7	140	887	1455	1711	0.0203	0.0033	0.0020	0.0017
> 400	0	2	362	886	1246	0.9600	0.0053	0.0021	**0.0015**
> 450	0	0	215	684	1046	–	0.0089	0.0029	0.0018

64
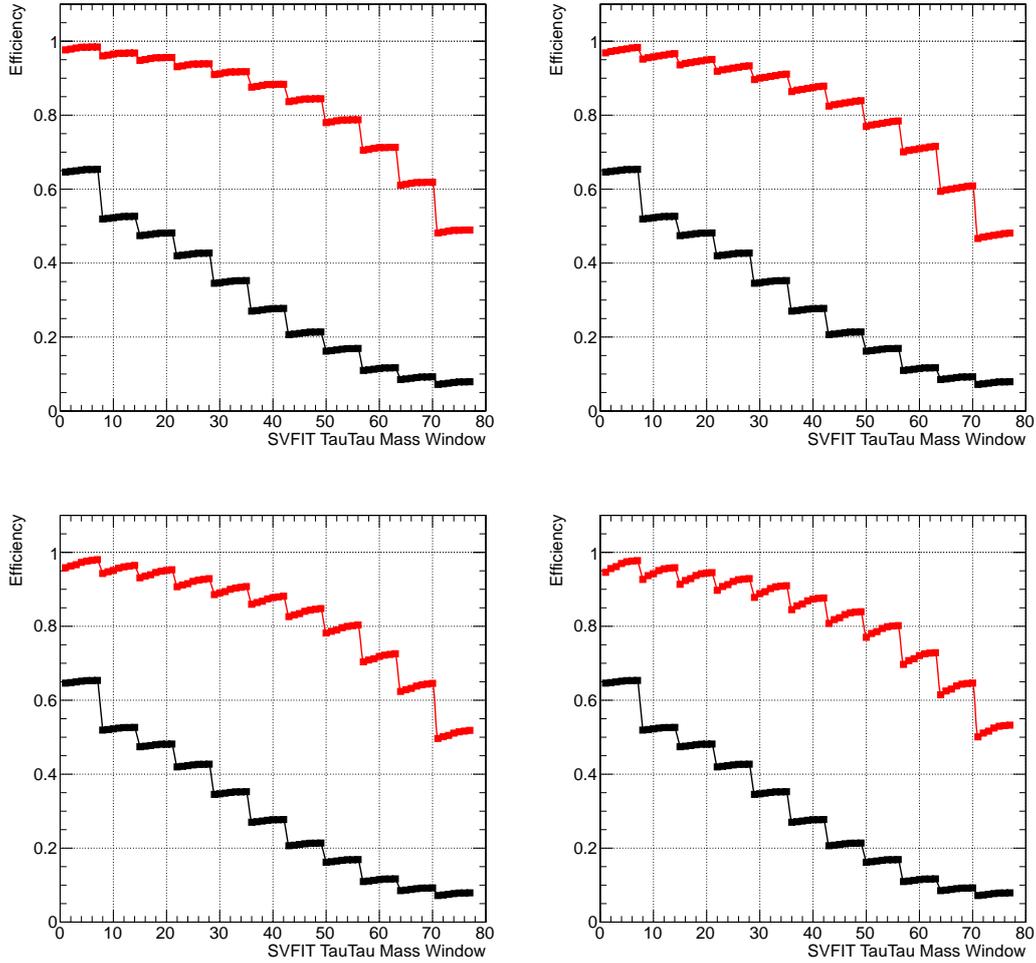
Figure 6.18: Efficiency in signal (red curves) and background (black curves) as a function of the SVFit $\tau\tau$ mass cuts (see Tab. 6.6 to refer to the x-axis labels in the plot). We consider signals with 1.0 TeV (top left), 1.5 TeV (top right), 2.0 TeV (bottom left), and 2.5 TeV (bottom right).

served events. In Tab. 6.9 we show the signal efficiency for this selection.

6.4 Comparison between observed events and MC simulations

In Tab. 6.8, in addition to the full selection, we list the baseline criteria. This last selection is useful for comparison between MC simulations and observed events because it results in higher number of events in both samples, and consequently well defined distributions. The full selection results in a very small number of expected events (order of 10 events for $L_{\text{int}} \sim 20/\text{fb}$).

In Figs. 6.23–6.25 some control plots comparing observed events and MC simulation are shown. The distributions were done after the baseline selection and in signal-free regions, i.e.

65
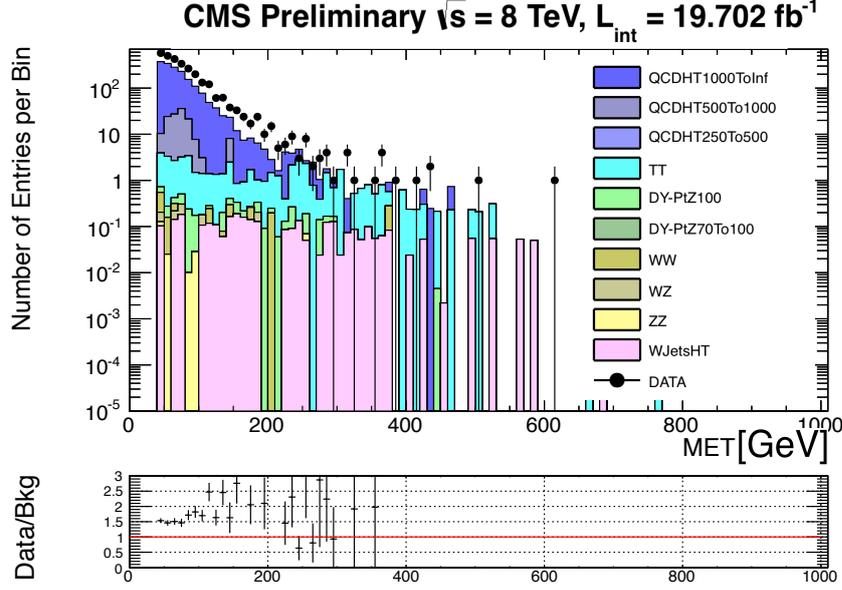
Figure 6.19: PFMET distribution in a QCD enriched sample defined with the full selection with at least one τ failing the isolation criterion.

Table 6.8: Summary of the event selection.

	Selection	Comments
	At least two identified τ_h candidates	
	$p_T^{lead} > 50$ GeV and $p_T^{sublead} > 20$ GeV	Baseline selection uses $p_T^{lead} > 20$ GeV
Taus	$|\eta| < 2.3$ and $|\Delta\phi(\vec{\not{E}}_T, \tau_h)| < 1.5$	
	$\Delta R(\tau_h, \tau_h) < 1.0$	
	$105 < m_{\tau,\tau} < 180$ GeV	Not applied in baseline selection
	At least one jet passing the Loose Jet ID	
Jets	$p_T^{lead} > 400$ GeV	Baseline selection uses $p_T^{jets} > 30$ GeV
	$|\eta| < 2.4$ and $|\Delta\phi(jet, \vec{\not{E}}_T)| > 2.0$	
	$70 <$ pruned jet mass < 110 GeV and $\tau_{21} < 0.75$	Not applied in baseline selection
	Veto the event if at least one isolated e/μ	
Events	PFMET > 80 GeV	Baseline selection uses PFMET > 40 GeV
	$m_{ZH} > 800$ GeV	

using only events that fail the pruned jet mass window requirement $70 < m_{jet}^P < 110$ GeV. This requirement is in agreement with the blinding strategy used in this analysis to avoid bias from anticipated observation of events in signal region. We observe a reasonable description of the observed events by the MC simulations, but it is possible to visualize some discrepancies. It must be stressed that, even at this level of selection, the low amount of events in the simulated samples already leads to discrepancies (mainly in the tail of the distributions), particularly in the QCD simulation; this generates spikes that appear as considerable deviations
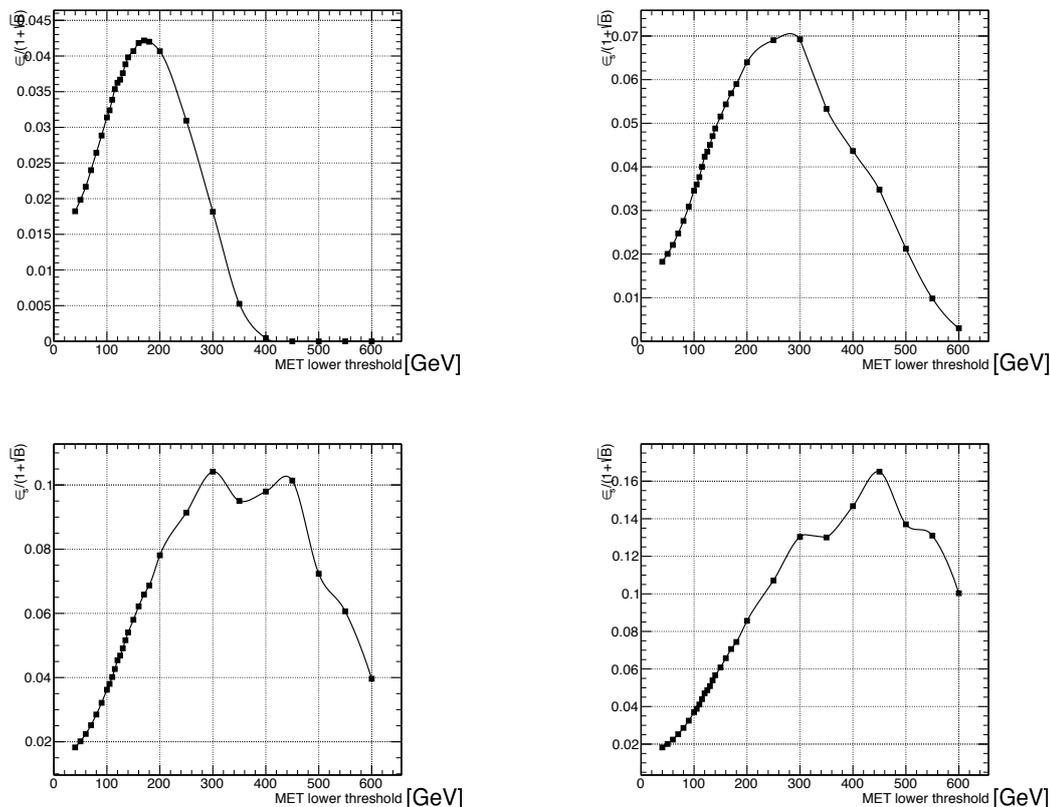
Figure 6.20: Punzi's significance as function of the PFMET threshold. The background is represented by a QCD enriched sample from observed events. We consider signals with $1.0\,\mathrm{TeV}$ (top left), $1.5\,\mathrm{TeV}$ (top right), $2.0\,\mathrm{TeV}$ (bottom left), and $2.5\,\mathrm{TeV}$ (bottom right).

Table 6.9: Number of signal events generated and signal yield after full selection (defined in Tab. 6.8). Dividing these values we obtain the total signal efficiency, including the branching fraction of the two τ leptons to decay hadronically. The last column shows the final efficiency of the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel.

Resonance mass (TeV)	Signal events generated	Signal yield	Final efficiency (%)
1.0	17958	1359	18.0
1.5	14099	1726	29.1
2.0	13080	1751	31.9
2.5	13091	1653	30.1

from the observed events.

QCD multijets is the highest background component when using the baseline selection. This background has cross sections calculated only at leading-order, which means we expect big uncertainties on the normalization of this process. It is possible to correct this normalization applying a k-factor estimated from a QCD enriched sample with the same selection as the baseline, but asking at least one of the τ_h candidates to fail the isolation discriminator. In Fig. 6.22 η distributions for the selected τ_h candidates and jets in the QCD enriched sample are shown. From this sample, a k-factor of 1.482 ± 0.003 is obtained comparing the
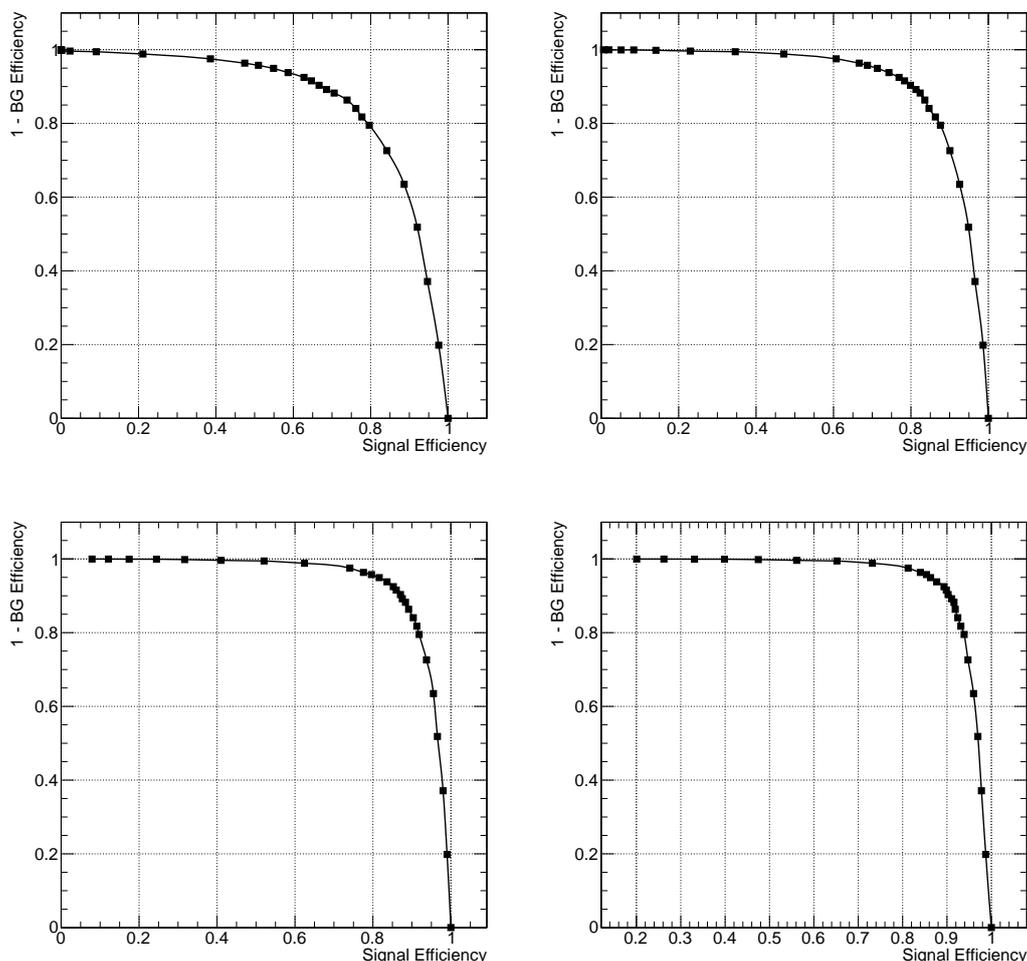
Figure 6.21: Background rejection as function of the signal efficiency for each PFMET threshold (the same cuts as in x-axis of Fig. 6.20). The background is represented by a QCD enriched sample from observed events. We consider signals with 1.0 TeV (top left), 1.5 TeV (top right), 2.0 TeV (bottom left), and 2.5 TeV (bottom right).

normalizations of MC simulation of QCD multijets with the observed events subtracted by the other background components from MC simulation (mainly W+jets and $t\bar{t}$).

6.5 Background estimation

This analysis has followed a "blinding policy", i.e., first we have defined and validated all the background estimate techniques looking only in kinematic regions that are expected to not have signal events. In these regions we have compared simulations with background estimates using observed events. After this validation we have "unblinded" the analysis, comparing our estimates with the observed events in the signal region.

In analyses that search for exotic signal in a specific kinematic region it is common to define a "signal-free region", denoted here by side-band (SB), from where it is possible to "propagate" the number of events in background to the signal-region (SR). In general these
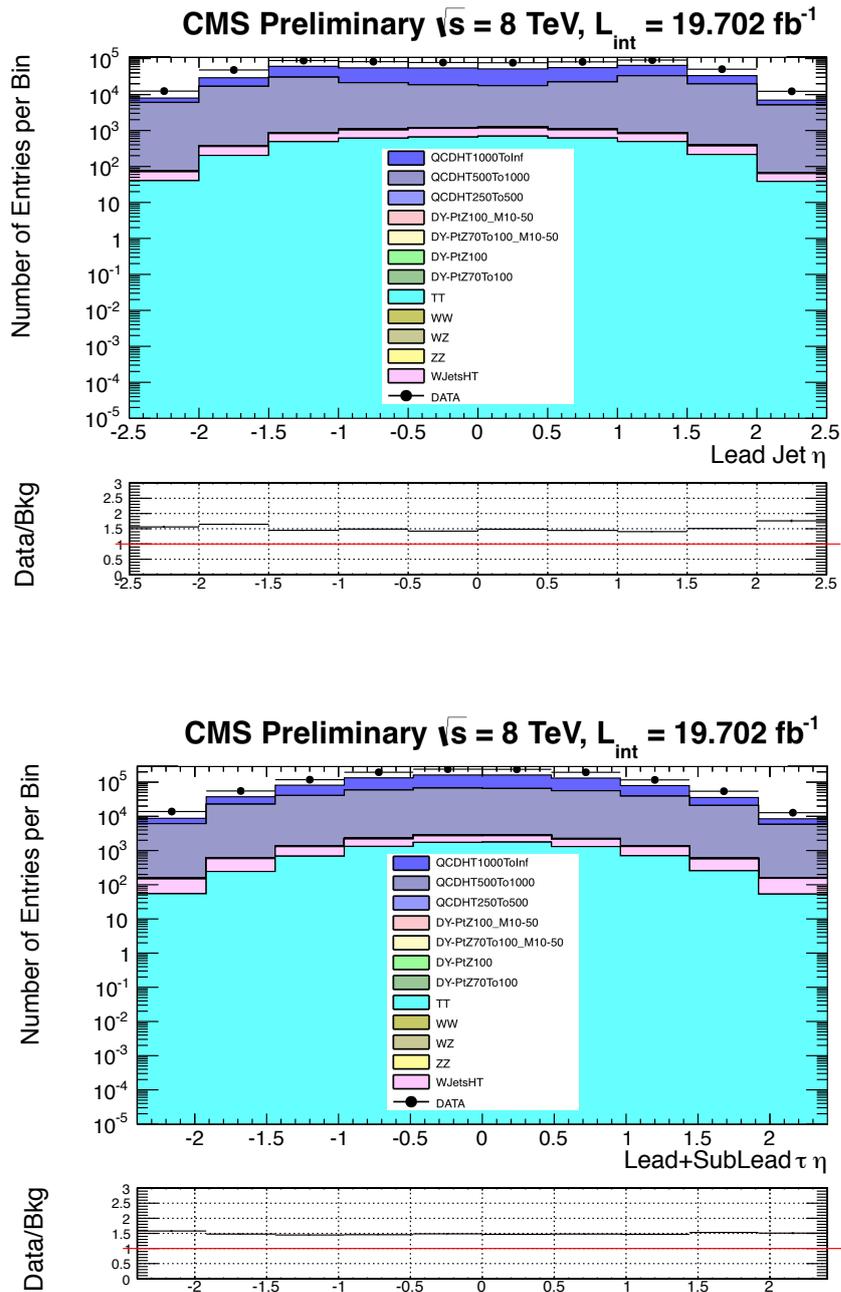
Figure 6.22: Distributions in a QCD enriched sample used to extract the QCD k-factor for the control plots in Figs. 6.23–6.25. Top: Jet candidates η. Bottom: τ_{h} candidates η.

SB regions are used to find both the normalization and shape of the background distributions in SR.
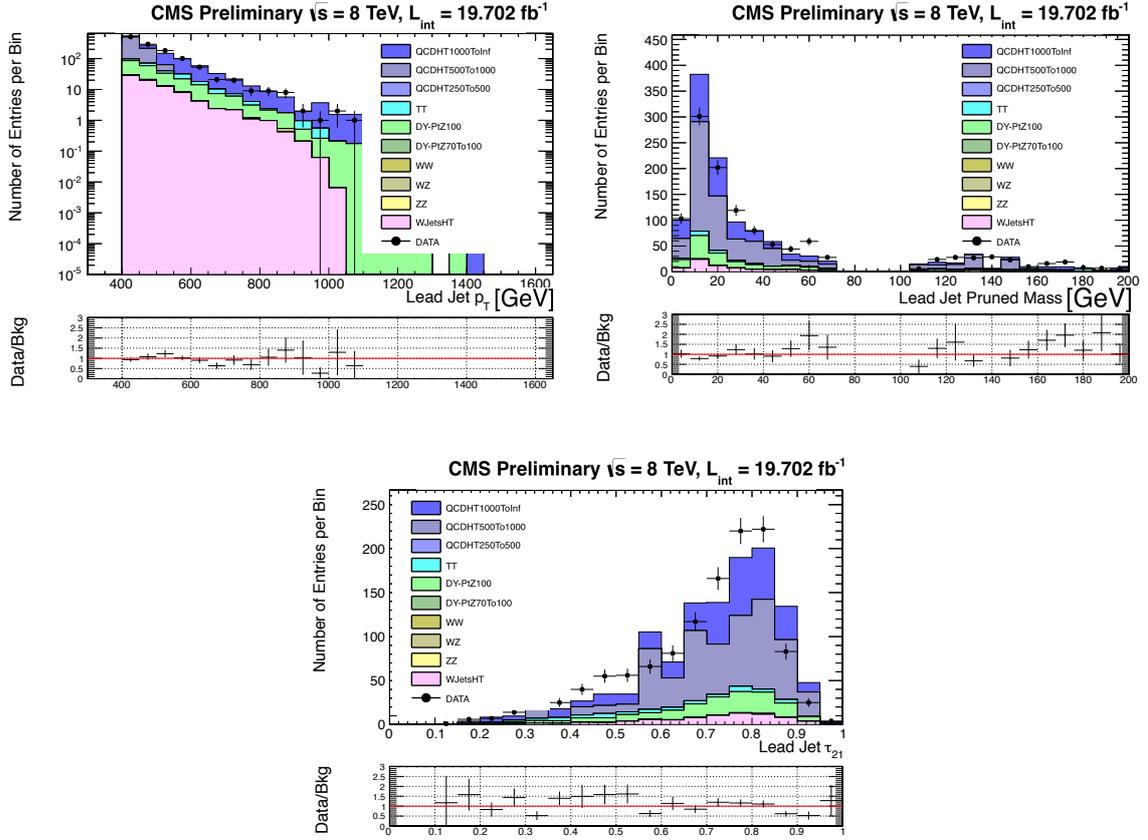
Figure 6.23: Comparison between MC simulation and observed events. Top left: leading jet p_T. Top right: jet pruned mass. Bottom: τ_{21} variable.
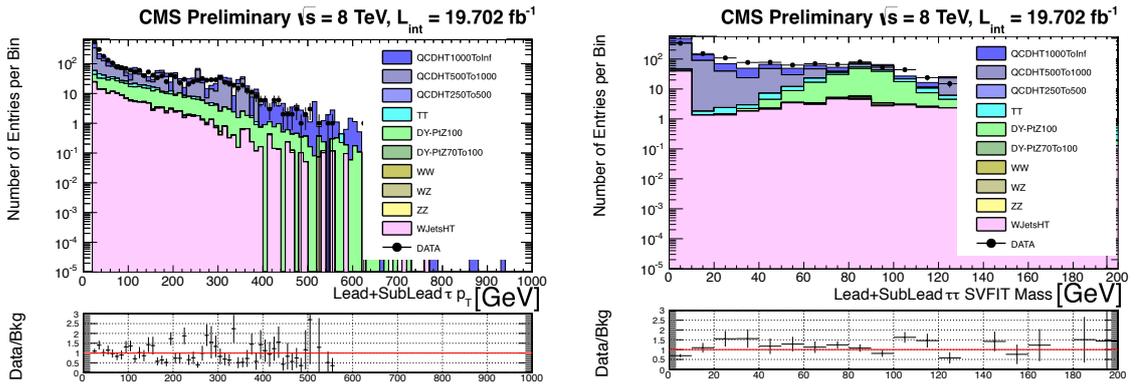
Figure 6.24: Comparison between MC simulation and observed events. Left: leading and subleading τ_h p_T. Right: SVFit $\tau\tau$ mass.

In this analysis, the SB is defined as the full selection but using only the events in the pruned jet mass window ($20 < m_{jet}^P < 70$ GeV). With this selection, the number of back-
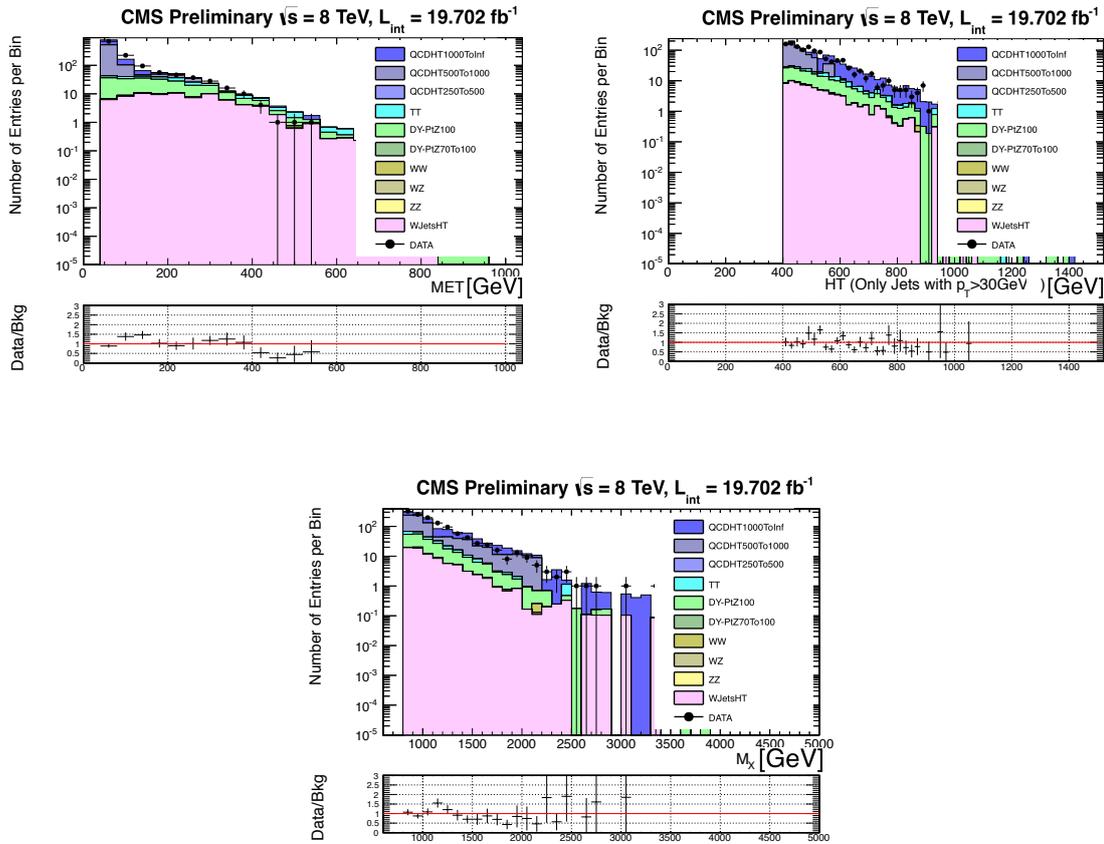
Figure 6.25: Comparison between MC simulation and observed events. Top left: PFMET variable. Top right: H_T variable, scalar sum of the jet p_T for jets with $p_T > 30\,\mathrm{GeV}$. Bottom: mass of the system Z-Jet $+ \tau\tau$.

ground events is suppressed and it is not suitable to perform a shape analysis. In addition, the MC simulation of these events does not seem reliable as we could appreciate in the control plots of Figs. 6.23–6.25. To avoid this problem, our strategy to estimate the background in the SR will be to use a single bin analysis (no shape) and avoiding as much as possible the usage of MC simulations.

For the background estimate we consider the "ABCD method". This method is defined as follows: given two uncorrelated variables ($var1$ and $var2$), if we can define four regions in the plane $var1 \times var2$ so that three regions are signal free ("regions BCD", for example) and one contain mostly the signal events ("region A" in this case), it is possible to estimate the number of events in "region A" using the events from "regions BCD" through: $N_A^{est} = N_D * (N_B/N_C)$. Note that in this method we always consider observed events to estimate the number of background events in the SR. In this analysis we propose to use the variables pruned jet mass (m_{jet}^P) and SVFit $\tau\tau$ mass ($m_{\tau,\tau}$) to define the "ABCD regions". The reason for this is that in theory they do not have any correlation between each other. The SVFit method basically depends on the kinematics of the τ_h and PFMET. Here we defined one selection that avoid any overlap between the Z-jet candidate and the two τ_h candidates [6]. To be more quantitative about this correlation, we calculate it using a ROOT method `GetCorrelationFactor` that returns a *Person Product-Moment Correlation Coefficient*, which is defined by the covariance between the two variables divided by the product of the RMS for each variable. Basically this coefficient measures the tendency for any linear dependence between the two variables. In Fig. 6.26 we present three 2D-plots of "$m_{jet}^P \times m_{\tau,\tau}$" for differents ranges of the two variables in MC simulations of the background events that passed the baseline selection defined in Tab. 6.8. The correlation factors obtained for three cases are considered very small, so that if exist any correlation between these two variables it will be very weak. The factors are 0.041 (top-left distrubution), -0.004 (top-right distribution), and 0.073 (bottom distribution).

In Fig. 6.27 we show a diagram describing the regions ABCD in $m_{jet}^P \times m_{\tau,\tau}$. The higher thresholds ($70 < m_{jet}^P < 110$ GeV and $105 < m_{\tau,\tau} < 180$ GeV) are chosen so that they are in agreement with the signal region and the lower thresholds ($m_{jet}^P > 20$ GeV and $m_{\tau,\tau} > 60$ GeV) are chosen to avoid contaminations from problematic backgrounds, like low-mass resonances. Applying the ABCD method in observed events using the full-selection, the results are shown on Tab. 6.10 (column "Observed events - Original").

---

[6]See distance in x-y plane between jets, τ_h, and PFMET defined in Tab. 6.8
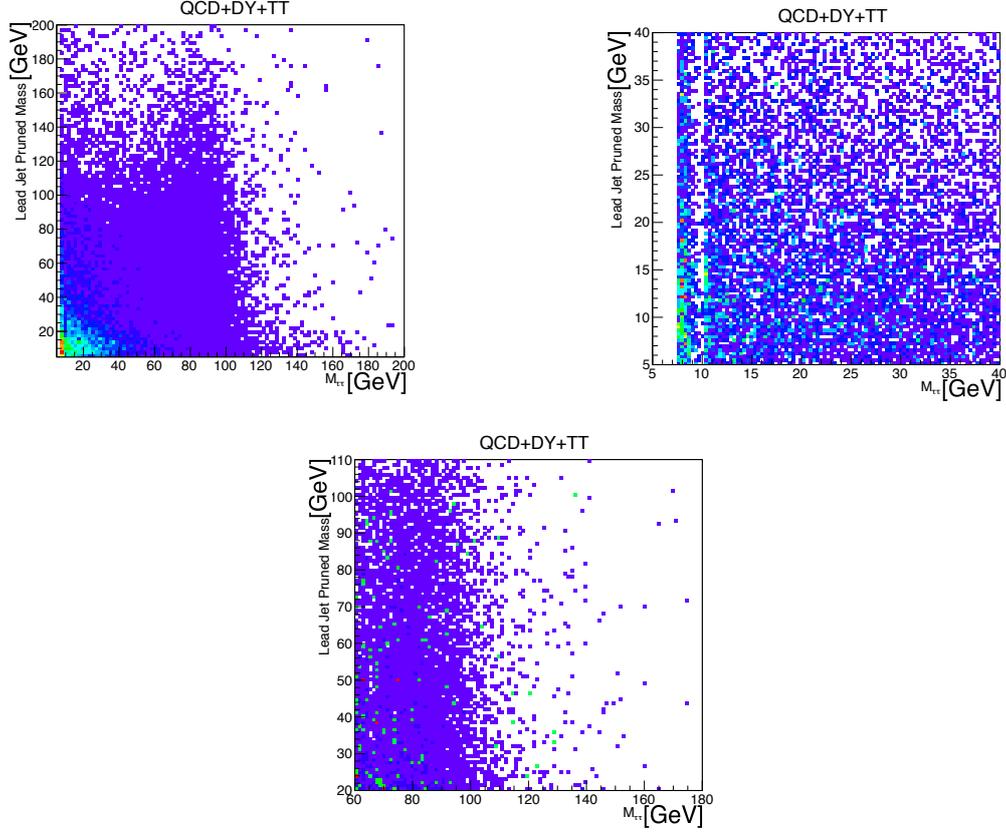
Figure 6.26: Correlation between the variables Jet pruned mass and $m_{\tau,\tau}$ in MC simulation of background events. The plot in the top-right is just a zoom in the low-mass region of the plot in the top-left. In the bottom part we present the correlation for the mass regions used in the ABCD method. We calculated the correlation factor for the three cases: 0.041 (top-left plot), -0.004 (top-right plot), and 0.073 (bottom plot). NB: there is a kind of strip with lack of events for around $m_{\tau,\tau} = 10$ GeV, this is a kinematical effect given that we have in the τ reconstruction level a cut in the p_{T} of the subjet in 10 GeV (see Subsec. 6.2.2).

To test the consistency of the ABCD method we re-estimate the events in "region A", but changing the lower thresholds in the mass selection for "regions BCD" as follows:

$$20 < m_{\mathrm{jet}}^{\mathrm{P}} < 70 \,\mathrm{GeV} \rightarrow 5 < m_{\mathrm{jet}}^{\mathrm{P}} < 70 \,\mathrm{GeV};$$

$$60 < m_{\tau,\tau} < 105 \,\mathrm{GeV} \rightarrow 5 < m_{\tau,\tau} < 105 \,\mathrm{GeV}.$$

In this way, we maintain the "region A". On Tab. 6.10 we present the results using this "test configuration" and compare with the original regions. We observe a good agreement between the two estimates.

Using the QCD enriched sample, where at least one τ_{h} fails isolation criterion, it is possible to test the method for these events, which, besides to be mainly composed by QCD events, have a very similar kinematics compared to the full selection. In addition, the enriched sample has high number of events to test the ABCD method and, given that this is not the "true sample", we can compare the estimate with the direct number of observed events, without violating the "blinding policy". On Tab. 6.11 we show these results and conclude that
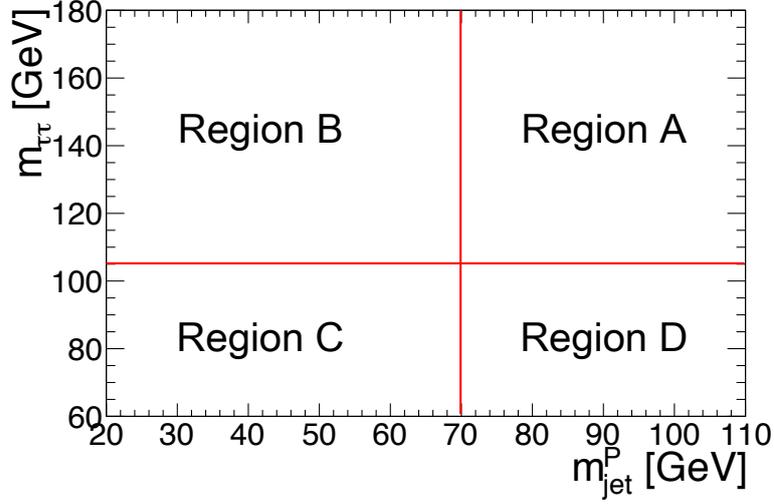
Figure 6.27: Diagram showing the ABCD regions in the $m_{\text{jet}}^{\text{P}}$ and $m_{\tau,\tau}$ variables.

Table 6.10: Number of events in each ABCD region in two different configurations: the "original" used in the ABCD method and one reducing the lower threshold in the mass cuts (test configuration). In addition, we show the number of events estimated in the signal region "NA (estimate)". We present only statistical uncertainties.

Regions	Observed events - original	Observed events - test configuration
NB	$9.0^{+4.1}_{-2.9}$	$14.0^{+3.7}_{-4.8}$
NC	43^{+8}_{-7}	128^{+12}_{-11}
ND	29^{+6}_{-5}	50^{+8}_{-7}
NA (estimate)	$6.1^{+3.2}_{-2.5}$	$5.5^{+2.1}_{-1.7}$

Table 6.11: Number of events in each region for observed events in a QCD enriched sample. We show the number of events estimated through the ABCD method in region A "NA(estimate)", which is compared with the direct number of events in this region "NA". We present only statistical uncertainties.

Regions	Observed events - QCD enriched sample
NB	3762^{+62}_{-61}
NC	11419^{+108}_{-107}
ND	6568^{+82}_{-81}
NA	2164^{+49}_{-49}
NA(estimate)	2101^{+47}_{-46}

the estimate in "region A", row "Region A (estimate)", is in agreement, within the statistical uncertainties, with the direct number of events in the same region, row "Region A".

To further test our estimation, we perform a closure test changing the pruned jet mass region from "$70 < m_{\text{jet}}^{\text{P}} < 110$ GeV" to "$m_{\text{jet}}^{\text{P}} > 150$ GeV", so that we can compare

the number of events estimated with the output in the "new region A" (very small signal contamination, then we can "unblind"). On Tab. 6.12 we present the results and conclude that the estimate and number of events observed in "region A" are in agreement considering the statistical uncertainties.

Table 6.12: Closure test for ABCD method. Number of events observed in each region. "Regions A and D" are re-defined changing the pruned jet mass selection from "$70 < m_{\text{jet}}^{\text{P}} < 110 \,\text{GeV}$" to "$m_{\text{jet}}^{\text{P}} > 150 \,\text{GeV}$". We consider only statistical uncertainties.

Regions	Observed events
NB	$9.0^{+4.1}_{-2.9}$
NC	43^{+8}_{-7}
ND	18^{+5}_{-4}
"NA"	$3.8^{+2.1}_{-1.7}$
"NA"(estimate)	$5.0^{+3.4}_{-2.2}$

The ABCD method relies on the assumption that there is a negligible number of signal events in the side-bands. However, it is possible to have some contamination. Fig. 6.28 shows the two dimensional distribution of $m_{\text{jet}}^{\text{P}} \times m_{\tau,\tau}$ for observed events after the full selection with a MC simulation of a signal superimposed ($M_X = 1.5$ TeV). We observe that in all the regions we have signal events. In order to see if the ABCD method is still reliable after a possible signal contamination in the side-bands, we redo the background estimate and expected limits adding signal events in the "regions BCD". Tab. 6.13 shows the background expectations for three situations: without any signal events in the SBs, adding signal events in SBs, and adding signal events multiplied by a factor of ten. In these three cases we have compatible results, leading us to conclude that the method is insensitive to signal contamination in SBs.

Table 6.13: Background estimate from ABCD method for three situations: without any signal events in the SBs, adding signal ($M_X = 1.5$ TeV) events in SBs, and adding signal events but multiplied by a factor of ten.

Regions	Results	Results (add signal in SBs)	Results (add $10\times$ signal in SBs)
B	$9.0^{+4.1}_{-2.9}$	$9.1^{+4.1}_{-3.0}$	$10.2^{+4.3}_{3.1-}$
C	43^{+8}_{-7}	43^{+8}_{-7}	44^{+8}_{-7}
D	29^{+6}_{-5}	29^{+6}_{-5}	33^{+7}_{-6}
"Estimate in A"	$6.1^{+3.2}_{-2.5}$	$6.2^{+3.3}_{-2.6}$	$7.8^{+3.8}_{-3.1}$
Exp. limit (fb)	$3.0^{+1.7}_{-1.6}$	$3.0^{+1.7}_{-1.6}$	$3.2^{+1.9}_{-1.1}$

An alternative method to be compared with the ABCD is described in Appendix A. This method uses a QCD enriched sample and two SB regions in order to estimate separately QCD multijets, Z/W+jets $+ t\bar{t}$, and SM dibosons. The background estimate from this method using only statistical uncertainties is $7.6^{+3.7}_{-2.8}$, which is in agreement with the ABCD method prediction of $6.1^{+3.2}_{-2.5}$.

In Fig. 6.29 we show the distributions of the mass of the system Z-jet $+ \tau\tau$ (m_{ZH}) for observed events and MC simulations of the background and signal for 1.5 TeV after full-selection. All the background components are from MC simulations except the QCD that
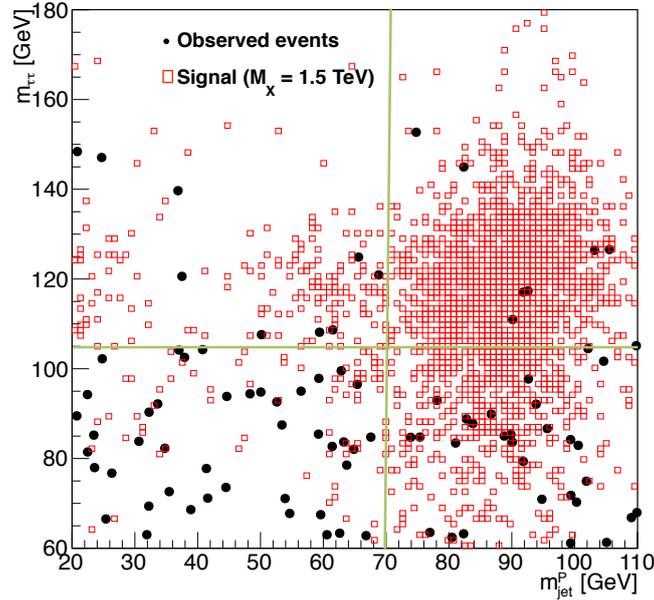
Figure 6.28: $m_{\text{jet}}^{\text{P}} \times m_{\tau,\tau}$ distribution for observed events (black points) and signal simulation of $M_X = 1.5$ TeV (red boxes). The green lines indicate the regions B (top left), A (top right), C (bottom left), and D (bottom right) used in the background estimation method. Each black point corresponds to one observed event for $L_{int} \approx 19.7$ fb^{-1}. The signal events are in arbitrary units, so that each red box corresponds to a signal event in the luminosity of the generated sample (L_{int}^{gen}), $L_{int}^{gen} >> L_{int}$.

considers the shape from a QCD enriched sample constructed using the same signal selection but requiring at least one τ_{h} to fail the isolation criterion.

6.6   Systematic uncertainties

The systematic uncertainties incorporate extra effects that affect the comparison among the observed events, background estimate and signal yield. These effects come from many sources, from detector based interferences to offline methods - like a fit procedure that models a distribution shape. In this section we will present the main effects that were supposed to interfere in the present analysis[7].

6.6.1   Background estimation method

In this analysis, we estimate the background using only observed events. The number of background events in signal region (n) is estimated from events in the side-band region (N) multiplying by an extrapolation factor (α). In this case, a possible distribution for n is the Gamma distribution (see Ref. [134] for a more detailed description). Therefore the

---

[7]I would like to thank Camilla Galloni from University of Zurich for the important contribution on these studies.
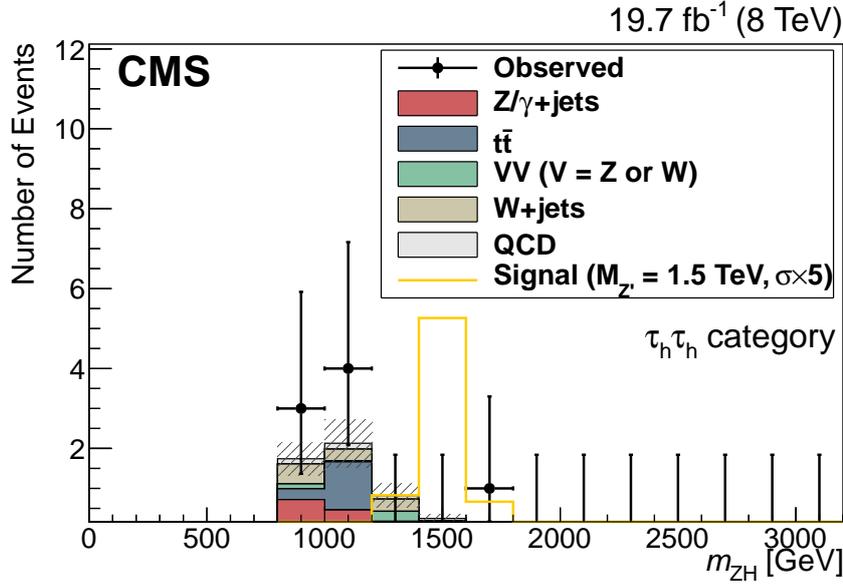
Figure 6.29: Observed distributions of m_{ZH} along with the corresponding MC expectations for signal and background. The signal cross-section is scaled by a factor of 5.

uncertainty on n is described by a Gamma distribution:

$$\Gamma(n) = \frac{1}{\alpha} \frac{(n/\alpha)^N}{N!} e^{-n/\alpha}$$

In this analysis we have a side-band with $N = 29$ events and an extrapolation factor of $\alpha = 0.21 \pm 0.10$.

6.6.2 Luminosity

From studies presented in Ref. [34], it is considered an uncertainty of 2.6% on the measurement of the LHC luminosity delivered to CMS in 2012 proton-proton collisions.

6.6.3 Pileup reweighting

To match the number of pile-up interactions in observed events, the events from MC simulations are reweighted. This is evaluated considering the instantaneous luminosity and the total inelastic proton-proton cross section, ~ 69.4 mb for 2012 LHC Run [135]. A systematic uncertainty is associated to this procedure: the analysis is redone after varying $\pm 5\%$ the total inelastic proton-proton cross section and the variation in signal efficiency is taken as systematic uncertainty. In Tab. 6.14 we list the relative variation for four resonance masses.

Table 6.14: Systematic uncertainties due to the pile-up reweigthing.

Resonance mass (TeV)	1.0	1.5	2.0	2.5
Relative uncertainty	1.0%	0.7%	0.1%	1.0%

6.6.4   V-tagging efficiency

An uncertanty is associated to the V-reconstruction procedure that has a different efficiency between MC simulation and observed events. A scale factor, $\mathrm{SF(V-tag)}$, which is the ratio between the efficiencies in observed events and MC simulation is found following the Ref. [14]:

$$\mathrm{SF(V-tag)} = 0.94 \pm 0.06.$$

The uncertainty of this scale factor is used as systematic error on the signal normalization.

6.6.5   Jet energy scale

For every jet in the event we correct the energy using the jet energy corrections increased and decreased of an amount equal to their uncertainties. We corrected independently Cambridge-Aachen jets with distance parameter 0.8 (CA8) and anti-k_T jets with distance parameter 0.5 (Ak5). To avoid overcorrecting the PFMET, we propagated to this quantity just the correction found using Ak5 jets with $p_\mathrm{T} > 10$ GeV. We take as systematic uncertainty the variation in the signal yield. The systematic uncertainties are summarized in Tab. 6.15.

Table 6.15: Systematic uncertainties associated to the jet energy scale.

	Jet Energy Scale							
	Up	Down	Up	Down	Up	Down	Up	Down
Resonance mass (TeV)	1.0		1.5		2.0		2.5	
Relative uncertainty	1.2%	-2.8%	2.1%	-1.8%	1%	-3.1%	2.5%	-2.2%

6.6.6   Jet transverse momentum resolution

We over-smear reconstructed jets in simulated events so that their p_T resolution would be the same as we measure it in observed events [8]. We have to distinguish between 2 cases:

- if a matched jet is found at generator level (genJet): We scale the reconstructed (corrected) jet p_T based on the p_T difference between matched reconstructed and genJet $p_\mathrm{T} \rightarrow max[0., p_\mathrm{T,genJet} + c * (p_\mathrm{T} - p_\mathrm{T,genJet})]$ where c is the core resolution scaling factor, i.e. the measured data/MC resolution ratio provided in [136].

- if a matched jet is not found at generator level: We randomly smear the reconstructed (corrected) jet p_T using a Gaussian of width $\sqrt{(c^2 - 1)} * \sigma_\mathrm{MC}$. This method only allows one to worsen the resolution ($c > 1$). We need to determine the jet resolution in MC simulation (σ_MC). We determine them separately for CA8 and Ak5 jets, in bins of $|\eta|$ for the CA8 and in bins of $|\eta|$ and p_T for the Ak5. The results are listed in Tab. 6.16.

We computed the impact of the uncertainty of the jet energy resolution using the previous procedure considering the uncertainties on the c factors, i.e. we repeated the procedure to smear the $p_\mathrm{T} > 100$ GeV with the c factors increased and decreased of their uncertainties. To avoid overcorrecting the PFMET, we propagated to the PFMET just the correction found using Ak5 jets with $p_\mathrm{T} > 10$ GeV. The systematic uncertainties in the analysis are taken as the yield variations in Tab. 6.17.

---

[8]It was observed in 2012 run that the jet energy resolution in data was worse than in simulation.

Table 6.16: Jet energy resolution in MC simulation of signal events.

	Jet energy resolution for Ak5 jets				
	$|\eta|< 0.5$	$0.5 < |\eta|< 1.1$	$1.1 < |\eta|< 1.7$	$1.7 < |\eta|< 2.3$	$2.3 < |\eta|< 5$
$10 < p_T < 30$ GeV	0.19	0.21	0.23	0.22	0.23
$30 < p_T < 100$ GeV	0.11	0.12	0.15	0.13	0.16
$p_T > 100$ GeV	0.06	0.06	0.07	0.07	0.08
	Jet energy resolution for CA8 jets				
	$|\eta|< 0.5$	$0.5 < |\eta|< 1.1$	$1.1 < |\eta|< 1.7$	$1.7 < |\eta|< 2.3$	$2.3 < |\eta|< 5$
$p_T > 100$ GeV	0.05	0.06	0.06	0.06	0.03

Table 6.17: Systematics errors due to the jet energy resolution.

	Jet energy resolution			
Resonance mass (TeV)	1.0	1.5	2.0	2.5
Relative uncertainty	-2.9%	-2%	-2.5%	-2%

6.6.7 Hadronic tau: standard reconstruction and identification

We use the uncertainties on the standard identification and reconstruction for the τ_h recommended by the collaboration: $syst(p_T) = 6\% + 0.2 \times p_T/(1000 \text{ GeV})$ [137]. For each resonance mass value, the impact of this systematic is computed assigning a weight to each event depending on the p_T of the τ_h ("$1 \pm syst(p_T)$" for each τ_h; both τ_h are considered in the all-hadronic channel). The percentage difference in the yields variation is taken as systematic uncertainty. Considering the τ_h spectra (Fig. 6.30) of the MC simulation of signal events, these uncertainties vary with the resonance mass from around 18% (1 TeV) to 26% (2.5 TeV). The percentage variations in the yields are reported in Tab. 6.18.

Table 6.18: Expected variation of the signal yields due to the systematic uncertainty of the τ identification creteria.

Resonance mass (TeV)	Yield variation [%]
1.0	17.9
1.5	20.6
2.0	23.5
2.5	26.2

6.6.8 Hadronic tau: energy scale

We consider an uncertainty on the τ_h energy scale of 3% [138]. To estimate the impact of this uncertainty in the analysis, we repeat our selection varying each τ_h p_T of 3% ($p_T \rightarrow p_T + 3\% * p_T$ while doing the event selection with energy scale increased and $p_T \rightarrow p_T - 3\% * p_T$ while doing the event selection with energy scale decreased). We take as systematic uncertainties the variation in the signal yields. We summarize the results in Tab. 6.19.
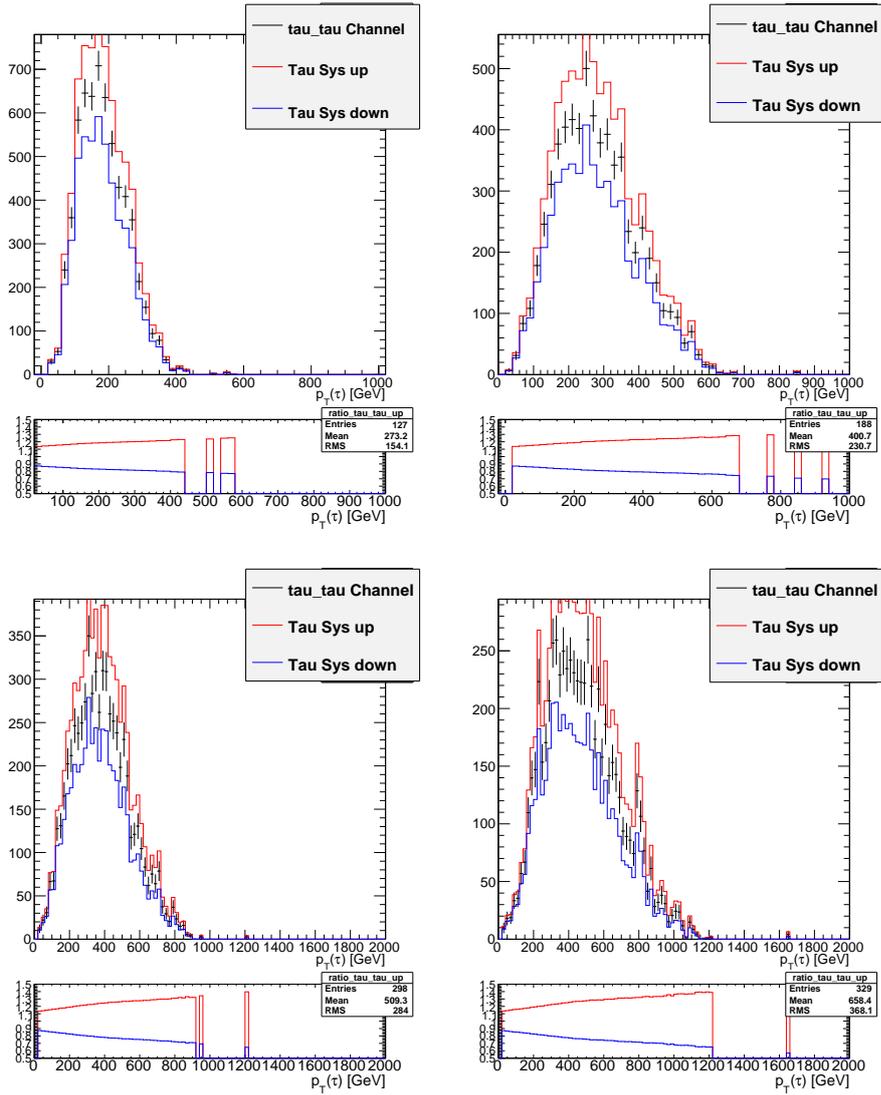
Figure 6.30: Spectra of the τ_h for the resonance masses of 1.0 TeV (top left), 1.5 TeV (top right), 2.0 TeV (bottom left), and 2.5 TeV (bottom right). In red and blue are showed the variation of the spectra due to the uncertainty on the τ_h standard identification.

6.6.9 Hadronic tau: reconstruction and identification with subjets

We assign an extra uncertainty taking into account the fact that the two subjets are collimated. The uncertainty is extracted by comparing the identification efficiencies using isolated τ leptons from $\mathrm{W}' \rightarrow \tau\nu$ MC simulations and the signal samples with boosted τ leptons reconstructed from subjets. The relative difference give us the additional systematic. From Fig. 6.31 we conclude that it is reasonable to consider a flat conservative value of 10% for τ_h candidates with $p_\mathrm{T} > 20$ GeV.

Table 6.19: Systematics uncertainties due to the tau energy scale.

	Tau energy scale							
	Up	Down	Up	Down	Up	Down	Up	Down
Resonance mass (TeV)	1.0		1.5		2.0		2.5	
Relative uncertainty	-4.2%	0.7%	-2.7%	1.8%	-3.2%	1.2%	-2.5%	1.3%
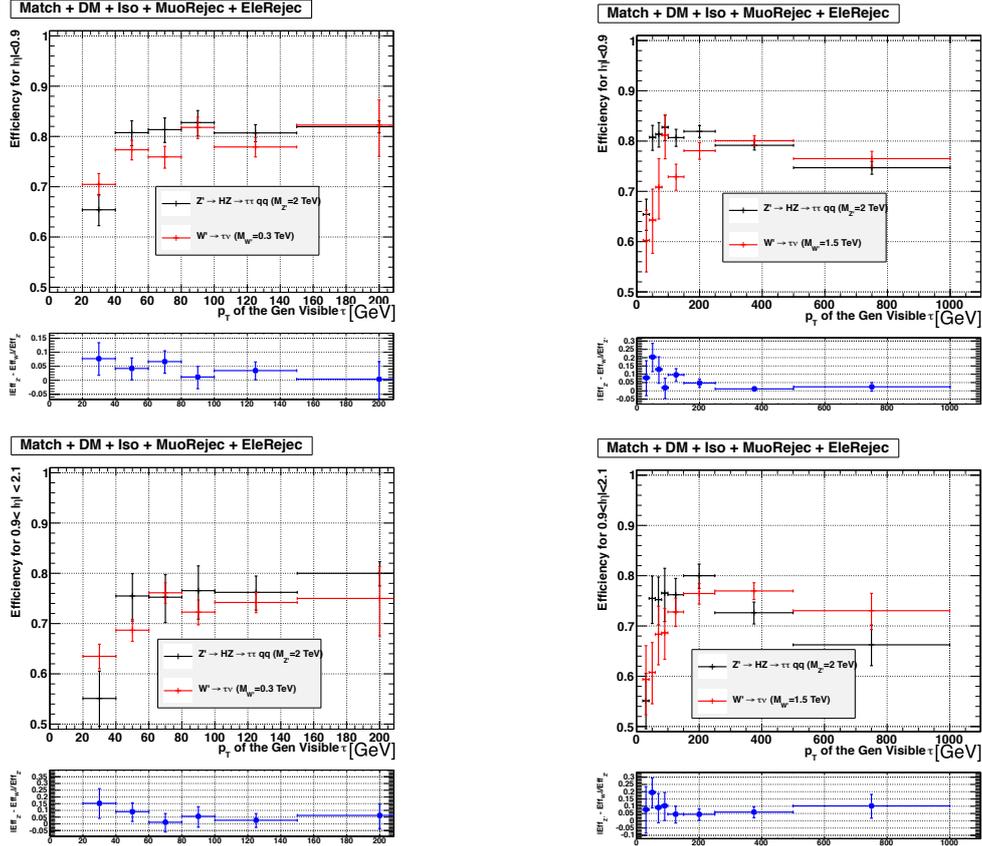
Figure 6.31: Efficiencies for the standard τ_{h} reconstruction (red) and the reconstruction from subjets (black). Top: efficiencies for $|\eta| < 0.9$. Bottom: efficiencies for $0.9 < |\eta| < 2.1$. We consider two W' samples ($M_{W'} = 0.3$ TeV and $M_{W'} = 1.5$ TeV) so that it is possible to estimate more precisely the uncertanties in the whole p_{T} range $20 < p_{\mathrm{T}} < 1000$ GeV.

6.6.10   Missing transverse energy

The uncertainties of the energy scale/resolution of jets are propagated to an uncertainty on the missing transverse energy. In particular, for each jet energy uncertainty, a "new" PFMET is used, defined as:

$$\vec{\cancel{E}}_T(\text{new}) = \vec{\cancel{E}}_T + \sum_{\text{reco objects}} [\,\vec{p}_{\,\mathrm{T}} - \vec{p}_{\,\mathrm{T}}(\text{new})\,]$$

where "new" stays for the new vector after the variations have been applied.

6.6.11 Summary of the systematics

In Tab. 6.20 a summary of all the systematic uncertainties considered in the analysis is shown. The minimum and maximum values among the four signal masses $(1.0, 1.5, 2.0,$ and 2.5 TeV$)$ are presented. These uncertainties are associated to the MC simulation. The uncertainties affect only the signal prediction because the background is estimated using just observed events.

Table 6.20: Summary of the systematics applied in the analysis. Minimum and maximum values for four signal masses $(1.0, 1.5, 2.0,$ and 2.5 TeV$)$ are reported.

Source	Relative uncertainty
Luminosity	2.6%
Pile-up	0.1%-1.0%
V-tag	6.4%
Jet scale	1.0%-3.1%
Jet resolution	2.0%-2.9%
Tau ID	17.9%-26.2%
Tau Scale	0.7%-4.2%
Tau - subjet	10%
PFMET	Included in jet energy uncertainties

6.7 Results

As discussed before, we validated the methods for background estimate in the signal region and after these cross-checks we unblinded the analysis, counting the number of observed events in the signal region. We expect $6.1^{+3.2}_{-2.5}$ background events and it was observed 8 events, which are statistically compatible. The relative uncertainty on the expected background is only statistical and is around 50%, this huge percentage is due to the small number of events in the side-bands.

A similar procedure was done for the other channels, where at least one of the τ leptons are expected to decay leptonically[9]. In Tab. 6.21 we list, for all the channels, the signal efficiencies, number of expected background and observed events after the final selection. We did not observe any statistical significant excess in the observed events compared to the expectations. From these numbers we computed expected and observed upper limits in the Z' production cross-section times branching fraction in ZH. These upper bounds are computed using the $\mathrm{CL_s}$ criterion [139, 140]. The test statistic is a profile likelihood ratio [134] and the systematic uncertainties are treated as nuisance parameters with the frequentist approach. The nuisance parameters are described with log-normal probability distribution functions, except for those related to the extrapolation from sideband events, which are expected to follow a Γ distribution (see Section 6.6). In Fig. 6.32 it is shown the expected and observed upper limits for the three categories (all-hadronic, all-leptonic, and semileptonic) considering

---

[9]These channels were investigated by Aniello Spiezia from University of Perugia. We have worked in a very strong collaboration debugging codes and defining analysis strategies.

all the systematic uncertainties[10]. The limits show the cross-sections that are excluded at 95% confidence level for each resonance mass hypothesis. Cross-sections above the black dotted line represent the expected excluded values and cross-sections above the continuous line are the observed excluded values. The green and yellow bands represent, respectively, $\pm 1\sigma$ and $\pm 2\sigma$ deviations of the expected limits. For lower resonance masses, below ~ 1.5 TeV, the all-hadronic and semileptonic channels showed better limits than in all-leptonic. For masses above ~ 1.5 TeV all the channels are comparable. This is reasonable, since the backgrounds and observed events are always compatible with zero [11] and the signal efficiency is higher (and similar among the channels) in these high-mass regions (see Tab. 6.21).

Table 6.21: Summary of the signal efficiencies, number of expected background events, and number of observed events for the six $\tau\tau$ channels. Only statistical uncertainties are included. For the all-leptonic and semileptonic channels, numbers of expected background events and observed events are evaluated for each mass point in m_{ZH} intervals corresponding to ± 2.5 times the expected resolution. For the all-hadronic channel we consider the number of expected background, signal, and observed events for $m_{\mathrm{ZH}} > 800$ GeV. When the expected background is zero, the 68% confidence level upper limit is listed.

	Mass (TeV)	$\tau_e\tau_e$	$\tau_e\tau_\mu$	$\tau_\mu\tau_\mu$	$\tau_e\tau_h$	$\tau_\mu\tau_h$	$\tau_h\tau_h$
$\mathcal{B}(\tau\tau)$		3.2%	6.2%	3.0%	23.1%	22.6%	41.9%
$\varepsilon_{\mathrm{sig}}(\%)$	0.8	2.8 ± 0.7	3.4 ± 0.5	4.2 ± 0.7	3.3 ± 0.3	4.4 ± 0.3	2.2 ± 0.2
	0.9	11 ± 1	16 ± 1	20 ± 2	14.3 ± 0.5	18.7 ± 0.6	11.5 ± 0.4
	1.0	17 ± 2	24 ± 1	38 ± 2	21.2 ± 0.6	29.3 ± 0.7	18.0 ± 0.5
	1.2	26 ± 2	30 ± 1	39 ± 2	28.3 ± 0.7	35.8 ± 0.7	23.0 ± 0.5
	1.5	30 ± 2	42 ± 2	53 ± 2	29.2 ± 0.8	38.1 ± 0.9	29.1 ± 0.7
	2.0	28 ± 2	39 ± 2	56 ± 3	31.1 ± 0.8	39.2 ± 0.9	31.9 ± 0.7
	2.5	27 ± 2	37 ± 2	42 ± 2	26.8 ± 0.8	37.0 ± 0.8	30.1 ± 0.7
N_{bkg}	0.8	0.3 ± 0.5	1.1 ± 0.8	1.6 ± 1.2	6.1 ± 2.0	6.7 ± 2.1	
	0.9	0.5 ± 0.4	1.7 ± 1.2	3.8 ± 2.1	9.8 ± 3.2	9.2 ± 2.9	
	1.0	1.4 ± 1.4	1.7 ± 1.0	2.0 ± 0.9	9.5 ± 3.5	7.6 ± 2.2	
	1.2	1.2 ± 1.2	1.2 ± 0.8	1.4 ± 0.6	5.0 ± 2.0	6.6 ± 2.3	$6.1^{+3.2}_{-2.5}$
	1.5	0.4 ± 0.4	0.07 ± 0.04	0.9 ± 0.4	4.3 ± 1.8	2.6 ± 0.9	
	2.0	<0.5	<0.4	0.7 ± 0.4	0.1 ± 0.1	<0.4	
	2.5	<2.1	<0.3	0.3 ± 0.1	0.18 ± 0.05	<0.5	
N_{obs}	0.8	1	1	2	3	10	
	0.9	2	2	3	4	13	
	1.0	2	2	5	2	13	
	1.2	0	1	3	5	12	8
	1.5	0	0	1	2	5	
	2.0	0	1	0	0	0	
	2.5	0	0	0	0	0	

6.7.1 Combination of all channels

All the channels were analysed so that the events in the signal regions do not overlap and we do consistently a final combination to extract the final upper limits. In Fig. 6.33 it is shown

---

[10]For the channels involving leptons, in addition to the hadronic taus and jets systematic uncertainties, we have uncertainties associated to electrons and muons.

[11]In all-hadronic channel this is not true since we consider a single bin to compute the limits, i.e. the number of observed events and background estimate are always the same.
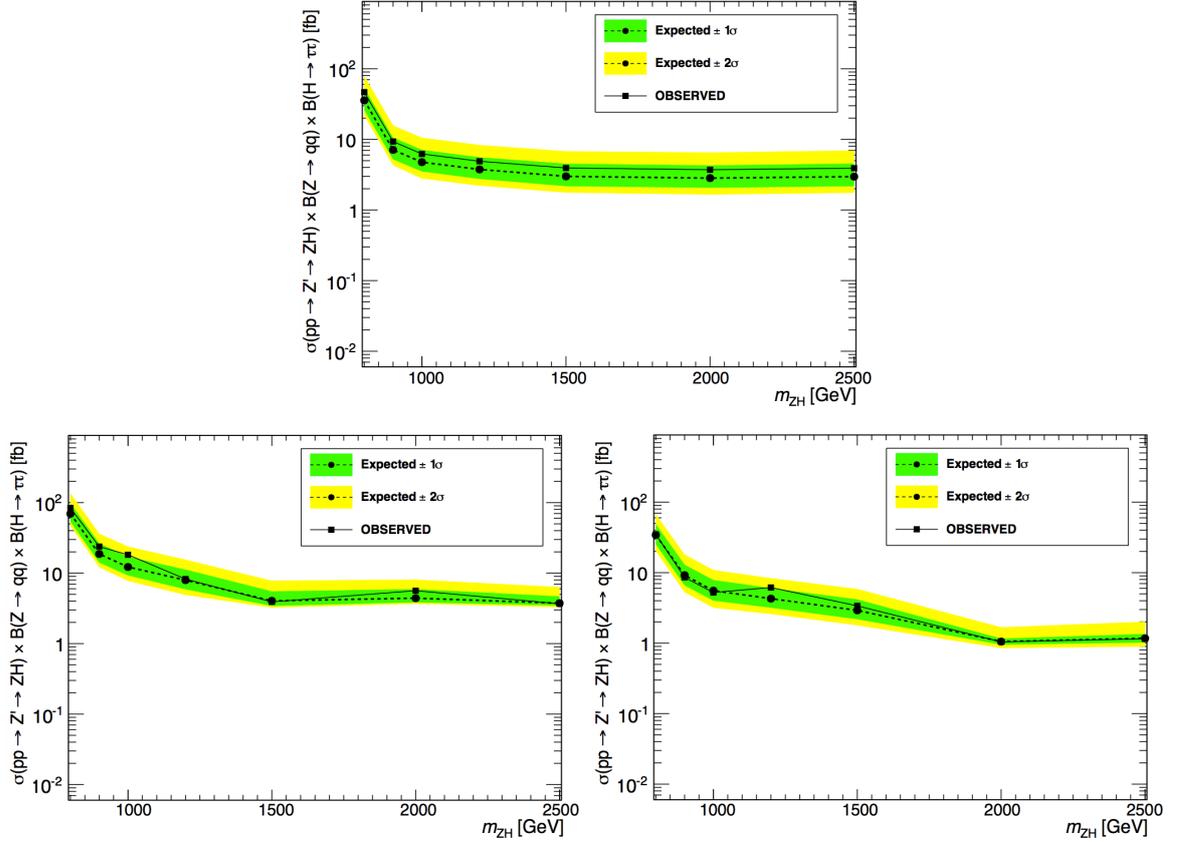
Figure 6.32: Expected and observed upper limit on the $\sigma(Z') \cdot \mathcal{B}(Z' \to ZH)$ as a function of the resonance mass for all search channels. Top: all-hadronic channel. Bottom left: all-leptonic channel. Bottom right: semileptonic channel.

the upper limits for the three categories combined. The following systematic uncertainties are considered fully correlated between the six channels: uncertainty on the luminosity, pileup reweighting, and the error on the V-reconstruction efficiency. The uncertainties on the lepton (e, μ, and τ) identification and energy scale/resolution are correlated between the channels that have at least one of the considered lepton. The other sources of systematic uncertainties are considered uncorrelated. Comparing the combined result with each channel from Fig. 6.32 we observe a considerable improvement in the limits. We add a curve describing the signal cross-section (in leading-order) for the HVT model showed in Subsection 6.1.1. We can not exclude any mass point considering the especific set of parameters: $g_V = 3$, and $c_F = -c_H = 1$.

In Fig. 6.34 it is shown a scan in the parameter space of the HVT model, with observed limits in the plane ($g_V c_H$, $g^2 c_F / g_V$). From this plot we observe that we exclude at 95% confidence level some values of the parameters of HVT model for masses 1.0 TeV and 1.5 TeV. As mentioned before, this analysis was developed with the narrow width assumption [12]. We show the regions in the parameter space where the narrow-width assumption is not valid, represented by the solid shaded region in the corners of the plot.

---

[12]Resonance width is expected to be much smaller than the experimental resolution, which for masses between $1.0 - 2.5$ TeV ranges from 6% to 10% of the resonance mass.
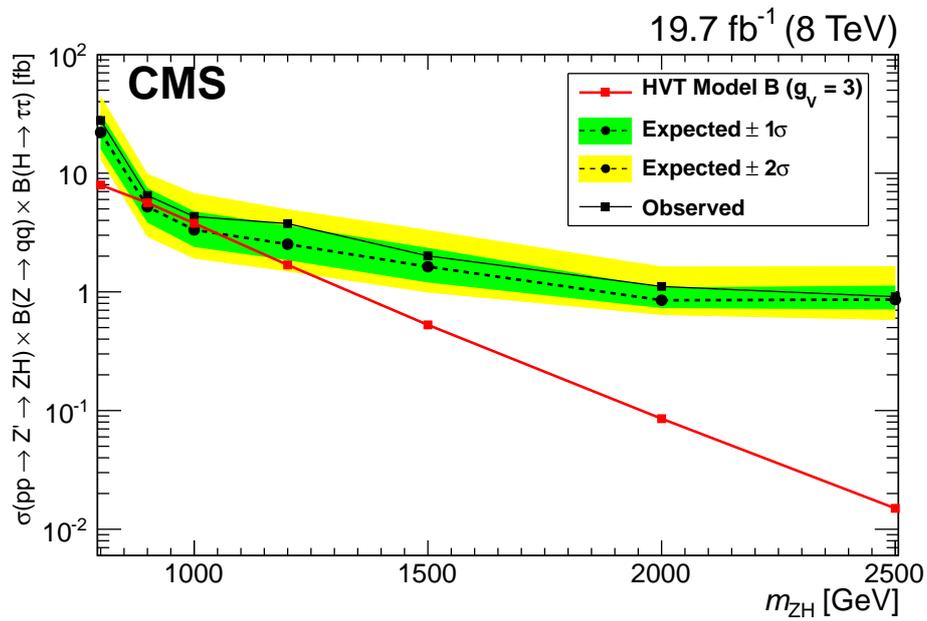
Figure 6.33: Expected and observed upper limits on the quantity $\sigma(Z') \cdot \mathcal{B}(Z' \rightarrow ZH)$ for the six analysis channels combined. Green and yellow bands correspond to ± 1 or $\pm 2\sigma$ variations on the expected upper limit, respectively.
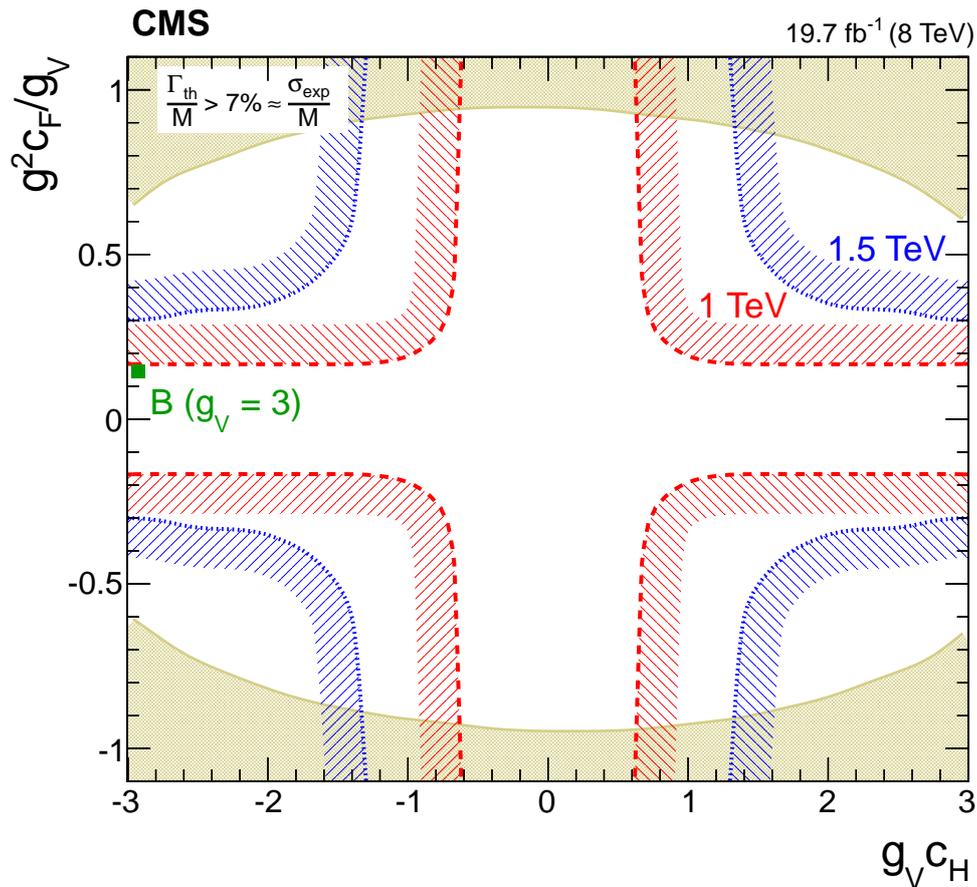
Figure 6.34: Exclusion regions in the plane of the HVT-model coupling constants ($g_V c_H$, $g^2 c_F / g_V$) for two resonance masses, 1.0 and $1.5 \, \text{TeV}$. The point B of the benchmark model used in the analysis, corresponding to $g_V = 3$ and $c_F = -c_H = 1$, is also shown. The boundaries of the regions of the plane excluded by this search are indicated by the dashed and dotted lines, and associated hatching. The areas indicated by the solid line and solid shading correspond to regions where the theoretical width is larger than the experimental resolution of the present search and thus the narrow-resonance assumption is not satisfied.

Chapter 7

Summary

The searches for exotic heavy resonances in the diboson channels are a very important topic in the LHC physics program because it covers a big set of beyond standard model theories where the predicted heavy resonances have suppressed branching fractions into fermions. This feature implies that most of the standard searches for heavy resonances via their direct decays into leptons have low sensitivity for these models, and consequently weaker limits on their parameter space. In addition, for high-mass resonances, these models present weaker limits from electroweak precision tests, which allows a considerable range of values in the parameter space.

ATLAS and CMS have dedicated groups working on the diboson signatures. The groups are in charge of the development of optimized reconstruction and identification criteria for high-momentum bosons, both in its hadronic and leptonic decays. This is a very important task, because the accelerators are constantly increasing the energy of the colliding particles, resulting in events with more high-momentum particles, including the SM bosons Z, W, and Higgs. In addition, the groups study methods for SM background predictions, selection optimization in order to increase the searches sensitivity, methods for setting limits on the models parameter space, and other tasks related to the analyses.

In this thesis, we presented one of the channels investigated by the group in CMS collaboration. We described a search for a heavy (mass above ~ 1 TeV) and narrow spin-1 resonance decaying into a Z and a Higgs bosons in the final state where the Z decays into quarks and the Higgs decays in a pair of τ leptons. This analysis was performed with the CMS detector at the LHC investigating 19.7/fb of integrated luminosity for proton-proton collisions at $\sqrt{s} = 8$ TeV.

The signal events in the detector are expected to be composed by a wide high-energy jet from the $Z \to q\bar{q}$ with a high Lorentz boost and two very collimated τ leptons from the boosted Higgs boson decay. Jet substructure techniques were used to identify the boosted Z boson decaying hadronically and a modified approach was considered to reconstruct the pair of τ leptons. After reconstructing and identifying the Z and Higgs bosons, we combined their kinematic quantities in order to reconstruct the heavy resonance candidate.

MC simulations and control samples of observed events were used to describe SM backgrounds. We compared many kinematic variables among observed events and simulations in order to check how good is our understanding about the physics of the proton-proton collisions in LHC. In general, it was showed that the MC simulations have a good agreement with the observed events.

Since the MC simulations have a low number of generated events in the signal region, dif-

ficulting a proper comparison with data samples, we have defined a method for background estimate using only observed events. This method considered only events in kinematic regions where the number of signal events are expected to be negligible compared with the number of SM backgrounds events (side-band regions). Due to the background composition after the full selection, it was possible to define two uncorrelated variables to extrapolate (with negligible bias) the number of background events from these side-band regions to the signal region.

We observed no deviations between our estimate of the SM background and the observed events. Then we set 95% CL_s upper limits in the resonance production cross-section times the branching fraction into $\mathrm{Z(q\bar{q})H}(\tau^-\tau^+)$ as a function of its mass. We excluded cross-sections in a range between 0.9 and 27.8 fb, depending on the resonance mass.

This channel with boosted τ leptons and jets showed to be an important component of the diboson searches in CMS, allowing the exclusion of regions in the parameter space of composite Higgs models. Our group in CMS started the studies for boosted τ leptons in the RunII period of LHC with proton-proton collisions at $\sqrt{s} = 13$ TeV. We are currently tuning the new criteria for boson reconstruction and identification and defining new analyses strategies. The new techniques for boosted Higgs and Z/W bosons reconstruction are important not only in searches for physics beyond the standard model, but for SM physics as well, and it will become even more important in future accelerators where the colliding particles energy will increase.

Appendix A

Background Estimate - alternative method

In order to validate the results from ABCD method defined in Sec. 6.5, we consider another method to estimate (mostly using observed events) the QCD, Z/W+jets + $t\bar{t}$, and SM diboson backgrounds. The method is defined in four steps:

- **Step1**: Estimate QCD events in the side-band (SB) using a QCD enriched sample.

- **Step2**: Using the QCD in SB from **Step1** and SM dibosons from MC, we estimate W/Z+jets and $t\bar{t}$ in signal region.

- **Step3**: Using the QCD in SB from **Step1** and a QCD enriched sample, we estimate the QCD in signal region.

- **Step4**: Using the MC simulation of SM diboson in signal region and summing with the other components from **Steps 2 and 3**, we get the total background prediction.

A.1 Step 1

In this step we estimate the QCD-multijets in SB region. This is an important step because in this region the MC simulation of QCD is not reliable due to the small number of events generated. As pointed before, the SB region is defined in a specific $m_{\text{jet}}^{\text{P}}$ window ($20 < m_{\text{jet}}^{\text{P}} < 70$ GeV) with all the other cuts fixed to the value used in the full selection from Tab. 6.8. In the SB region it is difficult to say which background dominates, but at a similar selection with $40 < \text{PFMET} < 80$ GeV, QCD has a higher rate compared to the other components. This feature is important because in this method we will suppose that in the region with PFMET below 80 GeV the QCD is represented by the number of observed events minus the other background components from MC simulation [1]. The QCD enriched sample is defined with the full selection but requiring at least one τ_{h} to fail the isolation criterion. This sample has a huge number of events and is dominated by QCD in integrated number of events for PFMET below and above 80 GeV as shown in Fig. A.1. We use the number of events from data-sample minus the other backgrounds components ($t\bar{t}$, W/Z+jets, SM dibosons) from MC simulation from both QCD enriched sample (control sample) and

---

[1] It was checked that the other background components represent around 6% of the events.

original selection with the pair of isolated taus (true sample) to obtain the number of QCD events in the SB. The formula below summarizes the method:

$$N_{\text{SB (80<PFMET)}}^{\text{QCD}} = \overbrace{\left[\frac{N_{\text{SB (80<PFMET)}}^{\text{DATA}} - N_{\text{SB (80<PFMET)}}^{\text{OBs from MC}}}{N_{\text{SB (40<PFMET<80)}}^{\text{DATA}} - N_{\text{SB (40<PFMET<80)}}^{\text{OBs from MC}}} \right]}^{\text{transport factor}}{}^{\text{control sample}}$$

$$\times \left[N_{\text{SB (40<PFMET<80)}}^{\text{DATA}} - N_{\text{SB (40<PFMET<80)}}^{\text{OBs from MC}} \right]^{\text{true sample}},$$

where "OBs" corresponds to Z/W+jets, $t\bar{t}$ and SM diboson. From this method, the QCD estimate in SB is $2.3^{+1.8}_{-1.2}$.

It is important to check if, at least in the MC simulation, the QCD in the SB for the true and enriched samples have compatible PFMET shapes so that we can use the transport factor in the formula above. However, in the "true SB" the number of events is very small in QCD MC simulation. In order to compare the shape of the distributions, in Fig. A.2 we show PFMET distributions for different events selections. In general, the shapes are in agreement within the statistical uncertainties.
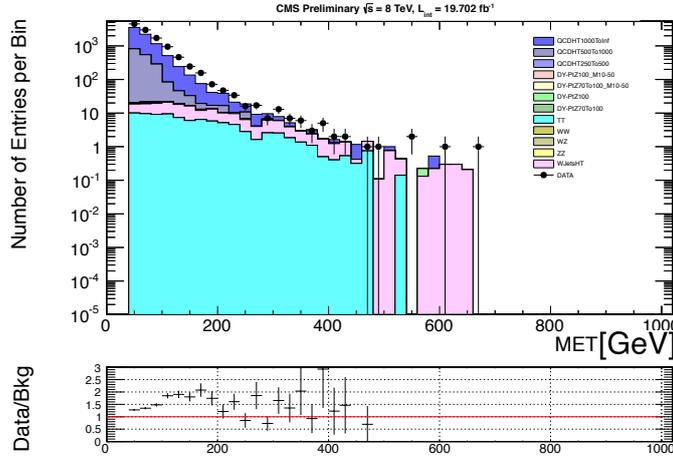
Figure A.1: PFMET distribution in a QCD enriched sample defined with the full selection in the SB ($20 < m_{\text{jet}}^{\text{P}} < 70$ GeV) with at least one τ_{h} failing the isolation criterion and PFMET > 40 GeV.

A.2   Step 2

Since we have the QCD estimate in SB ($N_{\text{SB}}^{\text{QCD}}$), using a similar method as described in Ref. [14] we obtain the (W/Z+jets + $t\bar{t}$) backgrounds in the signal region with the following expression:

$$N_{\text{SIG}}^{\text{Z/W+jets and } t\bar{t}} = (N_{\text{SB}}^{\text{DATA}} - N_{\text{SB}}^{\text{QCD}} - N_{\text{SB}}^{\text{SM diboson}}) \times \alpha^{\text{MC}},$$
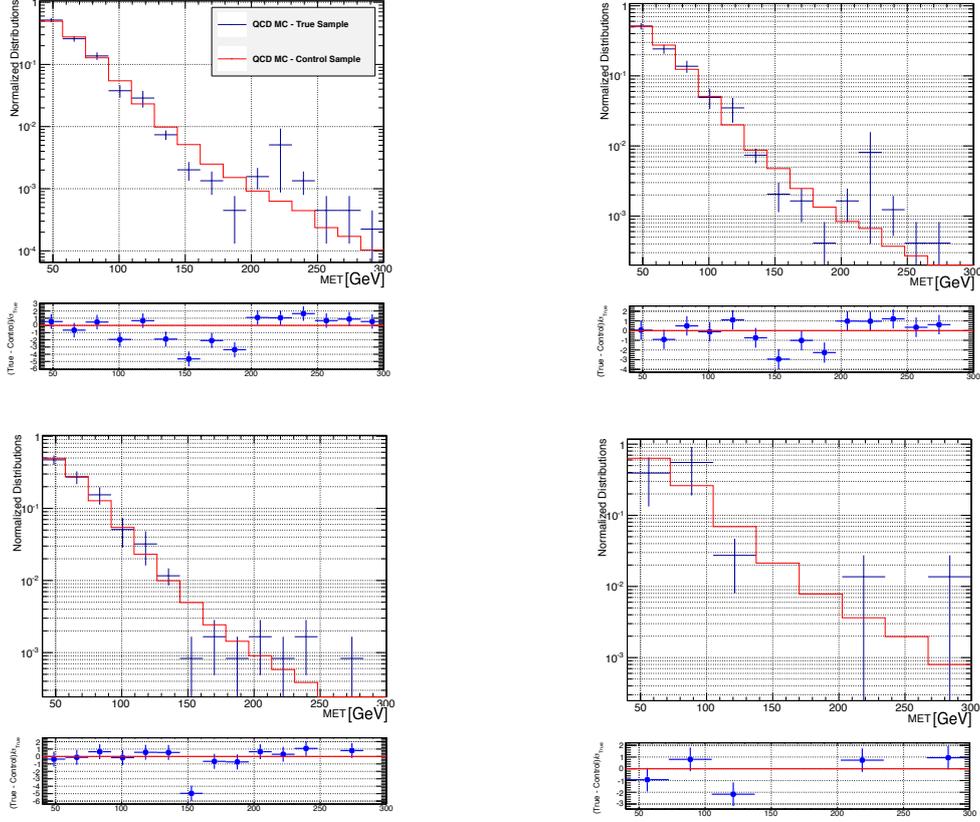
Figure A.2: PFMET shape comparison between the QCD enriched sample and the "true sample" using MC simulation of QCD with different selections. Top left: baseline selection. Top right: baseline selection plus $\tau_{21} < 0.75$. Bottom left: baseline selection plus $p_{\mathrm{T}}^{\mathrm{leading}\,\tau} > 50$ GeV. Bottom right: baseline selection plus $105 < m_{\tau,\tau} < 180$ GeV.

where:

- $N_{\mathrm{SB}}^{\mathrm{DATA}}$ is the observed events in side-band, while $N_{\mathrm{SB}}^{\mathrm{SM\ diboson}}$ is the SM diboson estimation taken from MC simulation.

- α^{MC} is the number of events in MC simulation for (W/Z+jets + $t\bar{t}$) in the signal region divided by the number of events for the same backgrounds in the side-band.

In Tab. A.1 we present the components of the formula above and the estimate of the (W/Z+jets + $t\bar{t}$) background in signal region, $5.9^{+3.5}_{-2.7}$. We show only statistical uncertainties.

A.3   Step 3

Using a similar idea as in Step1, we estimate the QCD in the signal region with the help of the QCD enriched sample. The procedure is defined by the formula below:

$$N_{\mathrm{SIG}}^{\mathrm{QCD}} = \overbrace{\left[\frac{N_{\mathrm{SIG}}^{\mathrm{DATA}} - N_{\mathrm{SIG}}^{\mathrm{OBs\,from\,MC}}}{N_{\mathrm{SB}}^{\mathrm{DATA}} - N_{\mathrm{SB}}^{\mathrm{OBs\,from\,MC}}}\right]}^{\text{transport factor}}{}^{\text{control sample}} \times \left[N_{\mathrm{SB}}^{\mathrm{QCD}}\right]^{\text{from Step1}},$$

Table A.1: Values of the factors used to estimate Z/W+jets and $t\bar{t}$ backgrounds in signal region. We show only statistical uncertainties.

Components	Results
N_{SB}^{DATA}	$9.0^{+4.1}_{-2.9}$
N_{SB}^{QCD}	$2.3^{+1.8}_{-1.2}$
$N_{SB}^{\text{SM diboson}}$	$0.016^{+0.244}_{-0.013}$
α^{MC}	$0.89^{+0.25}_{-0.24}$
$N_{SIG}^{\text{Z/W+jets and }t\bar{t}}$	$5.9^{+3.5}_{-2.7}$

where "OBs" corresponds to Z/W+jets, $t\bar{t}$ and SM diboson. Using the expression above we estimate QCD in signal region to be $1.3^{+1.0}_{-0.6}$. Here, as in Step1, it is important to compare the shapes of the pruned jet mass distributions between the true and enriched samples. In Fig. A.3 we compare the shapes for different selections. In general they have a resonable agreement considering the statistical uncertainties indicating that we can use the ratio from the QCD enriched sample to estimate the number of events for QCD in the signal region.

A.4 Step 4

We have QCD, W/Z+jets and $t\bar{t}$ in the signal region from previous steps. The SM diboson corresponds to a very small fraction of the background so it is reasonable to take it from MC simulation, $0.39^{+0.19}_{-0.22}$. In Tab. A.2 it is shown the total background estimate in the signal region, $7.6^{+3.7}_{-2.8}$. This result is in agreement with the estimate from ABCD method of $6.1^{+3.2}_{-2.5}$ (see Sec. 6.5).

Table A.2: Summary of the results. Only statistical uncertainties are included.

Components	Results
$N_{SIG}^{\text{Z/W+jets and }t\bar{t}}$	$5.9^{+3.5}_{-2.7}$
N_{SIG}^{QCD}	$1.3^{+1.0}_{-0.6}$
$N_{SIG}^{\text{SM diboson}}$	$0.39^{+0.19}_{-0.22}$
N_{SIG}^{Tot}	$7.6^{+3.7}_{-2.8}$

As in the case of the ABCD method, we test this procedure modifying the selection. We have applied this method for a fixed SVFit $\tau\tau$ mass $105 < m_{\tau,\tau} < 180$ GeV that corresponds to our signal region. We recompute the background defining a new mass window that is totally orthogonal to the signal region, $60 < m_{\tau,\tau} < 105$ GeV. We compare this estimate with the number of observed events. In Tab. A.3 we present the results for each component and the observed number of events. The values are in agreement within the statistical uncertainties, validating the method.
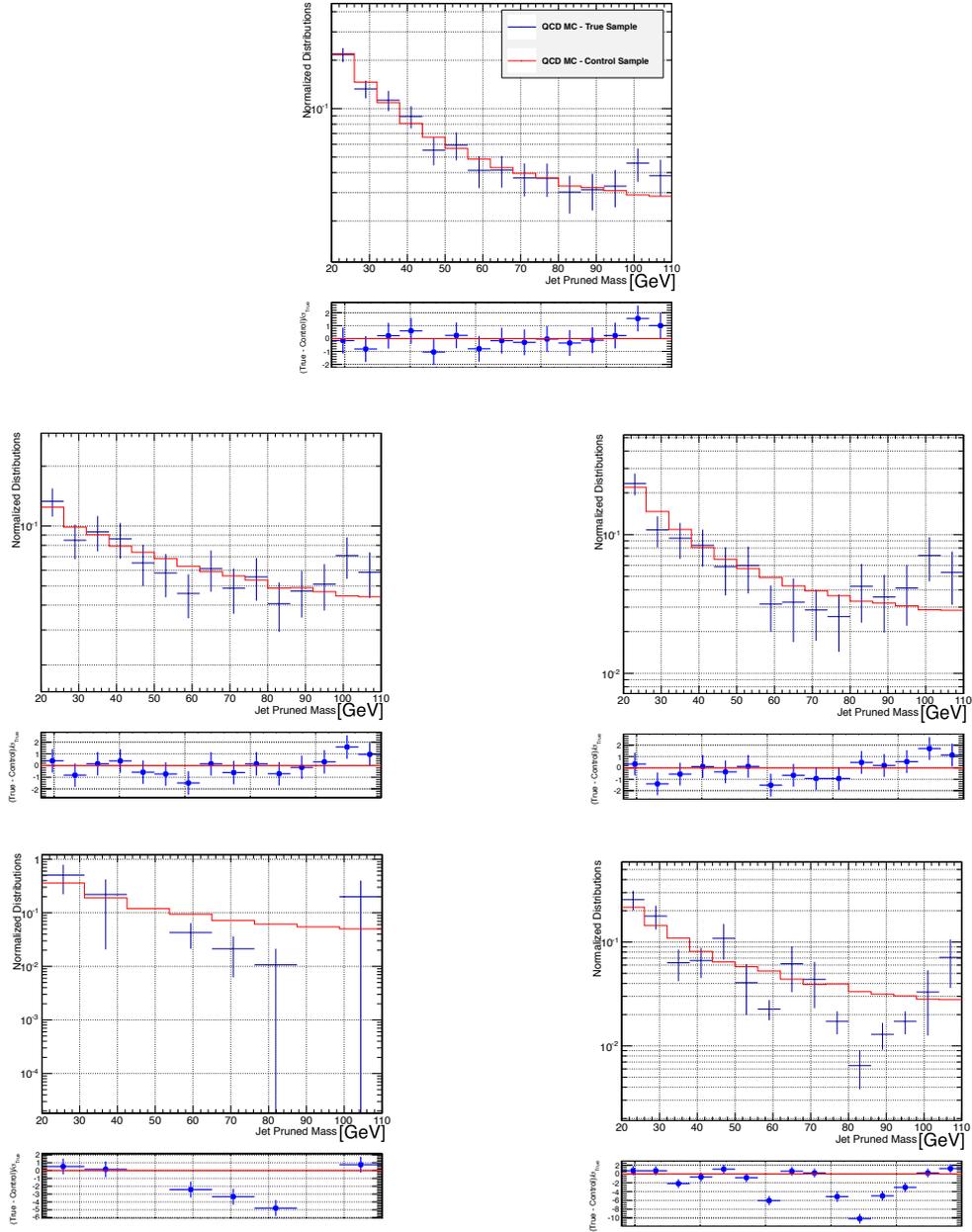
Figure A.3: Pruned jet mass shape comparison between the QCD enriched sample and the true sample using QCD MC simulation with different selections. Top: baseline selection. Middle right: baseline selection plus $\tau_{21} < 0.75$. Middle left: baseline selection plus $p_{\mathrm{T}}^{\mathrm{leading}\,\tau} > 50$ GeV. Bottom right: baseline selection plus $105 < m_{\tau,\tau} < 180$ GeV. Bottom left: baseline selection plus PFMET > 80 GeV.

Table A.3: Method applied in a kinematical region orthogonal to the signal region. We compare the total estimate "N^{Tot}" with the observed number events "N^{Data}".

Components	Results
$N_{\text{SIG}}^{\text{Z/W+jets and } tt}$	$17.4_{-3.2}^{+3.7}$
N^{QCD}	$5.1_{-1.1}^{+1.4}$
$N^{\text{SM diboson}}$	$1.19_{-0.26}^{+0.29}$
N^{Tot}	$23.6_{-3.4}^{+3.9}$
N^{Data}	29_{-5}^{+6}

Bibliography

[1] Aad, G. and others (ATLAS Collaboration), Phys. Lett. B **716** (2012) 1.

[2] Chatrchyan, S. and others (CMS Collaboration), Phys. Lett. B **716** (2012) 30.

[3] Arkani-Hamed, N., Dimopoulos, S., and Dvali, G., Phys. Lett. B **429** (1998) 263.

[4] Chamseddine, A. H., Arnowitt, R., and Nath, P., Phys. Rev. Lett. **49** (1982) 970.

[5] Barbieri, R., Ferrara, S., and Savoy, C. A., Phys. Lett. B **119** (1982) 343.

[6] Agashe, K., Contino, R., and Pomarol, A., Nucl. Phys. B **719** (2005) 165.

[7] Contino, R., Da Rold, L., and Pomarol, A., Phys. Rev. D **75** (2007) 055014.

[8] Randall, L. and Sundrum, R., Phys. Rev. Lett. **83** (1999) 3370.

[9] Leike, A., Phys. Rept. **317** (1999) 143.

[10] Rizzo, T. G., Z' phenomenology and the LHC, hep-ph/0610104, 2006.

[11] Contino, R., Marzocca, D., Pappadopulo, D., and Rattazzi, R., JHEP **10** (2011) 081.

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

[13] Khachatryan, V. and others (CMS Collaboration), JHEP **08** (2014) 173.

[14] Khachatryan, V. and others (CMS Collaboration), JHEP **1408** (2014) 174.

[15] Aad, G. and others (ATLAS Collaboration), Eur. Phys. J. C **75** (2015) 69.

[16] Aad, G. and others (ATLAS Collaboration), Eur. Phys. J. C **75** (2015) 209.

[17] Aad, G. and others (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, arXiv:1506.00962, 2015.

[18] Khachatryan, V. and others (CMS Collaboration), Search for massive WH resonances decaying to $\ell\nu b\bar{b}$ final state in the boosted regime at $\sqrt{s} = 8$ TeV, CMS-PAS-EXO-14-010, 2015.

[19] Aad, G. and others (ATLAS Collaboration), Eur. Phys. J. C **75** (2015) 263.

[20] Khachatryan, V. and others (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, arXiv:1506.01443, 2015.

[21] Aad, G. and others (ATLAS Collaboration), Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, arXiv:1506.00285, 2015.

[22] Pappadopulo, D., Thamm, A., Torre, R., and Wulzer, A., JHEP **1409** (2014) 060.

[23] Schmaltz, M. and Spethmann, C., JHEP **07** (2011) 046.

[24] Bellazzini, B., Csaki, C., Hubisz, J., Serra, J., and Terning, J., JHEP **11** (2012) 003.

[25] Livingston, M. S. and Blewett, J. P., *Particle Accelerators*, McGraw-Hill, 1962.

[26] Chao, A. W., Mess, K. H., Tigner, M., and Zimmermann, F., *Handbook of Accelerator Physics and Engineering*, World Scientific, 2013.

[27] The Large Hadron Collider,
`http://lhc-new-homepage.web.cern.ch/lhc-new-homepage/`.

[28] LHC Machine Outreach,
`http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/`.

[29] Evans, L., The Large Hadron Collider,
`http://iopscience.iop.org/1367-2630/9/9/335/pdf/njp7_9_335.pdf`.

[30] Public CMS luminosity information,
`https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults?rev=101`.

[31] Chatrchyan, S. and others (CMS Collaboration), JINST **3** (2008) S08004.

[32] Chatrchyan, S. and others (CMS Collaboration), JINST **5** (2010) T03021.

[33] Alvarez-Gaume, L. et al., *From the PS to the LHC; 50 years of Nobel memories in high-energy physics*, Springer, 2012.

[34] Khachatryan, V. and others (CMS Collaboration), CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update, CMS-PAS-LUM-13-001, 2013.

[35] CMS Physics - Technical Design Report Vol I: Detector Performance and Software,
`http://cmsdoc.cern.ch/cms/cpt/tdr`.

[36] CMS HCAL Technical Design Report,
`http://cmsdoc.cern.ch/cms/TDR/HCAL/hcal.html`.

[37] CMS The TriDAS Project - Technical Design Report Vol I: The Trigger Systems,
`http://cmsdoc.cern.ch/cms/TDR/TRIGGER-public/CMSTrigTDR.pdf`.

[38] Sherpa homepage,
`https://sherpa.hepforge.org/trac/wiki`.

[39] Ellis, S., Huston, J., Hatakeyama, K., Loch, P., and Tonnesmann, M., Prog. Part. Nucl. Phys. **60** (2008) 484.

[40] The CMS Offline WorkBook,
`https://twiki.cern.ch/twiki/bin/view/CMS/WorkBook`.

[41] Glashow, S. L., Nucl. Phys. **22** (1961) 579.

[42] Weinberg, S., Phys. Rev. Lett. **19** (1967) 1264.

[43] Salam, A., Eighth Nobel Symposium, edited by Almquvist and Wiksell (1968) 367.

[44] Altarelli, G., The Standard model of particle physics, hep-ph/0510281, 2005.

[45] Langacker, P., *The Standard Model and Beyond*, Taylor and Francis, 2009.

[46] Quigg, C., *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, Westview Press, 1983.

[47] Benvenuti, A. and others (Fermilab), Phys. Rev. Lett. **35** (1974) 252.

[48] Arnison, G. and others (UA1 Collaboration), Phys. Lett. B **122** (1983) 103.

[49] Arnison, G. and others (UA1 Collaboration), Phys. Lett. B **126** (1983) 398.

[50] Bagnaia, P. and others (UA2 Collaboration), Phys. Lett. B **129** (1983) 130.

[51] Abe, F. and others (CDF Collaboration), Phys. Rev. Lett. **74** (1995) 2626.

[52] ALEPH, DELPHI, L3, and OPAL, Technical Report No. CERN-EP/99-15 (unpublished) (1999).

[53] SLD Collaboration, S. W., Proceedings of the 24th Annual SLAC Summer Institute on Particle Physics: The Strong Interaction, From Hadrons to Partons, edited by L. Dixon, J. Chan, L. DePorcel (Springfield, NTIS) (1997) 405.

[54] Caso, C. et al., The European Physical Journal C **3** (1998) 1.

[55] Novaes, S., Standard model: An Introduction, hep-ph/0001283, 1999.

[56] Hernandez, P., Neutrino physics, arXiv:1010.4131, 2010.

[57] Yao, W. M. et al., Journal of Physics G **33** (2006).

[58] Aaij, R. and others (LHCb Collaboration), Phys. Rev. Lett. **115** (2015) 072001.

[59] Aad, G. and others (ATLAS & CMS Collaborations), Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and $8\,\text{TeV}$ with the ATLAS and CMS Experiments, arXiv:1503.07589, 2015.

[60] Khachatryan, V. and others (CMS Collaboration), Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV, arXiv:1412.8662, 2014.

[61] Aad, G. and others (ATLAS Collaboration), JHEP **1504** (2015) 117.

[62] Aad, G. and others (ATLAS Collaboration), Phys. Lett. B **726** (2013) 88.

[63] Chatrchyan, S. and others (CMS Collaboration), Phys. Rev. D **89** (2014) 092007.

[64] Heinemeyer, S. et al., Handbook of LHC Higgs Cross Sections: 3. Higgs Properties, arXiv:1307.1347, 2013.

[65] Pomarol, A., Beyond the Standard Model, arXiv:1202.1391, 2012.

[66] Drees, M., An Introduction to supersymmetry, hep-ph/9611409, 1996.

[67] Barbieri, R., Ten Lectures on the ElectroWeak Interactions, arXiv:0706.0684, 2007.

[68] Langacker, P., *The Standard Model and Beyond*, CRC Press, 2009.

[69] Martin, S. P., Adv. Ser. Direct. High Energy Phys. **21** (2010) 1.

[70] Peskin, M. E., Beyond the standard model, hep-ph/9705479, 1997.

[71] Contino, R., The Higgs as a Composite Nambu-Goldstone Boson, arXiv:1005.4269, 2011.

[72] Csaki, C., TASI lectures on extra dimensions and branes, hep-ph/0404096, 2004.

[73] Rizzo, T. G., Pedagogical introduction to extra dimensions, hep-ph/0409309, 2004.

[74] Contino, R., Marzocca, D., Pappadopulo, D., and Rattazzi, R., JHEP **1110** (2011) 081.

[75] Agashe, K., Davoudiasl, H., Perez, G., and Soni, A., Phys. Rev. D **76** (2007) 036006.

[76] Fitzpatrick, A. L., Kaplan, J., Randall, L., and Wang, L.-T., JHEP **0709** (2007) 013.

[77] Antipin, O., Atwood, D., and Soni, A., Phys. Lett. B **666** (2008) 155.

[78] Weinberg, S., *The Quantum Theory of Fields II*, Cambridge University Press, 1995.

[79] Chanowitz, M. S. and Gaillard, M. K., Nucl. Phys. B **261** (1985) 379.

[80] Khachatryan, V. and others (CMS Collaboration), Search for leptonic decays of W' bosons in pp collisions at $\sqrt{s} = 8$ TeV, CMS-PAS-EXO-12-060, 2013.

[81] Khachatryan, V. and others (CMS Collaboration), Phys. Lett. B **740** (2014) 83.

[82] Khachatryan, V. and others (CMS Collaboration), Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at $\sqrt{s} = 8$ TeV, CMS-PAS-EXO-12-024, 2013.

[83] Ciuchini, M., Franco, E., Mishima, S., and Silvestrini, L., JHEP **1308** (2013) 106.

[84] Barger, V., Keung, W. Y., and Ma, E., Phys. Rev. D **22** (1980) 727.

[85] Li, H.-L., Si, Z.-G., Yang, X.-Y., Yang, Z.-J., and Zheng, Y.-J., Phys. Rev. D **87** (2013) 115024.

[86] Khachatryan, V. and others (CMS Collaboration), Phys. Lett. B **748** (2015) 255.

[87] Dasgupta, M., Magnea, L., and Salam, G. P., JHEP **0802** (2008) 055.

[88] Ellis, S. D., Vermilion, C. K., and Walsh, J. R., Phys. Rev. D **80** (2009) 051501.

[89] Butterworth, J. M., Davison, A. R., Rubin, M., and Salam, G. P., Phys. Rev. Lett. **100** (2008) 242001.

[90] Ellis, S. D., Vermilion, C. K., and Walsh, J. R., Phys. Rev. D **81** (2010) 094023.

[91] Krohn, D., Thaler, J., and Wang, L.-T., JHEP **1002** (2010) 084.

[92] Thaler, J. and Van Tilburg, K., JHEP **03** (2011) 015.

[93] Butterworth, J., Cox, B., and Forshaw, J. R., Phys. Rev. D **65** (2002) 096014.

[94] Aad, G. and others (ATLAS Collaboration), Phys. Rev. D **86** (2012) 072006.

[95] Almeida, L. G. et al., Phys. Rev. D **79** (2009) 074017.

[96] Ellis, S. D., Hornig, A., Roy, T. S., Krohn, D., and Schwartz, M. D., Phys. Rev. Lett. **108** (2012) 182003.

[97] Larkoski, A. J., Salam, G. P., and Thaler, J., JHEP **1306** (2013) 108.

[98] Krohn, D., Schwartz, M. D., Lin, T., and Waalewijn, W. J., Phys. Rev. Lett. **110** (2013) 212001.

[99] Aad, G. and others (ATLAS Collaboration), JHEP **1309** (2013) 076.

[100] Chatrchyan, S. and others (CMS Collaboration), JHEP **1305** (2013) 090.

[101] Khachatryan, V. and others (CMS Collaboration), Identifying Hadronically Decaying Vector Bosons Merged into a Single Jet, CMS-PAS-JME-13-006, 2013.

[102] Aad, G. and others (ATLAS Collaboration), Performance of Boosted W Boson Identification with the ATLAS Detector, ATL-PHYS-PUB-2014-004, 2014.

[103] Altarelli, G., Mele, B., and Ruiz-Altaba, M., Z. Phys. C **45** (1989) 109.

[104] Baur, U., Hinchliffe, I., and Zeppenfeld, D., Int. J. Mod. Phys. A **2** (1987) 1285.

[105] Baur, U., Spira, M., and Zerwas, P., Phys. Rev. D **42** (1990) 815.

[106] Randall, L. and Sundrum, R., Phys. Rev. Lett. **83** (1999) 3370.

[107] Randall, L. and Sundrum, R., Phys. Rev. Lett. **83** (1999) 4690.

[108] Khachatryan, V. and others (CMS Collaboration), *V* Tagging Observables and Correlations, CMS-PAS-JME-14-002, 2014.

[109] Khachatryan, V. and others (CMS Collaboration), Pileup Removal Algorithms, CMS-PAS-JME-14-001, 2014.

[110] Alwall, J., Herquet, M., Maltoni, F., Mattelaer, O., and Stelzer, T., JHEP **06** (2011) 128.

[111] Alwall, J. et al., Comput. Phys. Commun. **176** (2007) 300.

[112] Sjostrand, T., Mrenna, S., and Skands, P. Z., JHEP **0605** (2006) 026.

[113] Jadach, S., Was, Z., Decker, R., and Kuhn, J. H., Comput. Phys. Commun. **76** (1993) 361.

[114] Agostinelli, S. et al., Nucl. Instrum. Meth. A **506** (2003) 250.

[115] Nason, P., JHEP **11** (2004) 040.

[116] Frixione, S., Nason, P., and Oleari, C., JHEP **11** (2007) 070.

[117] Alioli, S., Nason, P., Oleari, C., and Re, E., JHEP **06** (2010) 043.

[118] Alioli, S., Moch, S.-O., and Uwer, P., JHEP **01** (2012) 137.

[119] Cacciari, M., Salam, G. P., and Soyez, G., JHEP **0804** (2008) 063.

[120] Cacciari, M. and Salam, G. P., Phys. Lett. B **659** (2008) 119.

[121] Cacciari, M., Salam, G. P., and Soyez, G., Eur. Phys. J. C **72** (2012) 1896.

[122] Khachatryan, V. and others (CMS Collaboration), Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET, CMS-PAS-PFT-09-001, 2009.

[123] Khachatryan, V. and others (CMS Collaboration), Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector, CMS-PAS-PFT-10-001, 2010.

[124] Chatrchyan, S. and others (CMS Collaboration), Eur. Phys. J. C **70** (2010) 1165.

[125] Chatrchyan, S. and others (CMS Collaboration), JINST **9** (2014) P10009.

[126] Chatrchyan, S. and others (CMS Collaboration), JINST **7** (2012) P01001.

[127] Wobisch, M. and Wengler, T., Hadronization corrections to jet cross-sections in deep inelastic scattering, hep-ph/9907280, 1999.

[128] Ellis, S. D. and Soper, D. E., Phys. Rev. D **48** (1993) 3160.

[129] Catani, S., Dokshitzer, Y. L., Seymour, M., and Webber, B., Nucl. Phys. B **406** (1993) 187.

[130] Ellis, S. D., Vermilion, C. K., and Walsh, J. R., Phys. Rev. D **80** (2009) 051501.

[131] Khachatryan, V. and others (CMS Collaboration), JINST **6** (2011) P11002.

[132] Chatrchyan, S. and others (CMS Collaboration), Phys. Rev. Lett. **106** (2011) 231801.

[133] Punzi, G., Sensitivity of searches for new signals and its optimization, in *Phy-Stat2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, edited by Lyons, L., Mount, R. P., and Reitmeyer, R., 2003.

[134] ATLAS and CMS Collaborations, Procedure for the LHC Higgs boson search combination in summer 2011, Technical Report ATL-PHYS-PUB-2011-011, CERN, 2011.

[135] Estimating Systematic Errors Due to Pileup Modeling (CMS Collaboration) (unpublished),
`https://twiki.cern.ch/twiki/bin/view/CMS/`
`PileupSystematicErrors`.

[136] Jet energy resolution (CMS Collaboration) (to be published),
`https://twiki.cern.ch/twiki/bin/viewauth/CMS/`
`JetResolution`.

[137] Tau POG Twiki (CMS Collaboration) (to be published),
`https://twiki.cern.ch/twiki/bin/viewauth/CMS/`
`TauIDRecommendation`.

[138] Tau ID Systematics (CMS Collaboration) (to be published),
`https://indico.cern.ch/event/320294/contribution/1/`
`material/slides/0.pdf`.

[139] Read, A. L., J. Phys. G **28** (2002) 2693.

[140] Junk, T., Nucl. Instrum. Meth. A **434** (1999) 435.